Article

The collective praise intervention: A brief intervention highlighting prosocial behavior reduces hostility towards Muslims

Roman A. Gallardo,[1] Boaz Hameiri,[2] Samantha L. Moore-Berg[1,3] and Emile Bruneau[1,3]

Abstract

Muslims are consistently the target of dehumanization and hostile policies. Previous research shows that interventions that highlight the hypocrisy wherein people collectively blame entire groups for the heinous acts of individual members of outgroups but not ingroups are effective in reducing animosity towards Muslims. However, these interventions rely on aversive materials (e.g., terrorist acts), which can hinder the scalability of interventions due to individuals' tendency to resist/avoid challenging and aversive stimuli. In three preregistered studies (combined $N = 2,635$), we developed and tested a nonaversive, hypocrisy-based intervention that highlights the hypocrisy involved in attributing ingroup members' prosocial acts to the entire ingroup (i.e., Christians) but not doing the same for outgroup members (i.e., Muslims). Results indicated that this collective praise intervention reliably reduces dehumanization of Muslims, anti-Muslim policy support, and collective blame of Muslims. We argue that the intervention's use of nonaversive stimuli allows for both practical and scalable applications.

Muslims experience constant discrimination globally (e.g., China built concentration camps for over one million Muslims: Simmons, 2019; India passed a citizenship law that openly discriminates against Muslims: Samuel, 2019; European countries approved laws banning headscarves [hijabs] in many settings: Abdelgadir & Fouka, 2019). U.S. Muslims have similarly been the subject of discriminatory policies (e.g., the U.S. travel ban, which suspended entry to the US from Muslim-majority countries and resulted in the separation of thousands of Muslim families by preventing

overseas Muslims from reuniting with their families in the US; Niayesh, 2019). In fact, half of

[1]University of Pennsylvania, USA
[2]Tel Aviv University, Israel
[3]Beyond Conflict Innovation Lab, USA

Corresponding authors:
Roman A. Gallardo, Annenberg School for Communication, University of Pennsylvania, PA, USA.
Boaz Hameiri, The Evens Program in Conflict Resolution and Mediation, Tel Aviv University, Israel.
Emails: roman.gallardo@asc.upenn.edu; bhameiri@tauex.tau.ac.il

Muslim Americans state that in recent years it has become more difficult to be Muslim in America, and roughly three quarters of Muslims report experiencing "a lot of discrimination" (Lipka, 2017; Diamant, 2017). Research suggests that one driver of such discrimination is dehumanization of Muslims (Kteily & Bruneau, 2017b; Kteily et al., 2015). Thus, given the detrimental consequences of dehumanization, we sought to develop a novel intervention that combats dehumanization of Muslims and anti-Muslim policy support.

Dehumanization is the denial of full humanness (e.g., Haslam, 2006; Haslam et al., 2008) and predicts hostile attitudes, behaviors, and policies (e.g., Kteily & Bruneau, 2017b). For example, Kteily et al. (2015) found that non-Muslim Americans, on average, explicitly rated Muslims as less "evolved" than their own group, which predicted support for harmful policies (e.g., preventing new Mosques from opening; see also Bruneau & Kteily, 2017; Kteily & Bruneau, 2017a). One contributing factor to dehumanization of groups is whether group values are perceived to reflect warmth and competence (Haslam & Loughnan, 2014; Vaes & Paladino, 2010). For instance, when people perceive a group as lacking the ability to engage in prosocial behavior (e.g., providing care or help), they are more likely to dehumanize the group and thus express hostility towards it (Esses et al., 2013). As another contributing factor, a sense of threat contributes to dehumanization of groups (Kteily & Bruneau, 2017b). Therefore, interpreting a group as violent or threatening can potentially lead to dehumanization, whereas interpreting a group as warm or benevolent may reduce dehumanization.

Recent research suggests that preexisting views on a group and the valence of an act committed by a member(s) of this group are pivotal to how individuals view those who committed the act, as well as to the extent to which they will extrapolate their views to the entire group (for a recent meta-analysis, see Paolini & McIntyre, 2019). For example, Adelman et al. (2019) found an interaction between participants' preexisting attitudes towards

Muslims and the valence of a news story about Muslims they were exposed to on the extent to which they extrapolated the actions described in the news story to all Muslims. Specifically, participants who held negative preexisting attitudes towards Muslims and were given a negative (but not positive or neutral) news story held all Muslims collectively responsible for the negative action described in the story. Furthermore, participants who held positive preexisting attitudes towards Muslims and were given a positive news story held Muslims collectively responsible for the positive actions. Thus, individuals' preexisting views on groups interact with the nature of the acts committed by members of those groups, which predicts perceived collective responsibility of entire group for those actions.

Although cognitive biases exacerbate conflicts, making people aware of these cognitive biases may promote better intergroup relations (see Bar-Tal & Hameiri, 2020). For example, exposing individuals to inconsistencies or hypocrisies in their judgments of actions performed by their ingroup versus an outgroup can reduce intergroup hostility (Bruneau et al., 2018; Bruneau, Kteily, & Urbiola, 2020; Shulman et al., 2020). For instance, Bruneau et al., 2018; Bruneau, Kteily, & Urbiola, 2020) examined the hypocrisy involved in collectively blaming all Muslims but not all White people for individual acts of violence committed by individual group members. Specifically, Bruneau et al., 2018; Bruneau, Kteily, & Urbiola, 2020) used an interactive activity that employed a Socratic approach by having non-Muslim participants first report how much they blame themselves and collectively blame White Americans for terrorist attacks committed by self-identified White people. Next, participants were asked to report how much they blame individual Muslims and collectively all Muslims for a terrorist attack committed by self-identified Muslims. This collective blame hypocrisy (CBH) activity reliably reduced collective blame placed on Muslims and, in turn, anti-Muslim sentiments both in the US (Bruneau et al., 2018) and in Spain, with effects persisting for at least 1 year (Bruneau, Kteily, & Urbiola, 2020).

Although this hypocrisy-based intervention showed promising results, we argue that it relies on participants' motivation to identify biases in their thinking, on their motivation to correct these biases (Wegener & Petty, 1997), and on their willingness to take part in the interventions altogether (Bar-Tal & Hameiri, 2020). However, the use of aversive stimuli to induce perception of hypocrisy can lead to reduced motivation to fully engage with the stimuli (e.g., Hay et al., 2015; Sheppes & Levin, 2013). For instance, people might use different strategies to distract themselves from the aversive stimuli, to reduce negative feelings (Pliskin et al., 2018). Thus, using such stimuli could significantly hinder intervention scalability.

To address this limitation, we developed a novel intervention, the collective praise (CP) intervention, that aims to reduce hostility towards Muslims using nonaversive stimuli. Similar to previous work, the CP intervention exposes the hypocrisy in differentially interpreting similar acts committed by outgroup and ingroup members. Unlike previous work (e.g., Bruneau et al., 2018; Bruneau, Kteily, & Urbiola, 2020), the CP approach reveals a cognitive bias using positive, nonaversive stimuli. The CP approach highlights the inconsistency involved in not attributing prosocial behaviors of outgroup members to the entire outgroup but attributing prosocial behaviors of ingroup members to the entire ingroup.

Across three preregistered studies conducted with American participants, we designed and tested the effectiveness of the CP intervention to decrease dehumanization of Muslims, anti-Muslim policy support, and collective blame placed on Muslims. In Study 1, we tested different iterations of the CP intervention by manipulating the presentation order of Muslim and Christian vignettes to determine which presentation order most effectively reduces anti-Muslim hostility. In Study 2, we compared the effectiveness of the best performing CP iteration from Study 1 to that of the CBH intervention to assess the effectiveness of an intervention that uses positive stimuli (CP intervention) compared to an intervention that uses negative stimuli (CBH intervention).

Finally, in Study 3, we separated the different components of the CP intervention to identify the key ingredients driving its effects. We argue that if hypocrisy perception is driving the CP intervention's effects, both Muslim and Christian vignettes would be necessary to reduce hostility towards Muslims.

We hypothesized that the hypocrisy perception evoked in the CP intervention (vs. a control) would result in psychological discomfort, which would make people reevaluate their cognitive biases, resulting in less dehumanization of Muslims and decreased anti-Muslim policy support. As previous research suggests, dehumanization is an important contributor to anti-Muslim policy support (Kteily et al., 2016), thus we also hypothesized that the CP intervention would have an indirect effect on anti-Muslim policy support through dehumanization. Finally, following CBH work (Bruneau et al., 2018; Bruneau, Kteily, & Urbiola, 2020), we also hypothesized that the CP activity would lead to lower levels of collective blame placed on Muslims.

Study 1

The goal of Study 1 was to examine whether the CP approach (vs. a control) reduces hostility towards Muslims. To determine the most efficient CP intervention iteration, we operationalized it into three conditions. We predicted that taking part in one of the experimental conditions would improve attitudes towards Muslims, reflected in (a) reduced dehumanization, (b) reduced support for anti-Muslim policies, and (c) reduced collective blame put on Muslims. In all studies, comparisons between experimental conditions were exploratory.

Methods

Participants. We performed a power analysis using G*Power 3.1 (Faul et al., 2009), and found that when comparing two independent groups, 76 participants per condition were needed to obtain a medium effect ($d = 0.50$) with .95 power, and a sample of 436 per condition was required to

detect a small effect ($d = 0.20$). To detect small to medium effects, we recruited 200 participants per condition via Amazon's Mechanical Turk (MTurk), which, according to a sensitivity power analysis, allowed detection of a small effect size $d = 0.37$ with .95 power. Eight hundred participants completed our study and, after removing 17 participants who failed an embedded attention check question and nine participants who identified as Muslim, we had 774 participants in the final sample (47.8% female; $M_{age} = 36.02$, $SD = 11.68$; 37.2% Republican, 62.8 % Democrat; 67.3% White, 13.7% Black, 7.9% Asian, 5.9% Hispanic, 5.2% other; 54.7% Christian, 39.1% atheist/agnostic, 6.2% other). Participants were compensated US$1.00 for their participation.

Measures and procedure. Participants were randomly assigned to one of three experimental conditions: (a) Muslim–Christian ($n = 199$), (b) Christian–Muslim ($n = 196$), (c) random order ($n = 195$), or to a control ($n = 184$). All study demographic information by condition is presented in the supplemental material. In the Muslim–Christian condition, participants first read three different, randomly presented vignettes of recent prosocial acts performed by Muslim organizations, followed by three different, randomly presented vignettes of recent prosocial behaviors performed by Christian organizations. Each vignette was accompanied by a relevant image. After each individual vignette, participants responded on unmarked sliders anchored at 0 (*not at all*) and 100 (*very*) to the question "How much should the actions of these [Muslim/Christian] organizations reflect upon [Muslims/Christians], in general?" In the Christian–Muslim condition, we presented the vignettes in the reverse order, and in the random condition all vignettes were randomized. Control condition participants did not receive any activity.

Following the CP interventions or control, all participants completed our key outcome measures of dehumanization, anti-Muslim policy support (Kteily & Bruneau, 2017a; Kteily et al., 2015), and collective blame (Bruneau et al., 2018; Bruneau, Kteily, & Urbiola, 2020). Additionally,

we collected demographics. Study 1 was fully preregistered before data collection[1].

Dehumanization was assessed with two measures: Ascent of (Hu)Man (i.e., ascent dehumanization; Kteily et al., 2015) and a multi-item trait dehumanization (i.e., trait dehumanization; Bastian et al., 2013). Ascent of Hu(Man) assessed perceptions of how "evolved and civilized" Muslims and Christians (among other filler groups) are perceived to be. Target groups were presented in random order, and participants made such judgments using a sliding scale anchored at 0 ("least evolved" image) and 100 ("most evolved" image). The ascent dehumanization score was calculated by subtracting the ascent rating of Muslims from that of Christians. Trait dehumanization was assessed with a series of 15 items, including eight reverse-scored (r) ones (e.g., "refined" (r), "cultured" (r), "backward," "primitive"). Participants characterized Christians ($\alpha = .92$) and Muslims ($\alpha = .91$) on these traits using a 7-point scale (1 = *not at all*, 7 = *extremely so*). Trait ratings were averaged to create a separate trait dehumanization score for Christians and Muslims. The trait dehumanization score was calculated by subtracting the average rating of Christians from the average rating of Muslims. Finally, the ascent and trait dehumanization scores were standardized and averaged together to create a composite dehumanization score ($r = .56$, $p < .001$; for a similar approach, see Bruneau, Hameiri, et al., 2020).[2]

Anti-Muslim policy support was assessed by asking participants to indicate their support for nine policies targeting Muslims, adapted from Kteily and Bruneau (2017b) and partly inspired by campaign statements of (then presidential candidate) Donald Trump (e.g., "We should ban the opening of any new Mosques in this country"). Responses were given on a 7-point Likert-type scale (1 = *strongly disagree*, 7 = *strongly agree*). We averaged responses together to create a composite anti-Muslim policy support score ($\alpha = .96$).

Collective blame was assessed by presenting participants with a brief description of the Belgium terror attacks ("In March 2016, three Muslim extremists killed 32 people and wounded

Table 1. Means, standard deviations, and correlations for all variables: Study 1.

	M	SD	1	2	3	4	5	6
1. Dehumanization	−0.005	1.75	–					
2. Anti-Muslim policy support	2.63	1.70	.57***	–				
3. Collective blame	30.62	32.81	.42***	.69***	–			
4. Political affiliation	−0.53	2.23	.35***	.45***	.38***	–		
5. Age	36.02	11.68	.09**	.06	−.01	.10**	–	
6. Gender (F = 0, M = 1)	–	–	−.05	−.09*	−.10**	−.07*	.14***	–

Note. N = 774.
*p < .05. **p < .01. ***p < .001.

over 300 at the Brussels airport in Belgium. How responsible do you think Muslims are for the bombing of the airport in Brussels, Belgium in 2016?"; Bruneau et al., 2018) and then having them report how responsible they felt "Belgian Muslims" and "Muslims in general" were for the attacks, using unmarked sliders anchored at 0 (*not at all*) and 100 (*very much*). We used participants' responses to how responsible they felt "Muslims in general" were for the attacks as our collective blame variable. Demographics of race, age, gender, religion, and political affiliation were collected.

Results

Given the centrality of ideology in predicting Islamophobic attitudes and behaviors in the US (Ogan et al., 2014), we first examined whether participants differed by political affiliation across conditions. A one-way ANOVA revealed no differences in political affiliation across conditions, $F(3, 770) = 1.70$, $p = .166$, $\eta^2_p = .01$. However, since we detected a failure of equal randomization of political affiliation in Study 2, for consistency, we controlled for political affiliation throughout all statistical analyses to ensure results were driven by the manipulation rather than political affiliation. Results from all studies were similar without controlling for political affiliation (see supplemental material). See means, standard deviations, and variable intercorrelations in Table 1.

To test our main hypotheses, in all studies, we conducted a series of planned contrasts to examine the differences in dehumanization, anti-Muslim policy support, and collective blame between the experimental conditions and control. For means and standard deviations by condition, see Table 2. We also tested whether the intervention affected anti-Muslim policy support through dehumanization in all experimental conditions, using Hayes's (2018) PROCESS Model 4. We included path models for all significant conditions versus the control. In order to replicate the CBH work (Bruneau et al., 2018; Bruneau, Kteily, & Urbiola, 2020), and as preregistered, we examined collective blame as an outcome variable only.
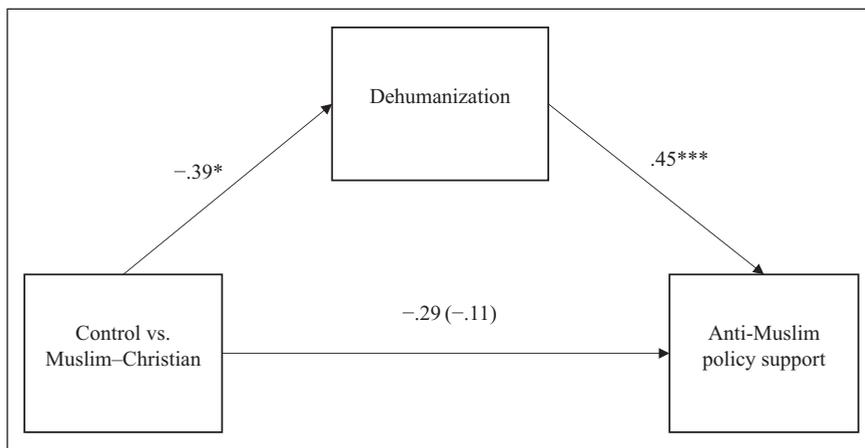
Comparing the Muslim–Christian condition to the control, dehumanization of Muslims was significantly lower, $t(740) = 2.29$, $p = .022$, $d = 0.21$, and there was no significant difference in anti-Muslim policy support, $t(769) = 1.95$, $p = .052$, $d = 0.18$. There was also a significant indirect effect of the Muslim–Christian condition (vs. the control) on anti-Muslim policy support through dehumanization (indirect effect = −.17, $SE = 0.09$, 95% CI [−0.35, −0.01]; see Figure 1). Finally, collective blame placed on Muslims was also significantly lower in the Muslim–Christian condition than in the control, $t(769) = 3.29$, $p = .001$, $d = 0.31$.

Comparing the Christian–Muslim condition to the control, there were no significant differences in dehumanization of Muslims, $t(740) = 0.43$, $p = .668$, $d = 0.04$; anti-Muslim policy support, $t(769) = 1.18$, $p = .238$, $d = 0.11$; or collective blame put on Muslims, $t(769) = 1.80$, $p = .072$, $d = 0.16$.

Table 2. Means and standard deviations for each measure, by condition: Study 1.

Condition	Dehumanization M (SD)	Anti-Muslim policy support M (SD)	Collective blame M (SD)
Control (n = 184)	0.09 (2.11)[a]	2.78 (1.76)[a]	35.83 (35.71)[a]
Muslim–Christian (n = 199)	−0.11 (1.53)[b]	2.48 (1.65)[a]	26.24 (30.22)[b]
Christian–Muslim (n =196)	0.05 (1.72)[ab]	2.60 (1.66)[a]	32.51 (33.95)[ab]
Random (n = 195)	−0.04 (1.58)[ab]	2.68 (1.74)[a]	28.26 (30.68)[b]

Note. Means not sharing the same superscript are significantly different from each other at $p < .05$; superscript comparisons are per variable.

Figure 1. Unstandardized path model assessing the effect of the Muslim–Christian intervention on anti-Muslim policy support through dehumanization: Study 1.

Note. *$p < .05$. ***$p < .001$.

Random condition

Comparing the random condition to the control, there were no significant differences in dehumanization of Muslims, $t(740) = 1.45$, $p = .146$, $d = 0.13$, or anti-Muslims policy support, $t(769) = 0.67$, $p = .502$, $d = 0.06$; however, collective blame placed on Muslims was significantly lower in the random condition than in the control, $t(769) = 2.33$, $p = .020$, $d = 0.22$.

Discussion

Study 1 demonstrated that participants in the Muslim–Christian condition (vs. the control) scored significantly lower on dehumanization and collective blame placed on Muslims. There was also an indirect effect of the Muslim–Christian condition (vs. the control) on anti-Muslim policies through dehumanization. Those in the Christian–Muslim and the random conditions did not differ from the control on either dehumanization of Muslims or on anti-Muslim policy support. Participants in the random condition scored significantly lower on collective blame; however, the Muslim–Christian condition had stronger effects on collective blame and showed more potential for reducing hostility towards Muslims across all outcome variables. Finally, there was no difference in collective blame between the Christian–Muslim and the control conditions. Results from Study 1 offer

partial support for our main hypothesis, as only the Muslim–Christian condition demonstrated consistent effects.

We argue that the Muslim–Christian condition was effective because participants in this condition experienced cognitive dissonance; they were not able to go back and change their ratings of Muslims' prosocial behaviors after rating Christians' prosocial behaviors. This is in contrast to the Christian–Muslim condition in which participants first rated Christians' prosocial behaviors before rating Muslims' prosocial behaviors. Rating Muslims after Christians potentially allowed participants to reduce their perceived hypocrisy by rating Muslims and Christians similarly. Although it is plausible that participants could have changed their ratings of Christians to correspond with how they rated Muslims in the Muslim–Christian condition, we argue that even if they had done so, it would not have translated to more negative views in general, and dehumanization in particular, on Christians, as individuals try to maintain a positive ingroup identity (Leippe & Eisenstadt, 1994; Tajfel & Turner, 1979).

Study 2

In Study 2, we sought to replicate the effects of the successful Muslim–Christian CP iteration and compare it to a similar task asking participants to consider the collective blame of Muslims and Christians for individual acts of violence committed by Muslim or Christian extremists, respectively (Bruneau et al., 2018; Bruneau, Kteily, & Urbiola, 2020). Thus, we compared the effectiveness of an intervention that uses positive stimuli (CP) to that of an intervention that uses negative stimuli (CBH). We also explored the effectiveness of an intervention that combined the CP and CBH approaches to examine the impact of combined positive and negative stimuli. Thus, Study 2 had three experimental conditions: Muslim–Christian (i.e., CP activity) only, CBH only, and the CP–CB combined condition. We preregistered that all conditions (vs. the control) would significantly reduce dehumanization of Muslims, anti-Muslim policy support, and collective blame placed on Muslims;

however, the analysis of which condition (vs. the control) would have the strongest effect on reducing hostility towards Muslims was exploratory. We speculated the CBH condition would be more effective than the CP condition in reducing anti-Muslim attitudes.

Additionally, Bruneau, Kteily, & Urbiola (2020) found that the CBH intervention was moderated by preference for consistency (PFC; Cialdini et al., 1995). Specifically, they found that participants higher in PFC were especially susceptible to the activity, which suggests the dissonance that participants experienced after being confronted with their hypocrisy was the driving mechanism behind the CBH intervention. Therefore, we tested whether PFC moderated the results for each experimental condition.

Additionally, we examined two alternative mechanisms. First, we examined whether being exposed to any intervention would stimulate participants' sense of being shamed, as the study attempts to make them aware of their hypocrisy in differentially rating Christians and Muslims, which could, in turn, increase reactance and reduce the effectiveness of the interventions (Brehm, 1966). Second, past research also suggests that processes of cognitive unfreezing can increase individuals' openness to information, and thus shift attitudes (e.g., Hameiri et al., 2018). Therefore, we examined whether being exposed to any intervention might challenge the validity of participants' previously held views regarding Muslims, thereby stimulating cognitive unfreezing of their beliefs. Thus, we tested feelings of being shamed as either a moderator or a mediator, and cognitive unfreezing as a mediator.

Methods

Participants. We used the same power and sensitivity analyses as in Study 1, which revealed that our final sample would be sufficiently powered to detect a minimum effect size of $d = 0.28$ with .95 power. Eight hundred and fifty-four participants completed our study. We excluded 35 participants for failing at least one of two embedded attention check questions and 12 participants who identified as Muslim, leaving 807

participants in our final sample (36.1% female; M_{age} = 35.72, SD = 10.38; 42% Republican, 57.9% Democrat; 65.3% White, 19.7% Black, 6.1% Asian, 5.3% Hispanic, 3.6% other; 58.0% Christian, 34.8% atheist/agnostic, 7.2% other). Participants were compensated US$1.50.

Measures and procedure. Participants were first randomly assigned to one of three experimental conditions: (a) Muslim–Christian (*n* = 206), (b) CBH (*n* = 200), (c) CP–CB (*n* = 200), or control (*n* = 201; see supplemental material for study demographics by condition). The Muslim–Christian condition was the same best performing iteration of the CP activity from Study 1. The CBH intervention was derived from Bruneau et al. (2018); Bruneau, Kteily, & Urbiola (2020) and was composed of two parts. First, participants were asked to read three different acts of violence committed by Christian individuals. To foreshadow the comparison of violence committed by Christian extremists to violence committed by Muslim extremists, participants learned that each perpetrator was motivated by his Christian beliefs. Then, participants rated the extent to which they felt responsible and the extent to which they felt Christians in general were responsible for this act of violence, using unmarked sliders anchored at 0 (*not at all*) and 100 (*very*). As a note, we adapted the original CBH intervention to highlight Christian identities and motivations in lieu of White supremacist motivations, to bring the CBH intervention to greater parity with the CP one. Next, we asked participants to report, using the same scale, how responsible they felt individual Muslims, as well as Muslims in general, were for acts of violence committed by Muslim extremists. The CP–CB intervention consisted of both the CP and the CBH blocks being presented in a randomized order, with vignettes randomized within blocks as well. Within the CP–CB intervention, we tested for any differences in our outcome variables between the two randomized versions (i.e., CBH followed by CP; CP followed by CBH) and found no differences. Thus, we combined responses across these conditions and treated them as one condition.

Following the intervention (or control), all participants completed the same dehumanization and collective blame measures as in Study 1, and a revised measure of anti-Muslim policy support. Additionally, we assessed PFC, sense of feeling shamed, and cognitive unfreezing to examine the psychological mechanisms driving the effects of each experimental condition. We also preregistered several exploratory variables that were not the focal point of this study and therefore are not reported. Study 2 was fully preregistered before data collection.[3]

Dehumanization (trait dehumanization: Christians: α = .88; Muslims: α = .91; correlation between ascent and trait dehumanization measures: r = .64, p < .001) and collective blame were assessed as in Study 1.

Anti-Muslim policy support was assessed as in Study 1 but with slightly different items. In particular, participants were asked to indicate their support for five policies targeting Muslims, adapted from previous scales (e.g., "We need to stop accepting Muslim immigrants into this country, period"; see Kteily & Bruneau, 2017a; Lee et al., 2009; Schildkraut, 2009). Responses were given on a 7-point Likert-type scale (1 = *strongly disagree*, 7 = *strongly agree*) averaged to create a composite anti-Muslim policy support score (α = .83).

Preference for consistency (PFC) was assessed using an abridged, six-item scale developed by Cialdini et al. (1995; e.g., "I want to be described by others as a stable, predictable person"), answered on a 7-point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*). We averaged responses together to create a composite PFC score (α = .79).

Cognitive unfreezing was assessed using a four-item scale derived from Hameiri et al. (2018), assessing the extent, from 0 (*not at all*) to 100 (*very much so*), to which the survey made them doubt their preexisting ideas about Muslims (i.e., "Muslims hate many of the freedoms that we hold dear in the US"; α = .94).

Sense of feeling shamed was assessed with one item asking participants to indicate their perception of how proud (1) versus how ashamed (7) of their American identity the people who

Table 3. Means, standard deviations, and correlations for all variables: Study 2.

	M	SD	1	2	3	4	5	6	7	8	9
1. Dehumanization	−0.003	0.86	–								
2. Anti-Muslim policy support	3.14	1.97	.43***	–							
3. Collective blame	36.58	35.42	.34***	.76***	–						
4. Preference for consistency	5.01	1.05	.12***	.25***	.22***	–					
5. Feeling shamed	4.07	1.70	.02	−.06	−.04	−.04	–				
6. Cognitive unfreezing	36.72	31.82	−.02	−.07*	−.06	−.08*	.22***	–			
7. Political affiliation	−0.34	2.38	.29***	.42***	.33***	.22***	.00	−.03	–		
8. Age	35.72	10.38	.08*	−.10**	−.10**	.02	.11**	.03	.03	–	
9. Gender (F = 0, M = 1)	–	–	−.05	−.16***	−.16***	.01	.03	.03	−.09**	.15***	–

Note. N = 807.
*p < .05. **p < .01. ***p < .001.

created this survey want them to feel. Demographics were collected as in Study 1.

Results

As in Study 1, we examined whether the conditions differed by political affiliation. A one-way ANOVA revealed that political affiliation significantly differed across conditions, $F(3, 800) = 3.36$, $p = .018$, $\eta^2_p = .01$. We therefore controlled for political affiliation. For means, standard deviations, and variable intercorrelations, see Table 3.

For means and standard deviations by condition, see Table 4. As in Study 1, we also tested whether the intervention versus control affected anti-Muslim policy support through dehumanization, using Hayes's (2018) PROCESS Model 4. Additionally, we tested for moderation of PFC and feeling shamed (only if no direct effects were found vs. the control on feeling shamed) on the relationship between the experimental conditions versus the control and our outcome variables, using Hayes's (2018) PROCESS Model 1. Finally, as part of additional analyses we conducted, we examined the effects of the different experimental conditions versus control on cognitive unfreezing and feeling shamed. If direct effects were found, we used Hayes's (2018) PROCESS Model 4 to test whether the interventions versus the control affected our outcome variables through cognitive unfreezing and feeling shamed.

Muslim-Christian condition

When comparing the Muslim–Christian condition to the control, dehumanization of Muslims was significantly lower, $t(799) = 2.41$, $p = .016$, $d = 0.23$, but there was no significant difference between these conditions in terms of anti-Muslim policy support, $t(799) = 1.49$, $p = .136$, $d = 0.14$. There was also a significant indirect effect of the Muslim–Christian condition (vs. the control) on anti-Muslim policy support through dehumanization (indirect effect = −.15, $SE = 0.06$, 95% CI [−0.28, −0.03]; see Figure 2A). Collective blame placed on Muslims was also significantly lower in the Muslim–Christian condition than in the control, $t(799) = 1.97$, $p = .049$, $d = 0.19$.
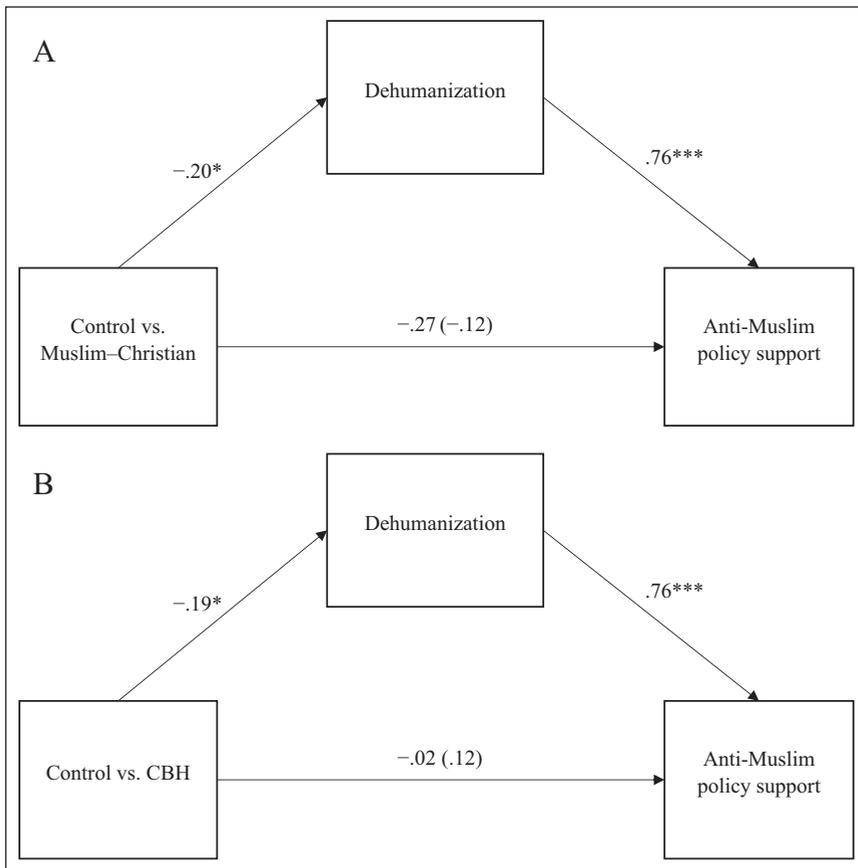
There was a significant Condition x PFC interaction on dehumanization of Muslims ($b = −0.15$, $SE = 0.08$, 95% CI [−0.30, < 0.01]), such that those with higher levels of PFC had lower scores on dehumanization in the intervention versus the control condition ($b = −0.35$, $SE = 0.11$, 95% CI [−0.57, −0.13]), whereas for low-PFC participants, there was no difference in dehumanization between these conditions ($b = −0.03$, $SE = 0.12$, 95% CI

Table 4. Means and standard deviations for each measure, by condition: Study 2.

Condition	Dehumanization M (SD)	Anti-Muslim policy support M (SD)	Collective blame M (SD)	Feeling shamed M (SD)	Cognitive unfreezing M (SD)
Control (n = 201)	0.13 (0.99)[a]	3.23 (2.03)[a]	44.48 (35.34)[a]	4.01 (1.58)[a]	30.98 (30.29)[a]
Muslim–Christian (n = 206)	−0.07 (0.70)[b]	2.96 (1.87)[a]	36.08 (33.89)[b]	4.06 (1.62)[a]	36.07 (32.17)[b]
CBH (n = 200)	−0.06 (0.90)[b]	3.20 (1.98)[a]	32.66 (35.91)[b]	4.06 (1.79)[a]	42.04 (31.30)[ab]
CP–CB (n = 200)	−0.02 (0.80)[ab]	3.18 (2.01)[a]	33.60 (35.66)[b]	4.15 (1.80)[a]	37.36 (32.65)[ab]

Note. Means not sharing the same superscript are significantly different from each other at $p < .05$; superscript comparisons are per variable. CBH = collective blame hypocrisy; CP–CB = collective praise–collective blame.

Figure 2. Unstandardized path model assessing the effects of the (A) Muslim–Christian and (B) the CBH interventions on anti-Muslim policy support through dehumanization: Study 3.

Note. $*p < .05$. $***p < .001$.

[−0.26, 0.20]). There was no significant Condition x PFC interaction on anti-Muslim policy support (b = −0.04, SE = 0.17, 95% CI [−0.37, 0.29]) or collective blame put on Muslims (b = 2.80, SE = 3.13, 95% CI [−3.34, 8.94]).

Collective blame hypocrisy (CBH) condition. Compared to the control, dehumanization of Muslims was significantly lower in the CBH condition, $t(799)$ = 2.33, p = .020, d = 0.20, but there was no significant difference between these conditions in anti-Muslim policy support, $t(799)$ = 0.12, p = .905, d = 0.01. There was also a significant indirect effect of the CBH condition (vs. the control) on anti-Muslim policy support through dehumanization (indirect effect = −.15; SE = 0.07, 95% CI [−0.29, −0.01]; see Figure 2B). Collective blame placed on Muslims was also significantly lower in the CBH condition than in the control, $t(799)$ = 2.75, p = .006, d = 0.26.

There was a significant Condition x PFC interaction on dehumanization of Muslims (b = −0.24, SE = 0.08, 95% CI [−0.40, −0.09]), such that those with higher levels PFC had lower scores on dehumanization in the intervention versus the control condition (b = −0.44, SE = 0.11, 95% CI [−0.67, −0.22]), whereas for low-PFC participants, there was no difference (b = 0.07, SE = 0.12, 95% CI [−0.16, 0.30]). There was no significant Condition x PFC interaction on anti-Muslim policy support (b = −0.03, SE = 0.17, 95% CI [−0.36, 0.31]) or collective blame placed on Muslims (b = 1.15, SE = 3.18, 95% CI [−5.09, 7.38]).

Collective praise–collective blame (CP–CB) condition. Comparing the CP–CB condition to the control, there was no significant difference in dehumanization of Muslims, $t(799)$ = 1.84, p = .066, d = 0.17, or in anti-Muslim policy support, $t(799)$ = 0.26, p = .793, d = 0.02, but collective blame placed on Muslims was significantly lower in the CP–CB condition, $t(799)$ = 2.20, p = .028, d = 0.21.

There were no significant Condition x PFC interactions on dehumanization of Muslims (b = −0.05, SE = 0.07, 95% CI [−0.19, 0.10]),

anti-Muslim policy support (b = 0.09, SE = 0.16, 95% CI [−0.23, 0.40]), or collective blame put on Muslims (b = 3.45, SE = 3.02, 95% CI [−2.47, 9.37]).

Additional analyses. Compared to the control, no experimental conditions showed significant differences in sense of feeling shamed (ps > .399), and only the CBH condition significantly increased cognitive unfreezing (p < .001). However, the indirect effect of the CBH intervention (vs. the control) on all outcome variables through cognitive unfreezing was not significant (see supplemental material). There was also no Condition x Feeling Shamed interaction on any outcome variable (ps > .070; see supplemental material).

Discussion

Study 2 showed that both the Muslim–Christian and CBH conditions reduced dehumanization of Muslims, as compared to the control, although the effects were descriptively larger in the Muslim–Christian condition. Compared to the control, no experimental condition significantly reduced anti-Muslim policy support, but both the Muslim–Christian and CBH conditions indirectly reduced anti-Muslim policy support through dehumanization. All experimental conditions significantly reduced collective blame assigned to Muslims.

Despite the efficacy of the separate Muslim–Christian and CBH interventions, the combined CP–CB condition did not reduce dehumanization or anti-Muslim policy support. We speculate that this lack of effect could be due to simultaneous exposure to two contrasting hypocrisy-based attitudes (e.g., collectively blaming outgroups but not ingroups, and attributing the prosocial behavior of ingroup members but not of outgroup members to their respective groups). It could be that this double exposure facilitated hypothesis awareness, which may have caused psychological reactance (Brehm, 1966), consequently rendering the intervention ineffective (cf. Wohl et al., 2015).

Unexpectedly, we did not find the CBH intervention to be more effective than the CP

intervention. One potential explanation for the partial replication of previous studies assessing the effects of CBH intervention (Bruneau et al., 2018; Bruneau, Kteily, & Urbiola, 2020) might be that, in past studies, participants generally showed high levels of anti-Muslim policy support. This may be because these past studies were conducted in close temporal proximity to major terrorist attacks conducted by Muslim extremists (e.g., a series of attacks by Muslim extremists associated with ISIS in Barcelona and Cambrils; Bruneau, Kteily, & Urbiola, 2020). As stated previously, a contributing factor to dehumanization is a sense of threat. The relatively low levels of anti-Muslim sentiment in our control group likely yielded less reduction in anti-Muslim sentiment in the interventions by comparison, which may have hindered our capacity to obtain significant differences in anti-Muslim policy support between the intervention and the control condition. However, descriptively, participants in the Muslim–Christian condition had the lowest levels of anti-Muslim policy support and, along with the CBH condition, yielded significant indirect effects on support for anti-Muslim policies through dehumanization, as compared to the control.

Also, we would like to acknowledge the difference in vignette order between the Muslim–Christian and CBH conditions. Specifically, the Muslim–Christian condition first presents acts carried out by Muslims and then by Christians, while in the CBH condition the order is reversed. However, the CBH intervention's vignette order has not been explored in previous research; thus, we are unsure how reversing the order of vignettes in the CBH intervention will affect the results. Research on the CBH approach (Bruneau et al., 2018; Bruneau, Kteily, & Urbiola, 2020) shows that when the CBH intervention is presented in the order Christians then Muslims, it is effective in reducing hostility towards Muslims. Thus, we opted to compare the tested iteration of the CBH intervention to the most effective CP iteration. Manipulating the CBH vignette order is a potential avenue for future research.

These results show that the Muslim–Christian intervention is more effective than the CP–CB

intervention, and similarly effective as the CBH. The PFC results shed some light on the psychological mechanism driving these effects, giving some indication that the Muslim–Christian intervention is most influential among participants with high PFC. We will return to this issue in the General Discussion section. Thus, with the effectiveness of the Muslim–Christian intervention established, in Study 3, we sought to further examine the underlying psychological mechanisms behind its effects by separating and testing the different components of the CP task.

Study 3

The goal of Study 3 was to obtain additional evidence of the process underlying the CP intervention. Thus, we separated the Muslim–Christian intervention into three experimental conditions—three Muslim vignettes only, three Christian vignettes only, and six Muslim vignettes only—and compared them and the full Muslim–Christian intervention to a control. We reasoned that if hypocrisy is the psychological process underlying the intervention's effects, then the Muslim–Christian condition should be the only effective condition in reducing hostility towards Muslims as compared to the control. If the driving process is persuasive new positive information about Muslims (Esses et al., 2013), then exposing participants to three, and also to six, vignettes about Muslims engaging in prosocial behaviors should also result in reduced hostility towards Muslims. As for the three-Christian-vignettes condition, we reasoned that presenting Christians with examples of Christian organizations engaging in prosocial behavior might serve as a group affirmation intervention (Sherman et al., 2007), which will lower participants' defenses and thus lower hostility towards Muslims. We predicted that at least one of the experimental conditions involving vignettes of Muslim organizations engaging in prosocial actions would improve attitudes towards Muslims, reflected in (a) reduced dehumanization, (b) reduced support for anti-Muslim policies, and (c) reduced collective blame placed on Muslims, as compared to the control.

In addition to PFC, as in Study 2, we also examined two secondary hypotheses regarding the mechanism driving the effects of the CP intervention. First, previous research shows that increasing perceived outgroup heterogeneity is associated with more favorable intergroup attitudes (e.g., Brauer & Er-Rafiy, 2011). Thus, we examined whether being exposed to information about positive actions conducted by Muslims would challenge participants' perceptions that Muslims are a homogeneous group. Second, past research suggests that exposure to information that receives not much attention by the mainstream media will lead participants to perceive the media as more biased, which is associated with more favorable intergroup attitudes (Moore-Berg et al., 2021). Thus, we examined the role of perceived media bias as a potential mediator.

Methods

Participants. Based on the same power analysis as in Study 1, our final sample is sufficiently powered to detect a minimum effect size of $d = 0.27$ with .95 power. We recruited 1,100 participants to complete our survey. We excluded 29 participants for failing at least one of the two embedded attention check questions, one participant who failed to complete the survey, and 16 participants who identified as Muslim, leaving 1,054 participants in our final sample (47.0% female; M_{age} = 37.0, SD = 11.14; 37.3% Republican, 62.7% Democrat; 71.5% White, 14.1% Black, 5.8% Asian, 5.1% Hispanic, 3.5% other; 55.7% Christian, 36.3% atheist/agnostic, 8% other). Participants were compensated US$1.00.

Measures and procedure. Participants were randomly assigned to one of four experimental conditions: (a) Muslim–Christian ($n = 213$), (b) three Muslim vignettes ($n = 209$), (c) three Christian vignettes ($n = 214$), (d) six Muslim vignettes ($n = 204$), or to the control ($n = 214$). Demographics by condition are reported in the supplemental material.

The Muslim–Christian condition was the same as in Studies 1 and 2. Participants in the three-Muslim-vignettes condition were exposed to three out of six randomized vignettes of prosocial acts performed by Muslim organizations. Participants in the three-Christian-vignettes condition read three of the six randomized prosocial acts performed by Christian organizations. Participants in the six-Muslim-vignettes condition read all six vignettes of prosocial acts performed by Muslim organizations in randomized order. Following the intervention (or control), participants completed the same dehumanization (trait dehumanization: Christians: $\alpha = .91$; Muslims: $\alpha = .94$; correlation between ascent and trait dehumanization measures: $r = .59$, $p < .001$), collective blame, and demographic measures as in Studies 1 and 2, and the same anti-Muslim policy support ($\alpha = .96$) and PFC items ($\alpha = .96$) as in Study 2. We also introduced two potential mediators: media bias and outgroup homogeneity. Study 3 was preregistered before data collection.[4]

Outgroup homogeneity was assessed by asking participants how much they agree or disagree with four items partially adapted from Badea et al. (2012; e.g., "Generally, Muslims are similar to each other"). Responses were given on a 7-point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*). We averaged responses to create an outgroup homogeneity score ($\alpha = .73$).

Media bias was assessed by asking participants to indicate "In general, how biased do you think the American news media is against or for the following groups." Participants reported how much media bias Muslims face (among four filler groups in random order), using unmarked sliders anchored from 0 (*very negatively biased against*) to 100 (*very positively biased for*), with 50 set in the middle (*no bias*). As our measure of bias, we used levels of perceived media bias expressed towards Muslims.

Results

Although a one-way ANOVA revealed no difference in political affiliation across conditions, $F(4, 1046) = 0.31$, $p = .871$, $\eta^2_p < .01$, we controlled for political affiliation in all analyses, as in Studies 1 and 2. For means, standard deviations, and variable intercorrelations, see Table 5.

Table 5. Means, standard deviations, and correlations for all variables: Study 3.

	M	SD	1	2	3	4	5	6	7	8	9
Dehumanization	0.003	1.78	–								
Anti-Muslim policy support	2.59	1.70	.52***	–							
Collective blame	31.53	32.41	.38***	.68***	–						
Preference for consistency	4.94	1.06	.22***	.19***	.12***	–					
Outgroup homogeneity	3.50	1.17	.45***	.51***	.45***	.17***	–				
Media bias	39.20	29.43	.22***	.44***	.40***	.07*	.24***	–			
Political affiliation	−0.55	2.28	.36***	.42***	.33***	.12***	.32***	.31***	–		
Age	37.00	11.14	.05	.02	.02	.07*	.06*	.10***	.03	–	
Gender (F = 0, M = 1)	–	–	−.10**	−.09**	−.07*	.01	−.06	−.16***	−.07*	.10***	–

Note. N = 1,054.
*p < .05. **p < .01. ***p < .001.

**Table 6.** Means and standard deviations for each measure, by condition: Study 3.

Condition	Dehumanization M (SD)	Anti-Muslim policy support M (SD)	Collective blame M (SD)	Outgroup homogeneity M (SD)	Media bias M (SD)
Control (n = 214)	0.19 (1.90)[ac]	2.68 (1.67)[a]	31.92 (31.68)[a]	3.54 (1.17)[ab]	37.80 (28.03)[a]
Muslim–Christian (n = 213)	−0.15 (1.54)[b]	2.47 (1.66)[a]	28.71 (31.29)[a]	3.34 (1.18)[ab]	38.04 (28.31)[a]
Three-Muslim-vignettes (n = 209)	−0.10 (2.01)[ab]	2.69 (1.80)[a]	33.45 (33.18)[a]	3.63 (1.13)[a]	40.51 (30.74)[a]
Three-Christian-vignettes (n = 204)	0.32 (1.82)[c]	2.62 (1.71)[a]	30.96 (32.85)[a]	3.59 (1.25)[a]	41.88 (28.86)[a]
Six-Muslim-vignettes (n = 204)	−0.26 (1.50)[b]	2.48 (1.65)[a]	32.35 (32.94)[a]	3.40 (1.08)[ab]	37.59 (31.00)[a]

Note. Means not sharing the same superscript are significantly different from each other at *p < .05*; superscript comparisons are per variable.

For means and standard deviations by condition, see Table 6. We used Hayes's (2018) PROCESS Model 4 with 5,000 iterations to test whether the intervention (vs. the control) affected anti-Muslim policy support through dehumanization. Additionally, we used Hayes's (2018) PROCESS Model 1 to test for interactions between each experimental condition (vs. the control) and PFC on all outcome variables.

Finally, as additional analyses, we examined the effects of the experimental conditions versus the control on perceived outgroup homogeneity and media bias against Muslims. If direct effects were found, we used Hayes's (2018) PROCESS Model 4 to test whether the interventions versus the control affected our outcome variables through outgroup homogeneity and media bias.

Figure 3. Unstandardized path model assessing the effects of the (A) Muslim–Christian and the (B) six-Muslim-vignettes intervention on anti-Muslim policy support through dehumanization: Study 3.
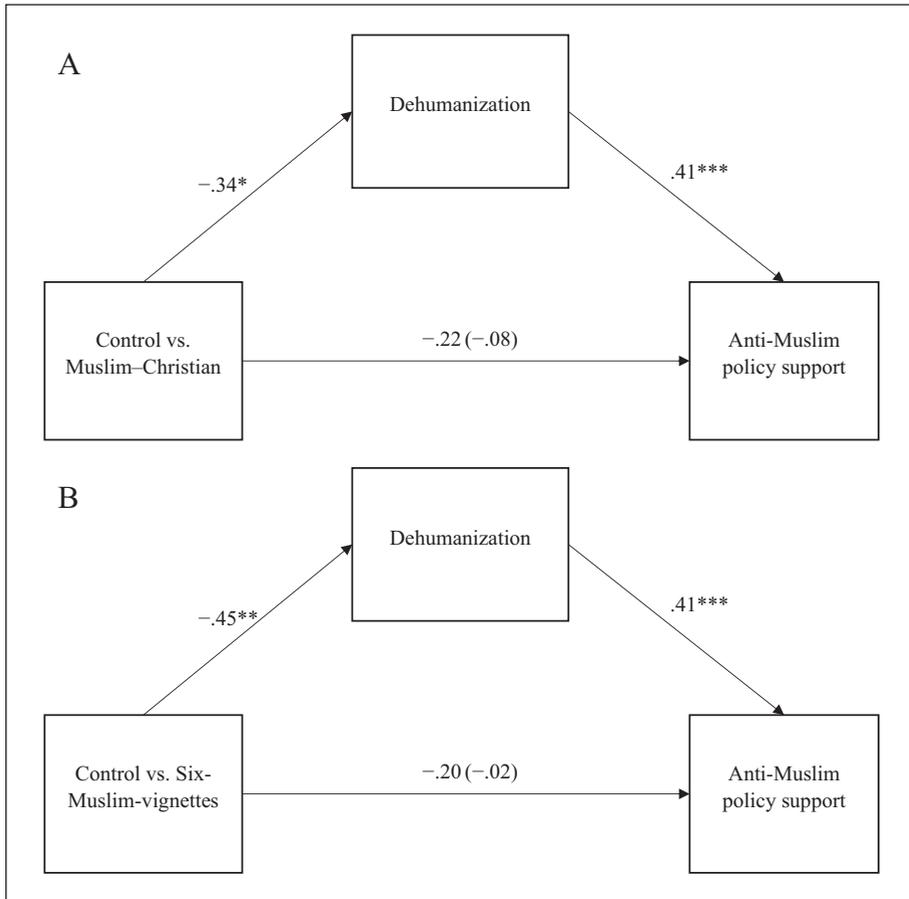
Muslim–Christian condition. Comparing the Muslim–Christian condition to the control, dehumanization of Muslims was significantly lower in the Muslim–Christian condition, $t(1019) = 2.10$, $p = .036$, $d = 0.20$; however, there was no significant difference in anti-Muslim policy support between these conditions, $t(1045) = 1.39$, $p = .164$, $d = 0.12$. Replicating Studies 1 and 2, there was a significant indirect effect of the Muslim–Christian condition (vs. the control) on anti-Muslim policy support through dehumanization (indirect effect $= -.14$, $SE = 0.06$, 95% CI [-0.26, -0.01]; see

Figure 3A). There was no significant difference in collective blame placed on Muslims between these conditions, $t(1045) = 0.96$, $p = .338$, $d = 0.09$. There was no significant Condition x PFC interaction on dehumanization of Muslims ($b = -0.04$, $SE = 0.02$, 95% CI [-0.09, 0.01]), anti-Muslim policy support ($b = -0.02$, $SE = 0.02$, 95% CI [-0.06, 0.02]), or collective blame put on Muslims ($b = -0.15$, $SE = 0.44$, 95% CI [-1.02, 0.73]).

Three-Muslim-vignettes condition. There was no significant difference in dehumanization of Muslims,

$t(1019) = 1.82$, $p = .069$, $d = 0.15$, anti-Muslim policy support, $t(1045) = -0.07$, $p = .944$, $d = -0.01$, or collective blame placed on Muslims, $t(1045) = -0.41$, $p = .684$, $d = -0.04$, between the three-Muslim-vignettes condition and the control. There was also no significant Condition x PFC interaction on dehumanization of Muslims ($b = -0.02$, $SE = 0.02$, 95% CI [-0.07, 0.03]), anti-Muslim policy support ($b = 0.04$, $SE = 0.02$, 95% CI [-0.01, 0.08]), or collective blame assigned to Muslims ($b = 0.77$, $SE = 0.46$, 95% CI [-0.13, 1.66]).

Three-Christian-vignettes condition. There were no significant differences in dehumanization of Muslims, $t(1019) = -0.79$, $p = .430$, $d = -0.07$; anti-Muslim policy support, $t(1045) = 0.36$, $p = .718$, $d = 0.03$; or collective blame assigned to Muslims, $t(1045) = -0.04$, $p = .972$, $d = -0.003$, between the three-Christian-vignettes and the control condition. There were also no significant Condition x PFC interactions on dehumanization of Muslims ($b = -0.02$, $SE = 0.03$, 95% CI [-0.07, 0.03]), anti-Muslim policy support ($b = 0.01$, $SE = 0.02$, 95% CI [-0.04, 0.05]), or collective blame placed on Muslims ($b = 0.51$, $SE = 0.47$, 95% CI [-0.42, 1.44]).

Six-Muslim-vignettes condition. Comparing the six-Muslim-vignettes condition to the control, dehumanization of Muslims was significantly lower, $t(1019) = 2.78$, $p = .006$, $d = 0.27$, in the six-Muslim-vignettes intervention, and there was no significant difference in anti-Muslim policy support between these conditions, $t(1045) = 1.28$, $p = .200$, $d = 0.12$. There was a significant indirect effect of the six-Muslim-vignettes intervention (vs. the control) on anti-Muslim policy support through dehumanization (indirect effect $= -.18$, $SE = 0.06$, 95% CI [-0.31, -0.06]; see Figure 3B). There was no difference in collective blame placed on Muslims between the six-Muslim-vignettes and control conditions, $t(1045) = 0.12$, $p = .904$, $d = 0.01$. Also, there were no significant Condition x PFC interactions on dehumanization of Muslims ($b = -0.03$, $SE = 0.02$, 95% CI [-0.08,

0.02]), anti-Muslim policy support ($b = 0.02$, $SE = 0.02$, 95% CI [-0.03, 0.06]), or collective blame assigned to Muslims ($b = 0.69$, $SE = 0.46$, 95% CI [-0.21, 1.59]).

Additional results. Compared to the control, all experimental conditions showed no significant differences in outgroup homogeneity or media bias (ps $> .087$).

Discussion

Consistent with Studies 1 and 2, Study 3 demonstrated that the Muslim–Christian condition significantly reduced dehumanization of Muslims versus the control, and that it had a significant indirect effect on anti-Muslim policy support through dehumanization. However, there was no difference in collective blame between the Muslim–Christian and control conditions. Furthermore, the six-Muslim-vignettes condition showed similar effects as the Muslim–Christian condition—it significantly reduced dehumanization of Muslims and had an indirect effect (vs. the control) on anti-Muslim policy support through dehumanization. This provides some evidence that to reduce animosity towards Muslims, it is sufficient to present new information about them engaging in prosocial acts. However, the three-Muslim-vignettes condition was not effective at reducing hostility towards Muslims on any outcome measure, which suggests that if the intervention relies solely on presenting new information about Muslims engaging in prosocial behavior, more than three vignettes are required to challenge anti-Muslim views people may hold.

Futhermore, we initally hypothesized that hypocrisy perception would be the driving force behind the CP intervention's effects, as in the CBH intervention (Bruneau, Kteily, & Urbiola, 2020). However, the fact that we found the six-Muslim-vignettes intervention to be effective and that PFC did not moderate any of the effects makes viable the hypothesis that the effects of the Muslim–Christian CP intervention were

driven by exposure to inconsistent information about Muslims engaging in prosocial acts (see Bar-Tal & Hameiri, 2020). We will address these psychologcial processes in the General Discussion section.

Mini Meta-Analyses

To gauge the overall effect of the Muslim–Christian intervention versus the control on dehumanization of Muslims, anti-Muslim policy support, and collective blame placed on Muslims across studies, we ran three mini meta-analyses (Goh et al., 2016). We used fixed effects in which the mean effect size was weighted by sample size. The meta-analyses showed that the effects across studies were relatively small, but nonetheless significant, for dehumanization (mean $d = 0.21$, $Z = 3.67$, $p < .001$, two-tailed), anti-Muslim policy support (mean $d = 0.15$, $Z = 2.53$, $p = .011$, two-tailed), and collective blame (mean $d = 0.21$, $Z = 3.62$, $p < .001$, two-tailed). Thus, across studies, the Muslim–Christian iteration of the CP task reduced dehumanization of Muslims, anti-Muslim policy support, and collective blame assigned to Muslims, compared to the control.

General Discussion

Across three preregistered studies that included various iterations of the CP intervention and compared them to an alternative intervention, we found that exposure to our brief intervention consisting of three vignettes of prosocial behaviors enacted by Muslims followed by three similar vignettes of prosocial behaviors enacted by Christians consistently reduced dehumanization of Muslims. Moreover, mini meta-analyses indicated that the Muslim–Christian CP intervention significantly reduced levels of anti-Muslim policy support and collective blame placed on Muslims when collapsing across all studies, albeit effects were not consistently significant when examining results from each study separately. Finally, across all studies, the CP intervention indirectly reduced levels of anti-Muslim policy support through dehumanization. Consistent with previous

assessments of hypocrisy-based interventions (Bruneau et al., 2018; Bruneau, Kteily, & Urbiola, 2020; Shulman et al., 2020), the current work highlights the effectiveness of such interventions in reducing outgroup hostility. Unlike previous interventions, the CP approach utilizes positive, rather than negative or aversive, information to induce hypocrisy perception.

An intervention that consistently reduces dehumanization of Muslims is of the utmost value, as dehumanization predicts the most detrimental intergroup outcomes (e.g., support for hostile policies) above and beyond prejudice (e.g., Kteily & Bruneau, 2017b; Kteily et al., 2015, 2016; Wilde et al., 2014). The CP intervention reduced dehumanization, and indirectly reduced anti-Muslim policy support through dehumanization in all three studies, further highlighting the value of an intervention that can short-circuit dehumanization. Although indirect analyses do not confirm the causal relationship between dehumanization and anti-Muslim policy support, previous literature has argued that dehumanization predicts a variety of hostile outcomes (Bruneau et al., 2018; Kteily et al., 2015).

The CP intervention follows in the footsteps of previous studies that highlight hypocrisy perception as a way to change behavior, adding an important layer to the literature. Recently, the CBH intervention showed that highlighting the hypocrisy involved in collectively blaming entire groups for the heinous acts of individual members of outgroups but not of ingroups can promote better intergroup relations (Bruneau et al., 2018; Bruneau, Kteily, & Urbiola, 2020). Similarly, the CP intervention also highlights hypocrisy perception to improve intergroup relations. However, it highlights the hypocrisy involved in attributing prosocial acts of ingroup individuals to the entire ingroup but not doing the same for outgroup members, suggesting that nonaversive stimuli can be at least equally, if not more, effective than aversive stimuli in a hypocrisy-based paradigm.

Building on the existing literature, the CP intervention is, to the best of our knowledge, the first hypocrisy-based, nonaversive intervention to successfully reduce intergroup hostility. Beyond

the contributions of these findings to the theoretical understanding of hypocrisy-based interventions, our work has valuable practical implications as well. In lab studies, participants are usually given an incentive (e.g., compensation) for participation; however, to reach a broader audience, willingness to engage with interventions needs to be voluntary (Bar-Tal & Hameiri, 2020; Hameiri et al., 2016). When considering work on hypocrisy perception interventions that rely on negative stimuli, some individuals might be reluctant to voluntarily engage with such stimuli when no external incentive is offered (e.g., Hay et al., 2015). Another practical problem is that, in many cases, those who can administer these interventions (e.g., advertising agencies, NGOs) do not want to be associated with negative stimuli (e.g., terrorist attacks) out of fear of backlash (cf. Paluck & Cialdini, 2014). Thus, we argue that the CP approach is promising as a potential real-world intervention because it consists of nonaversive stimuli, which will likely make participants and organizations less resistant to it.

Despite these meaningful theoretical and practical implications, the current research is not without limitations. First, while the CP intervention significantly reduced levels of anti-Muslim policy support and collective blame put on Muslims when collapsing across all studies, effects were not consistently significant when examining each study independently. One explanation for these inconsistent results may be the lack of attacks by Muslim extremists during study deployment. This might have yielded lower levels of anti-Muslim attitudes across conditions (including the control), which provided less room for the intervention to significantly decrease these attitudes. Indeed, the context in which an intervention is administered plays a pivotal role in its effectiveness (Bar-Tal & Hameiri, 2020; Ditlmann et al., 2017). Specifically, we argue that an intervention designed to shift attitudes requires participants to hold relevant preexisting attitudes towards the targeted group. If participants do not hold salient attitudes towards the topic of the intervention, or are preoccupied with other current issues (e.g., COVID-19), the information

provided will be less relevant to them and perhaps less influential.

Although we established the CP intervention's successfulness, more research is needed to understand whether hypocrisy is the underlying psychological process behind the effects. To examine this, we tested whether PFC moderated our results (see Bruneau, Kteily, & Urbiola, 2020). In Study 2, we found that among high-PFC participants, the Muslim–Christian intervention (our most effective CP intervention), as compared to the control, was more effective at reducing dehumanization. However, when comparing high- and low-PFC participants within the Muslim–Christian condition, we found no differences. Also, we failed to replicate this finding in Study 3. This provides only tentative support for the hypothesis that the CP intervention effects are driven by hypocrisy perception.

Furthermore, we originally theorized that the Christian vignettes would be essential to the CP intervention. However, in Study 3, we found that the six-Muslim-vignettes intervention also significantly reduced levels of dehumanization as compared to the control. This finding might indicate that the three vignettes with Christian protagonists in the Muslim–Christian intervention are redundant, and thus that hypocrisy is not driving its effects. Further, all other examined mechanisms (i.e., feeling shamed, cognitive unfreezing, outgroup homogeneity, and media bias) did not shed light on the psychological process behind the CP intervention's effects. Thus, although we are confident that the CP intervention effectively reduces anti-Muslim hostility, we cannot conclusively establish the underlying psychological process.

One alternative mechanism to hypocrisy is that participants might have changed their attitudes towards Muslims due to exposure to information that was inconsistent with their beliefs and attitudes. The general public is constantly exposed to information that portrays Muslims as violent and aggressive (Kearns et al., 2019; Stone, 2017). For example, an analysis of news coverage by three U.S. networks (CBS, FOX, and NBC) revealed that 75% of news reports about Muslims

focused on terrorist activities and conflict (Stone, 2017). Thus, the CP intervention's effect could be due to hypocrisy perception, persuasive inconsistent information, or both. In fact, hypocrisy perception and persuasive inconsistent information might differentially be the driving psychological processes for different people (see Bar-Tal & Hameiri, 2020; Halperin & Schori-Eyal, 2020). We included variables (i.e., cognitive unfreezing, outgroup homogeneity, and perceived media bias) to explore this hypothesis, but our intervention did not have a significant effect on them. Similarly, the different types of hypocrisy-based interventions (i.e., the CP and CBH approaches) might be differentially effective for different people (e.g., for people who are more or less willing to engage with aversive stimuli; Pliskin et al., 2018) and under certain conditions. Future work should further examine this.

In summary, previous hypocrisy-based interventions have reduced hostility towards outgroups but relied on aversive stimuli. The current work introduces the first hypocrisy-based intervention to reduce hostility towards an outgroup using nonaversive stimuli. The current research adds an important layer to the literature on hypocrisy-based interventions to reduce intergroup conflict by suggesting that nonaversive stimuli are similarly effective in reducing outgroup hostility as aversive stimuli are. A nonaversive intervention is useful because it is less likely to receive backlash, and thus it is more easily scalable. We hope that the current work will encourage researchers and practitioners to take this approach that utilizes positive stimuli to reduce intergroup hostility, especially towards minority groups.

Funding

ORCID iDs

Roman A. Gallardo ⓘD https://orcid.org/0000-0002 -3601-1225

Boaz Hameiri ⓘD https://orcid.org/0000-0002-0241 -9839

Samantha L. Moore-Berg ⓘD https://orcid.org/0000 -0003-2972-2288

Supplemental material

Supplemental material for this article is available online.

Notes

1. See https://osf.io/8utp5; for the complete survey and data, see https://osf.io/qj7y6/?view_only=8 077d611fe664b448afd8d369375c2c5.
2. Analyzing the Ascent of (Hu)Man and trait dehumanization measures separately yielded similar patterns of results.
3. See https://osf.io/e672r; for the complete survey and data, see https://osf.io/srmpb/?view_only= 3728e962c89f41cdbb69ce36db12ba2b.
4. See https://osf.io/hqby4; for the complete survey and data, see https://osf.io/hgrdp/?view_onl y=1f9069b83d9e4b23829d22a759e748e5.

References

Abdelgadir, A., & Fouka, V. (2019, June 3). How will Austria's new headscarf ban affect Muslims? *The Washington Post*. https://www.washingtonpost. com/politics/2019/06/03/how-will-austrias- new-headscarf-ban-affect-muslims/

Adelman, L., Yogeeswaran, K., & Lickel, B. (2019). They're all the same, sometimes: Prejudicial attitudes toward Muslims influence motivated judgments of entitativity and collective responsibility for an individual's actions. *Journal of Experimental Social Psychology*, *80*, 31–38. https://doi. org/10.1016/j.jesp.2018.10.002

Badea, C., Brauer, M., & Rubin, M. (2012). The effects of winning and losing on perceived group variability. *Journal of Experimental Social Psychology*, *48*, 1094– 1099. https://doi.org/10.1016/j.jesp.2012.03.006

Bar-Tal, D., & Hameiri, B. (2020). Interventions to change well-anchored attitudes in the context of intergroup conflict. *Social and Personality Psychology Compass*, *14*, Article e12534. https://doi. org/10.1111/spc3.12534

Bastian, B., Denson, T. F., & Haslam, N. (2013). The roles of dehumanization and moral outrage in retributive justice. *PLoS ONE*, *8*, Article e61842. https://doi.org/10.1371/journal.pone.0061842

Brauer, M., & Er-Rafiy, A. (2011). Increasing perceived variability reduces prejudice and discrimination. *Journal of Experimental Social Psychology*, *47*, 871–881. https://doi.org/10.1016/j.jesp.2011.03.003

Brehm, J. W. (1966). *A theory of psychological reactance.* Academic Press.

Bruneau, E., Hameiri, B., Moore-Berg, S. L., & Kteily, N. (2020). Intergroup contact reduces dehumanization and meta-dehumanization: Cross-sectional, longitudinal and quasi-experimental evidence from 16 samples in 5 countries. *Personality and Social Psychology Bulletin.* Advance online publication. https://doi.org/10.1177/0146167220949004

Bruneau, E., & Kteily, N. (2017). The enemy as animal: Symmetric dehumanization during asymmetric warfare. *PLoS ONE*, *12*, Article 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*, 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*, 45–54. https://doi.org/10.1038/s41562-019-0747-7

Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology*, *69*, 318–328. https://doi.org/10.1037/0022-3514.69.2.318

Diamant, J.. (2017) American Muslims are concerned – but also satisfied with their lives. *Pew Research Center.* https://www.pewresearch.org/fact-tank/2017/07/26/american-muslims-are-concerned-but-also-satisfied-with-their-lives/

Ditlmann, R. K., Samii, C., & Zeitzoff, T. (2017). Addressing violent intergroup conflict from the bottom up? *Social Issues and Policy Review*, *11*, 38–77. https://doi.org/10.1111/sipr.12027

Esses, V. M., Medianu, S., & Lawson, A. S. (2013). Uncertainty, threat, and the role of the media in promoting the dehumanization of immigrants and refugees. *Journal of Social Issues*, *69*, 518–536. https://doi.org/10.1111/josi.12027

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

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*, 535–549. https://doi.org/10.1111/spc3.12267

Halperin, E., & Schori-Eyal, N. (2020). Towards a new framework of personalized psychological interventions to improve intergroup relations and promote peace. *Social and Personality Psychology Compass*, *14*, 255–270. https://doi.org/10.1111/spc3.12527

Hameiri, B., Nabet, E., Bar-Tal, D., & Halperin, E. (2018). Paradoxical thinking as a conflict-resolution intervention: Comparison to alternative interventions and examination of psychological mechanisms. *Personality and Social Psychology Bulletin*, *44*, 122–139. https://doi.org/10.1177/0146167217736048

Hameiri, B., Porat, R., Bar-Tal, D., & Halperin, E. (2016). Moderating attitudes in times of violence through paradoxical thinking intervention. *Proceedings of the National Academy of Sciences of the USA*, *113*, 12105–12110. https://doi.org/10.1073/pnas.1606182113

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, *10*, 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology*, *65*, 399–423. https://doi.org/10.1146/annurev-psych-010213-115045

Haslam, N., Loughnan, S., Kashima, Y., & Bain, P. (2008). Attributing and denying humanness to others. *European Review of Social Psychology*, *19*, 55–85. https://doi.org/10.1080/10463280801981645

Hay, A. C., Sheppes, G., Gross, J. J., & Gruber, J. (2015). Choosing how to feel: Emotion regulation choice in bipolar disorder. *Emotion*, *15*, 139–145. https://doi.org/10.1037/emo0000024

Hayes, A. F. (2018). *An introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Press.

Kearns, E. M., Betus, A. E., & Lemieux, A. F. (2019). Why do some terrorist attacks receive more media attention than others? *Justice Quarterly*, *36*, 985–1022. https://doi.org/10.1080/07418825.2018.1524507

Kteily, N., & Bruneau, E. (2017a). Backlash: The politics and real-world consequences of minority group dehumanization. *Personality and Social*

Psychology Bulletin, *43*, 87–104. https://doi.org/10.1177/0146167216675334

Kteily, N., & Bruneau, E. (2017b). Darker demons of our nature: The need to (re)focus attention on blatant forms of dehumanization. *Current Directions in Psychological Science*, *26*, 487–494. https://doi.org/10.1177/0963721417708230

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*, 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*, 343–370. https://doi.org/10.1037/pspa0000044

Lee, S. A., Gibbons, J. A., Thompson, J. M., & Timani, H. S. (2009). The Islamophobia Scale: Instrument development and initial validation. *The International Journal for the Psychology of Religion*, *19*, 92–105. https://doi.org/10.1080/10508610802711137

Leippe, M. R., & Eisenstadt, D. (1994). Generalization of dissonance reduction: Decreasing prejudice through induced compliance. *Journal of Personality and Social Psychology*, *67*, 395–413. https://doi.org/10.1037/0022-3514.67.3.395

Lipka, M. (2017). *Muslims and Islam: Key findings in the U.S. and around the world*. Pew Research Center. https://www.pewresearch.org/fact-tank/2017/08/09/muslims-and-islam-key-findings-in-the-u-s-and-around-the-world/

Moore-Berg, S. L., Hameiri, B., Falk, E., & Bruneau, E. (2021). *Reducing Islamophobia: An assessment of psychological mechanisms that underlie successful anti-Islamophobia videos*. Manuscript submitted for publication.

Niayesh, V. (2019, September 26). Trump's travel ban really was a Muslim ban, data suggests. *The Washington Post*. https://www.washingtonpost.com/politics/2019/09/26/trumps-muslim-ban-really-was-muslim-ban-thats-what-data-suggest/

Ogan, C., Willnat, L., Pennington, R., & Bashir, M. (2014). The rise of anti-Muslim prejudice: Media and Islamophobia in Europe and the United States. *International Communication Gazette*, *76*, 27–46. https://doi.org/10.1177/1748048513504048

Paluck, E. L., & Cialdini, R. B. (2014). Field research methods. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 81–98). Cambridge University Press.

Paolini, S., & McIntyre, K. (2019). Bad is stronger than good for stigmatized, but not admired outgroups: Meta-analytical tests of intergroup valence asymmetry in individual-to-group generalization experiments. *Personality and Social Psychology Review*, *23*, 3–47. https://doi.org/10.1177/1088868317753504

Pliskin, R., Halperin, E., Bar-Tal, D., & Sheppes, G. (2018). When ideology meets conflict-related content: Influences on emotion generation and regulation. *Emotion*, *18*, 159–170. https://doi.org/10.1037/emo0000317

Samuel, S. (2019, December 12). *India just redefined its citizenship criteria to exclude Muslims*. Vox. https://www.vox.com/future-perfect/2019/12/12/21010975/india-muslim-citizenship-bill-national-register

Schildkraut, D. J. (2009). The dynamics of public opinion on ethnic profiling after 9/11: Results from a survey experiment. *American Behavioral Scientist*, *53*, 61–79. https://doi.org/10.1177/0002764209338786

Sheppes, G., & Levin, Z. (2013). Emotion regulation choice: Selecting between cognitive regulation strategies to control emotion. *Frontiers in Human Neuroscience*, *7*, Article 179. https://doi.org/10.3389/fnhum.2013.00179

Sherman, D. K., Kinias, Z., Major, B., Kim, H. S., & Prenovost, M. (2007). The group as a resource: Reducing biased attributions for group success and failure via group affirmation. *Personality and Social Psychology Bulletin*, *33*, 1100–1112. https://doi.org/10.1177/0146167207303027

Shulman, D., Halperin, E., Kessler, T., Schori-Eyal, N., & Reifen Tagar, M. (2020). Exposure to analogous harmdoing increases acknowledgment of ingroup transgressions in intergroup conflicts. *Personality and Social Psychology Bulletin*. Advance online publication. https://doi.org/10.1177/0146167220908727

Simmons, K. (2019, October 4). *Inside Chinese camps thought to be detaining a million Muslim Uighurs*. NBC News. https://www.nbcnews.com/news/world/inside-chinese-camps-thought-detain-million-muslim-uighurs-n1062321

Stone, M. (2017). *Snake and stranger: Media coverage of Muslims and refugee policy*. Harvard Kennedy School Shorenstein Center on Media, Politics and Pub-

lic Policy. https://shorensteincenter.org/media-coverage-muslims-refugee-policy/

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Brooks/Cole.

Vaes, J., & Paladino, M. P. (2010). The uniquely human content of stereotypes. *Group Processes & Intergroup Relations*, *13*, 23–39. https://doi.org/10.1177/1368430209347331

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. *Advances in experimental social psychology*, *29*, 141–208. https://doi.org/10.1016/S0065-2601(08)60017-9

Wilde, V. K., Martin, K. D., & Goff, P. A. (2014). Dehumanization as a distinct form of prejudice. *TPM: Testing, Psychometrics, Methodology in Applied Psychology*, *21*, 301–307. https://doi.org/10.4473/TPM21.3.5

Wohl, M. J., Cohen-Chen, S., Halperin, E., Caouette, J., Hayes, N., & Hornsey, M. J. (2015). Belief in the malleability of groups strengthens the tenuous link between a collective apology and intergroup forgiveness. *Personality and Social Psychology Bulletin*, *41*, 714–725. https://doi.org/10.1177/0146167215576721