Personality and Social Psychology Bulletin
2018, Vol. 44(3) 430–448

DOI: 10.1177/0146167217744197
journals.sagepub.com/home/pspb
SAGE

Interventions Highlighting Hypocrisy Reduce Collective Blame of Muslims for Individual Acts of Violence and Assuage Anti-Muslim Hostility

Emile Bruneau[1,2], Nour Kteily[3], and Emily Falk[1]

Abstract

Collectively blaming groups for the actions of individuals can license vicarious retribution. Acts of terrorism by Muslim extremists against innocents, and the spikes in anti-Muslim hate crimes against innocent Muslims that follow, suggest that reciprocal bouts of collective blame can spark cycles of violence. How can this cycle be short-circuited? After establishing a link between collective blame of Muslims and anti-Muslim attitudes and behavior, we used an "interventions tournament" to identify a successful intervention (among many that failed). The "winning" intervention reduced collective blame of Muslims by highlighting hypocrisy in the ways individuals collectively blame Muslims—but not other groups (White Americans, Christians)—for individual group members' actions. After replicating the effect in an independent sample, we demonstrate that a novel interactive activity that isolates the psychological mechanism amplifies the effectiveness of the collective blame hypocrisy intervention and results in downstream reductions in anti-Muslim attitudes and anti-Muslim behavior.

Keywords

collective blame, collective responsibility, vicarious retribution, prejudice, Islamophobia, intervention

Received February 9, 2017; revision accepted October 29, 2017

> . . . until [Muslims] recognize and destroy their growing jihadist cancer they must be held responsible.
>
> —Tweet from Rupert Murdoch, Chairman of Fox News, January 9, 2015

On April 15, 2013, Dzhokhar and Tamerlan Tsarnaev detonated two bombs at the finish line of the Boston Marathon. The Tsarnaev brothers cited the treatment of Muslims by the U.S. military overseas as motivation for the attacks (Wilson, Miller, & Horwitz, 2013). However, all the three people who were killed and over 200 injured from the blast were American civilians not directly involved in U.S. military interventions overseas. Since the Boston Marathon bombings, a handful of other attacks by Muslims on Americans have been launched, also targeting civilians. On the other side, each terror attack committed by Muslims has been followed by rhetoric (like the tweet above) explicitly blaming all Muslims for any attack, and a multifold increase in anti-Muslim hate crimes committed by non-Muslim Americans against Muslim Americans (Ingraham, 2015). The pattern of attack and reprisal against innocents from each group reveals a particular psychological calculus of intergroup conflict: People have a tendency to hold groups collectively responsible for the actions of individual group members, which justifies "vicarious retribution" against any group member to exact revenge (Lickel, Miller, Stenstrom, Denson, & Schmader, 2006).

Previous research has demonstrated the relevance of collective blame in organizational settings, showing that companies, schools, and even loosely affiliated groups of people are held responsible for the harmful actions of individual group members (Chiu, Morris, Hong, & Menon, 2000; Manchi Chao, Zhang, & Chiu, 2008; Menon, Morris, Chiu, & Hong, 1999; Singh et al., 2012; Zemba, Young, & Morris, 2006). Much of this research has focused on identifying differences in the tendency to engage in collective blame in Eastern versus Western cultures (e.g., Chiu et al., 2000; Manchi Chao et al., 2008), on discerning the psychological precursors of collective blame, including perceived outgroup homogeneity and entitativity (e.g., Denson, Lickel, Curtis, Stenstrom, &

[1]Annenberg School for Communication, University of Pennsylvania, Philadelphia, USA
[2]Beyond Conflict Innovation Lab, Boston, MA, USA
[3]Kellogg School of Management, Northwestern University, Chicago, IL, USA

Corresponding Author:
Emile Bruneau, Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA 19104, USA.
Email: ebruneau@asc.upenn.edu

Ames, 2006; Lickel et al., 2006), and on examining the consequences of collective blame—namely, exacting revenge on people from an offending group who were uninvolved of the offense (i.e., "vicarious retribution"; Lickel et al., 2006; Stenstrom, Lickel, Denson, & Miller, 2008). However, this previous work has given less consideration to collective blame in intergroup contexts, despite its clear potential to contribute importantly to intergroup hostility.

In a first study, we sought to examine the proclivity for and correlates of collective blame within the realm of intergroup relations. Specifically, we assessed the degree to which individuals collectively blamed Muslims for acts of mass violence committed by small groups of Muslims and tested the extent to which variability in collective blame predicted anti-Muslim attitudes and beliefs (prejudice and dehumanization), support for antagonistic policies toward Muslims, and hostile behavior toward Muslims. Having identified the importance of collective blame, in the remaining studies we focused our efforts on understanding how it might be reduced. Our initial approach was to gather a range of potential intervention strategies and test them against each other in an "intervention tournament."

Intervention Tournament

Consistent with the tenets of "action research" (Lewin, 1946), we sought to determine not only "what works" to reduce collective blame of Muslims, but also "what works *best*." We therefore focused our efforts not on a single intervention, but instead evaluated the efficacy of a number of interventions simultaneously in an "intervention tournament" (for similar approaches, see J-PAL Policy Bulletin, 2012; Lai et al., 2014; Lai et al., 2016). The interventions in the current study were videos created by nonscientists that tapped into at least one identifiable psychological process. Each of the videos was chosen because it spanned different styles of delivery (didactic, narrative, satire) and had, in our estimation, theoretically distinct psychological content. This allowed us to map specific psychological theories onto each of the videos (albeit more tightly in some cases than others). We hypothesized that each of the videos could reduce collective blame, either directly or indirectly, and reasoned that once we had identified a successful intervention (or set of interventions), we could then more deliberately explore the mechanism underlying it (or them).

Overview of Research

In Study 1, we sought to establish the importance of collective blame to an important contemporary intergroup conflict by showing that the degree to which Americans collectively blame Muslims for acts of terrorism is associated with anti-Muslim attitudes, policy support, and behavior.

In Study 2a and Supplemental Study 1, we examined the efficacy of interventions aimed at reducing anti-Muslim attitudes and behaviors. In a "forecasting tournament" (Supplemental Study 1), we had one group of participants predict the effect of each intervention on collective blame. In an "intervention tournament" (Study 2a), we used a second group of participants to determine the actual effect of each video on collective blame (then compared the actual effects with the forecasted effects, in supplemental analyses). After identifying a "winning" approach in the intervention tournament, we replicated the effects in an independent sample (Study 2b).

Because we had less control over the specific content of the intervention videos, we could only speculate about the processes by which any given video may have been effective. One of the videos that emerged as most effective in the intervention tournament highlighted, among other things, the hypocrisy in collectively blaming Muslims but not other groups (e.g., Christians) for the actions of a few group members. This approach resembles the classic cognitive dissonance hypocrisy paradigm (Aronson, Fried, & Stone, 1991), in which hypocrisy is induced by the combination of two factors: (a) Having individuals advocate for a position, and then (b) making them aware of failure to act in accord with that position. One way to resolve the resulting cognitive dissonance (Festinger, 1962) is to change one's behavior to act in accord with the advocated position. We theorized that the video in the intervention tournament worked because highlighting hypocrisy in collective blame induced dissonance, which could be resolved by reducing collective blame of Muslims. To verify this, we developed in Studies 3a and 3b a novel interactive activity that was specifically designed to target the proposed psychological mechanism through a Socratic exercise. We tested the effectiveness of this activity relative to two other theoretically distinct and intuitively promising activities (Study 3a) and replicated the effects of the activity in a second study (Study 3b).

Thus, this research took a full-cycle approach by (a) identifying collective blame as an important psychological process associated with anti-Muslim attitudes and behavior using correlational data, (b) seeking a successful intervention to causally mitigate collective blame, and then (c) testing the mechanism for the successful intervention by developing a new targeted intervention focused on the proposed key "ingredient."

Study 1

In Study 1, we sought to determine the prevalence and correlates of collective blame in an intergroup context by examining non-Muslim Americans' collective blame of Muslims for terror attacks. In a cross-sectional study, non-Muslim American participants reported how much they blamed Muslims for the terror attacks in Paris in November 2015 that killed 130 people and injured hundreds more. We then examined the association between collective blame and hostile attitudes, beliefs, and behavior toward Muslims.

We reasoned that holding all Muslims responsible for a terror attack would be associated with endorsement of anti-Muslim attitudes and beliefs. We focused on anti-Muslim prejudice and blatant dehumanization of Muslims. We also reasoned that collective blame of Muslims would be associated with more downstream support for anti-Muslim policies and anti-Muslim behavior (e.g., willingness to sign anti-Muslim petitions).

Method

For this study and all following studies, we determined our sample size a priori, did not exclude any data from analyses, and included in our analyses all manipulations and measures, except where explicitly specified.

Participants. For this correlational study with a range of variables, we aimed to collect a relatively large sample of 200 participants from Amazon's Mechanical Turk. Of the 200 non-Muslim Americans who completed the survey, seven failed an attention check question embedded in the survey, leaving 193 participants (104 female, M_{age} = 35.69, SD = 11.38). The final sample was 49.7% Christian, 2.6% Jewish, 2.1% Buddhist, 0.5% Hindu, 33.2% atheist/agnostic, and 3.6% "Other," with 8.3% of participants providing no response. Ethnically, the sample was 80.3% White, 6.2% Asian, 4.7% Hispanic, 6.7% Black, 0.5% Native American, and 1.6% "Other."

Procedure and stimuli. Participants completed a survey that assessed the key measure of collective blame, measures of anti-Muslim attitudes and beliefs (i.e., blatant dehumanization and prejudice), support for anti-Muslim policies, and two anti-Muslim behavioral measures.

Collective Blame was assessed by presenting participants with a brief description of the Paris terror attacks ("In November 2015, terror attacks in Paris killed 130 people and wounded hundreds. How responsible do you think Muslims are for the attacks in Paris?") and then having them report how responsible they felt "Muslims in general" and "French Muslims" were for the attacks using unmarked sliders anchored at 0 (*not at all*) and 100 (*very much*). Note that this study occurred in the weeks after the attacks, when media coverage of them was ubiquitous.

Blatant Dehumanization was assessed by asking participants how well a series of eight dehumanizing traits/trait pairs (e.g., "savage," "unsophisticated," "barbaric, cold-hearted"; Bastian & Haslam, 2010; Kteily, Bruneau, Waytz, & Cotterill, 2015) applied to Muslims (α = .95).

Prejudice was assessed by standardizing and combining responses to two different prejudice measures: feeling thermometers (Haddock, Zanna, & Esses, 1993) and a multi-item measure of Islamoprejudice (Imhoff & Recker, 2012). With the feeling thermometers, participants reported their affective prejudice toward a range of groups—including Americans and Muslims—on unmarked sliders anchored at 0 (*very cold/unfavorable*) and 100 (*very warm/favorable*). We took as our measure of anti-Muslim prejudice the difference in response to Americans versus Muslims. Islamoprejudice was assessed using a 15-item scale developed by Imhoff and Recker (2012). The 15-item scale includes nine items (e.g., "Islam is an archaic religion, unable to adjust to the present") that reflect Islamoprejudice and have previously been associated with anti-Muslim intentions. These are differentiated from six further items intended to reflect a secular critique of Islam that have been shown to be unassociated with anti-Muslim intentions (Imhoff & Recker, 2012) and which we therefore excluded from analyses. Responses were made on unmarked sliders anchored at 0 (*completely disagree*) and 100 (*completely agree*). A factor analysis revealed that one of the Islamoprejudice items loaded more strongly with the Secular Concern items; we therefore created an Islamoprejudice scale with the remaining eight items (α = .83). To create a single measure of prejudice, feeling thermometer and Islamoprejudice were each z scored and then combined (r = .66, $p < .001$).

Anti-Muslim Policy Support was assessed by asking participants to indicate their support for nine policies targeting Muslims taken from Kteily and Bruneau (2017); sample items included "We should ban the wearing of the Islamic veil" and "We should ban the opening of any new Mosques in this country." Responses were made on 7-point Likert-type scales anchored at 1 (*strongly disagree*) and 7 (*strongly agree*). Several of the policies were adapted directly from campaign statements of (then Presidential candidate) Donald Trump (α = .95).

Punitive Counterterrorism (behavior) was assessed using a measure adapted from Kteily et al. (2015) in which participants were asked to allocate funds in any proportion to two antiterrorism programs: "Build libraries and schools in Muslim-majority communities throughout the United States" (i.e., "preventative counter-terrorism") and "Increase surveillance and policing capabilities in Muslim-majority communities throughout the United States" (i.e., "punitive counter-terrorism"). We took the proportion of funds distributed to punitive counterterrorism as the outcome measure.

Signing Anti-Muslim Petitions (behavior) was assessed by giving participants the opportunity to sign six petitions urging congressional members to implement anti-Muslim policies (Kteily & Bruneau, 2017; Kteily et al., 2015). Three of the petitions focused on Muslim refugees (sample item: "Urge congressional members to deny entry to any Muslim refugees who seek to come to the United States"); the other three petitions were associated with Muslims more generally (sample item: "Urge congressional members to introduce surveillance programs targeting Mosques in the United States"). Responses were coded as +1 for signatures in support of anti-Muslim petitions, –1 for signatures to the counterpetitions, and 0 for no signature (α = .93).

Table 1. Descriptive Statistics and Variable Intercorrelations in Study 1.

	1	2	3	4	5	6
1. Collective blame	—					
2. Blatant dehumanization	.71***	—				
3. Prejudice	.73**	.78***	—			
4. Punitive counterterror	.72***	.72***	.84***	—		
5. Anti-Muslim policies	.68***	.63***	.81***	.85***	—	
6. Anti-Muslim petitions	.56***	.54***	.60***	.68***	.58***	—
M	35.65[a]	3.66[b]	0.00[c]	38.12[a]	2.86[b]	−0.08[d]
SD	34.84	1.52	1.00	37.63	1.77	0.53

[a]Scale: 0 to 100.
[b]Scale: 1 to 7.
[c]Scale: *z* score.
[d]Scale: −1, 0, +1.
*$p < .05$. **$p < .01$. ***$p < .001$.

Results

The primary goals of Study 1 were to examine the degree to which Muslims are held collectively responsible for an individual terror attack, and the correlates of collective blame. The mean response on the 100-point scale was 35.65 (*SD* = 34.84) for the terror attacks in Paris in 2015. Confirming our predictions, we found that the tendency to hold Muslims collectively responsible was significantly correlated with each of the other measures (Table 1). Thus, those who collectively blamed Muslims were also more likely to feel prejudiced against Muslims, dehumanize them, support anti-Muslim policies, donate to surveillance over education in Muslim communities to prevent terrorism, and sign petitions targeting Muslims. We observed a similar pattern of results when we examined participants' tendency to collectively blame French Muslims (rather than "Muslims in general").

Study 2a

Study 1 confirmed that the degree to which Americans collectively blame Muslims for acts of terrorism is associated with anti-Muslim attitudes and behavior. This study thus verifies that collective blame is relevant in the intergroup domain and holds the potential to importantly influence downstream policy attitudes and behavior. In Study 2a, we shifted our attention to determining how collective blame can be reduced by determining the efficacy of eight different video interventions.

We conducted two separate studies using the videos. In the main experiment (Study 2a), participants reported collective blame, other anti-Muslim attitudes (blatant dehumanization, prejudice, Islamoprejudice), and outcomes associated with vicarious retribution as part of an "intervention tournament." As a secondary aim, we had a separate sample of participants (Supplementary Study 1) report their lay predictions of each video's effectiveness using a "forecasting tournament," in which participants *predicted* how much other non-Muslim Americans would collectively blame Muslims for individual acts of terrorism after watching each video.

The forecasting tournament was included to control for potential experimenter bias (i.e., consciously or unconsciously pitting a "favored" video against a set of weak, low-quality, or inferior alternatives) but mostly to extend insights about individuals' (lack of) ability to forecast the success of interventions (e.g., Cialdini, 2003; Noar, 2006) to the realm of intergroup relations. It is possible that the prevalence of intergroup hostility provides individuals with good lay intuitions about its underlying causes (and how to reduce hostility). However, it is also possible that individuals are poor at identifying effective approaches for improving intergroup relations because these interventions may be operating in ways that are not easily accessible to lay perceivers (e.g., through unconscious processes). Addressing this question is theoretically important because it extends the generalizability of prior demonstrations of poor forecasting to intergroup processes, and practically important because finding that individuals are poor judges of what works when it comes to reducing hostility would have major implications for nongovernmental organizations (NGOs) and others who develop interventions based on intuition or focus groups, but do not test for their effects.

For the tournaments, we chose eight 2- to 4-min videos that included psychological elements that could directly or indirectly reduce collective blame. In one video (Video 1), a Muslim woman revealed the hypocrisy of blaming Muslims as a group for Muslim extremists, but not blaming Christians as a group for Christian extremists. This approach is loosely aligned with a hypocrisy paradigm that has been successfully employed to generate cognitive dissonance and induce prosocial behaviors (Dickerson, Thibodeau, Aronson, & Miller, 1992; Fried & Aronson, 1995; Stone & Fernandez, 2008). We therefore thought it plausible that the induction or revelation of the hypocrisy of collectively blaming some groups but not

others could reduce collective blame of Muslims. Five videos (Videos 1, 2, 3, 5, and 6) targeted collective blame indirectly by challenging the homogeneity and/or entitativity of Muslims. As group homogeneity and entitativity serve as direct precursors to collective blame (Denson et al., 2006; Lickel, Schmader, & Hamilton, 2003), reducing these perceptions could plausibly erode collective blame. Video 2 challenged the perception of Muslim homogeneity by presenting the diversity of the Muslim experience as part of an engaging TED Talk, and Videos 1, 5, and 6 challenged Muslim homogeneity by providing counterstereotypical exemplars, such as an assertive (vs. submissive) Muslim woman (Videos 1 and 6), or a soft-spoken Muslim cleric talking about his love for, and deference to, his wife (Video 5). Another video challenged perceived homogeneity of Muslims directly with didactic arguments during a confrontational television news interview (Video 3).

Many of the videos also included elements that we hypothesized would reduce anti-Muslim attitudes in general, but which could also affect collective blame specifically. For example, three of the videos (Videos 1, 5, and 6) gave participants the opportunity to hear directly from Muslims about their own experiences, allowing perspective-taking, which has been shown in many studies to improve intergroup attitudes and foster prosocial behavior (Bruneau & Saxe, 2012; Galinsky & Moskowitz, 2000). As the perspectives often illustrated nuanced views and challenged stereotypes, the shared perspectives could also reduce collective blame indirectly by reducing perceived homogeneity.

Another video (Video 7) provided normative examples of Americans espousing and engaging in pro-Muslim behaviors: A White mother helping her child donate to a vandalized mosque, and an interview with a man who attended a 2nd amendment rally across the street from a mosque wearing a "Fuck Islam" T-shirt who spoke of the transformation he experienced after accepting an invitation from the imam of the mosque to observe a service. Social proof has been shown previously to strongly influence behavior (McDonald & Crandall, 2015), and we thought it plausible that seeing others engage in pro-Muslim behaviors might influence participants' views of Muslims. The videos could specifically reduce collective blame through social proof—showing people who do not hold Muslims collectively responsible (even if they once did).

Two videos (Videos 4 and 6) challenged common beliefs about Muslims (that they hate America/Americans, and that Muslim immigrants would strain the economy) by citing data countering these views. Notably, in providing data from Pew surveys showing that people in the Muslim world generally respect and appreciate America/Americans, Video 6 challenged negative meta-perceptions—a technique that has been shown to reduce reciprocal hostility (Kteily, Hodson, & Bruneau, 2016). Finally, six videos (Videos 1, 2, 3, 4, 5, and 8) challenged the stereotype that Islam is a uniquely violent religion. For example, in Video 8, a documentarian read to people on the street passages from the "Quran" that condone violence, only to reveal that it was actually an *Old Testament* wrapped in a *Quran* book cover. Challenging the view that Muslim religious teachings are uniquely violent may indirectly reduce the perception that Muslims are intrinsically supportive of violence, which could reduce the tendency to blame all Muslims for the violent actions of individual group members.

It is worth noting that the psychological mechanisms listed above are speculative—reflecting what we theorized were the main psychological "ingredients" of each video—and not necessarily exhaustive. Moreover, individual videos sometimes included elements potentially operating via multiple psychological mechanisms, and the style of a given video or its protagonist could enhance or detract from the efficacy of any given intervention. Therefore, we suggest that this type of intervention tournament should not be used as evidence *against* the utility of any particular theoretical approach to reduce collective blame. Rather, we viewed this method as a first step to identify promising approaches, ruling *in* a subset of potentially relevant psychological factors that could then be subjected to further analysis and targeted verification. We take this approach here.

We randomly assigned non-Muslim American participants to view one of the intervention videos, an "empty" control condition (in which participants saw no-video), or a "negative control" condition. For the "negative control," participants watched a video in which a Muslim woman provided "criticism from within," suggesting that Islam is inherently violent, and that there is a clash between Muslim cultures and the West. We hypothesized that this video would *increase* collective blame of Muslims and hostility toward them, by framing them as an inherently violent group dedicated to aggressing against Western targets. Notably, evidence showing that decreasing collective blame improves attitudes toward Muslims *or* evidence showing that increasing collective blame worsens attitudes would support our theoretical suggestion that collective blame can cause anti-Muslim sentiments (although the former would, of course, be more useful for the practical purpose of promoting intergroup harmony). We also note that the arguments presented in the "negative control" are conceptually similar to many arguments presented on mainstream U.S. media in the aftermath of violent attacks by Muslims in the United States, making the inclusion of the "negative" control video both a theoretically informative and practically relevant point of comparison against which to assess the interventions. For links to each video and a summary of the information they contain, see Table 2.

Participants

We performed a power analysis using G*Power 3.1, and found that obtaining a small effect size ($d = .30$), with an alpha of .05 and power of .95 would require at least 135 participants per

Table 2. Summaries and Potential Psychological Mechanisms for the Videos Used for the Intervention Tournament (Study 2a) and Forecasting Tournament (Supplemental Study 1).

Link and summary of condition video	Potential psychological mechanisms	Length
Negative Control: Muslims Responsible https://player.vimeo.com/video/159133535 Interview with a Syrian-born woman who attacks Islam and Muslims as backward, primitive, violent, and at odds with Western civilization.	[Increased] homogeneity; stereotyping	3:12
Video 1: Collective Blame Hypocrisy https://player.vimeo.com/video/158199836 *Al Jazeera* interview with Linda Sarsour, a Muslim American woman who discusses the tendency to blame all Muslims for terror attacks, but not blame Christians for extremism by individual Christians.	Cognitive dissonance; perspective-taking; counterstereotyping; decrease homogeneity	2:07
Video 2: Homogeneity 1 https://player.vimeo.com/video/159133534 TED talk by a Muslim American man describing (and showing) his photo-journalistic journey through diverse Muslim communities around the world.	Decrease homogeneity; decrease entitativity; counterstereotyping	4:24
Video 3: Homogeneity 2 https://player.vimeo.com/video/158199837 CNN interview with Reza Aslan, a Muslim American scholar who challenges the view (expressed by the hosts) that policies in one Muslim-majority country should characterize "Muslims" or "Islam."	Decrease homogeneity; decrease entitativity; counterstereotyping; cognitive dissonance	4:04
Video 4: Counterstereotyping 1 https://player.vimeo.com/video/159133531 A segment from the satirical news program, The Daily Show, in which host John Oliver calls out media bias in their negative coverage of Muslims and Muslim violence.	Counterstereotyping; humanization; collective guilt	3:52
Video 5: Counterstereotyping 2 https://player.vimeo.com/video/158199845 A short video that witnesses an Egyptian imam and his wife describing their loving and respectful relationship to a small group of Muslims.	Counterstereotyping; decrease homogeneity; decrease entitativity; perspective-taking	2:41
Video 6: Challenge Meta-Perceptions https://player.vimeo.com/video/159133532 MSNBC interview with Dalia Mogahed, a Muslim American researcher who presents data from Pew surveys illustrating that Muslims view Americans and America favorably.	Improve meta-perceptions; perspective-taking; decrease homogeneity; decrease entitativity	3:35
Video 7: Normative Prosocial https://player.vimeo.com/video/160259623 Two news clips: A White conservative who describes his change of heart after visiting a mosque he was protesting across from; a White boy and his mother interviewed after donating money to a vandalized mosque.	Social proof (prosocial norms)	3:23
Video 8: Counterstereotyping 3 https://player.vimeo.com/video/159133527 "Gotcha" interviews where people respond to "Quran" passages that are primitive and intolerant by modern social norms, and then revealing that the passages were actually from a *Bible* in a *Quran* book cover.	Counterstereotyping; cognitive dissonance	3:16

condition. To allow for the loss of data from people who failed an embedded attention check, we recruited 180 participants for each of the 10 conditions in the study. Thirty-five participants failed the check question, leaving 1,765 participants in the final analyses (49.8% female, M_{age} = 34.75, SD = 11.3). The final sample was 46.2% Christian, 1.7% Jewish, 1.7% Buddhist, 0.7% Hindu, 43.8% atheist/agnostic, and 5.8% "Other." Ethnically, the sample was 77.8% White, 6.0% Asian, 5.8% Hispanic, 7.0% Black, 0.6% Native American, 0.2% Arab, 2.1% biracial, and 0.5% "Other."[1]

Procedure and Stimuli

Participants were randomly assigned to view one of the eight videos, the negative control video, or the no-video control condition. After viewing one of the videos (or not, in the no-video control condition), participants completed a survey, which included the key measure of collective blame, two measures assessing attitudes and beliefs about Muslims (dehumanization, prejudice) and two outcome measures (anti-Muslim policy support and support for punitive counterterrorism).

Table 3. Study 2a: Means (*SD*) and ANOVAs for Each Measure.

Condition	Collective blame	Blatant dehumanization	Prejudice	Punitive counterterrorism	Anti-Muslim policies
Scale	*0-100*	*100-+100*	*z score*	*0-100*	*1-7*
Muslims responsible (*N* = 177)	**40.08[a] (35.75)**	**15.48[a] (27.12)**	**.378[a] (1.08)**	65.15 (37.34)	3.13 (1.78)
No-video control (*N* = 174)	29.78 (34.18)	8.56 (25.18)	.068 (1.06)	67.14 (34.69)	2.95 (1.66)
Collective blame Hypocrisy (*N* = 176)	**20.66[a] (28.53)**	5.90 (21.96)	−.136[b] (0.91)	69.63 (33.56)	2.71 (1.61)
Homogeneity 1 (*N* = 178)	30.87 (33.27)	5.26 (18.57)	−.053 (0.88)	65.82 (36.53)	2.90 (1.72)
Homogeneity 2 (*N* = 171)	27.99 (32.46)	5.29 (22.59)	−.062 (0.93)	68.40 (34.42)	2.87 (1.63)
Counterstereotyping 1 (*N* = 177)	27.97 (33.29)	6.85 (20.15)	−.022 (1.00)	68.89 (36.76)	2.81 (1.65)
Counterstereotyping 2 (*N* = 179)	30.46 (34.77)	8.84 (25.54)	.081 (1.04)	66.17 (36.31)	2.94 (1.74)
Challenge meta-perceptions (*N* = 175)	23.40[b] (31.07)	9.31 (21.52)	−.086 (1.06)	69.59 (35.61)	2.85 (1.74)
Normative prosocial (*N* = 175)	23.12[b] (31.18)	5.85 (21.55)	−.112 (1.01)	68.27 (35.29)	2.80 (1.68)
Counterstereotyping 3 (*N* = 181)	29.63 (33.48)	8.96 (22.96)	−.069 (0.93)	65.30 (36.25)	2.97 (1.66)
ANOVA *F*(9, 1764)	4.72*** η^2 = .024	3.19** η^2 = .016	4.13** η^2 = .021	.42	.83

[a](and bold) Means that are significantly different from no-video controls ($p < .05$).
[b]Means that are marginally different from no-video controls ($p < .10$).

Collective Blame was assessed as in Study 1, but with respect to the Brussels Airport terror attack, which had occurred weeks prior to the study ("How responsible do you think Muslims in general are for the attacks at the Brussels Airport?").

Dehumanization was assessed using the Ascent Dehumanization scale (Kteily et al., 2015), which presents participants with the popular "Ascent of Man" diagram and asks them to determine where target groups fall on the scale, from the quadrupedal early human ancestor (0) to fully "evolved" modern human (100). We took as our measure of dehumanization the difference in reported "evolvedness" between Americans and Muslims.

Prejudice was assessed as in Study 1, by standardizing and averaging the Islamoprejudice composite (α = .89) and feeling thermometer ratings; $r = .74$, $p < .001$. *Anti-Muslim Policy Support* (α = .92) and *Punitive Counterterrorism* were also assessed as in Study 1.[2]

Results

Interventions tournament. For descriptive statistics and variable intercorrelations for the control condition, see Table S1. Mean results for each condition, ANOVAs, and *t* tests are presented in Table 3.

In the intervention tournament, participants were assigned to view one of the videos (or not, in the no-video control condition) and were then asked to report their collective blame of Muslims. We found a significant effect of condition on collective blame, $F(9, 1719) = 4.72$, $p < .001$, $\eta^2 = .024$. We then conducted a series of planned *t* tests to examine the differences in collective blame between those in the control condition and those who watched each of the videos. Only Video 1 ("Collective Blame Hypocrisy") significantly reduced collective blame ($M = 20.66$, $SD = 28.53$) relative to the no-video control condition, $t(343) = 2.69$, $p = .008$, $d = .29$. Collective blame was also significantly lower for those in the Collective Blame Hypocrisy condition versus those in four of the remaining seven intervention conditions ($ts > 2.1$, $ps < .04$). Mean collective blame scores were marginally lower among participants who viewed Videos 6 ("Challenge Meta-Perceptions") and 7 ("Normative Prosocial") relative to those in the no-video control condition ($ts > 1.8$, $ps < .075$).

On the contrary, participants who viewed the negative control video ("Muslims Responsible") reported significantly higher collective blame ($M = 40.08$, $SD = 35.75$) versus those in the no-video control condition, $M = 29.78$, $SD = 34.18$; $t(349) = 2.76$, $p = .006$, $d = .30$, and versus those in

each of the other intervention conditions (ts > 2.4, ps < .02). See Table 3 for means responses to all measures across condition and Figure S1 for a graphical depiction of mean collective blame ratings across all videos.

We also examined mean responses to the two negative attitudes and beliefs (blatant dehumanization, prejudice), and the two outcome measures (support for punitive counterterrorism, support for anti-Muslim policies). There was a main effect of condition for the attitudes and beliefs, blatant dehumanization: $F(9, 1757) = 3.19$, $p = .001$, $\eta^2 = .016$; prejudice: $F(9, 1763) = 4.13$, $p < .001$, $\eta^2 = .021$, but not for the outcome measures (Fs < 1). We conducted planned t tests on each of dehumanization and prejudice to see which intervention(s) drove the effect.

Those in the Collective Blame Hypocrisy condition (Video 1) reported marginally lower levels of prejudice ($M = –.136$, $SD = 0.91$) than no-video controls, $M = .068$, $SD = 1.06$; $t(355) = 1.95$, $p = .052$, $d = .21$. At the same time, negative controls reported greater prejudice ($M = .378$, $SD = 1.08$) than no-video controls, $t(359) = 2.75$, $p = .006$, $d = .29$, Collective Blame Hypocrisy, $t(354) = 4.87$, $p < .001$, $d = .52$, and all other conditions (ts > 2.6, ps < .008).

The main effect of dehumanization was driven primarily by the negative control: Those in the negative control dehumanized Muslims more ($M = 15.48$, $SD = 27.12$), compared with the no-video controls, $M = 8.56$, $SD = 25.18$; $t(357) = 2.51$, $p = .013$, $d = .27$, and compared with those in the Collective Blame Hypocrisy condition, $M = 5.90$, $SD = 21.96$; $t(351) = 3.64$, $p < .001$, $d = .39$. Dehumanization reported by those in the Collective Blame Hypocrisy condition was lower than no-video controls, but not significantly so, $t(354) = 1.06$, $p = .289$.

We conceptualized collective blame as a belief that could shape anti-Muslim policy support and behavior both directly and also indirectly by increasing negative attitudes and beliefs about Muslims (i.e., prejudice, dehumanization). In particular, to the extent that individuals collectively blame all Muslims for mass violence committed by a few, people could come to feel more dislike toward Muslims and see the group in more dehumanized terms (e.g., as "savages"). Given that prejudice and dehumanization are both known to (independently) predict hostile outcomes (e.g., Kteily et al., 2015), any effect of collective blame on these constructs could have downstream consequences on their anti-Muslim policy support. Focusing on the effects of the most effective video (i.e., Collective Blame Hypocrisy—Video 1), we used sequential mediation models (PROCESS, Hayes, 2012; Model 6) for each outcome measure to test the effect of condition (collective blame hypocrisy vs. control) on anti-Muslim policies, with collective blame as a first mediator and prejudice or dehumanization as subsequent mediators (to isolate their unique effects, we controlled for dehumanization when examining prejudice and vice versa). Specifically, we examined the indirect effects of the intervention on the outcome measures via collective blame. These included the indirect effect of the intervention from collective blame directly to the outcome measures (i.e., independent of prejudice and dehumanization), as well as the sequential indirect effects from the intervention to the outcomes through collective blame and then each of prejudice and dehumanization.

We found that collective blame (CB) directly mediated the effect of condition on both distal outcome measures (condition → CB → anti-Muslim policies and punitive counterterrorism). There were also significant indirect effects from condition to the outcome measures via collective blame's link to prejudice (condition → CB → prejudice → anti-Muslim policies and behavior). Notably, the indirect effect of condition on outcomes through prejudice alone (condition → prejudice → anti-Muslim policies) was not significant. The indirect effect from condition to the outcomes via collective blame's link to dehumanization (condition → CB → dehumanization → anti-Muslim policies) was not significant, nor was the indirect effect of condition on outcomes through dehumanization alone (condition → dehumanization → anti-Muslim policies). See Figures S2 and Table S2.[3]

Forecasting Tournament. To test the potential discrepancy between lay perceptions of our interventions' effectiveness and their actual effectiveness, we assigned a separate group of Americans ($N = 938$) to view a video and predict its effect on collective blame (see Supplemental Study 1).

In this forecasting tournament, participants reported their predictions about how the video they were randomly assigned to watch would affect Americans' collective blame of Muslims relative to control levels of collective blame:

> When asked how responsible they think Muslims in general are for the Paris attacks in 2015, American mTurkers report an average response of 30 on a 100-point scale, where 0 = *not at all responsible* and 100 = *completely responsible*. After watching this video, how much do you think American mTurkers will hold Muslims responsible for the Paris attacks?

Participants were then provided a slider anchored at 0 (*not at all responsible*) and 100 (*completely responsible*), with the slider starting point set at 30 on the scale (i.e., the mean response for control participants in the interventions tournament).

Interestingly, the actual effect of the collective blame hypocrisy video on collective blame was significantly greater than the effect predicted by forecasters, $t(272) = 2.44$, $p = .015$, $d = .30$. See Supplementary Study 1 for details and Figure S3 and Table S3 for a summary of results.

Discussion

Overall, the intervention tournament revealed a strategy (Collective Blame Hypocrisy) that significantly reduced collective blame and marginally reduced prejudice. Although

this intervention did not have direct effects on anti-Muslim policies (i.e., distal outcome measures), we did find evidence of significant indirect effects of the intervention on outcomes via its reduction of collective blame.

Two other approaches resulted in marginally significant reductions in collective blame: An interview with a Muslim woman who was friendly, intellectual, and assertive (i.e., counterstereotypical), and who presented evidence from Pew surveys challenging negative meta-perceptions among Americans with respect to Muslims (Video 6—"Challenge Meta-Perceptions"). This is consistent with previous research, which demonstrated a reduction in anti-Muslim attitudes among participants who were provided with a purportedly real newspaper article that included information from these same surveys suggesting that Muslims saw Americans in a humanizing light (Kteily et al., 2016). A second intervention that resulted in marginally lower collective blame was a video that presented two examples of White Americans engaging in prosocial gestures toward Muslims (Video 7—"Normative Prosocial"). These effects are consistent with research on the impact of social norms on prejudice (e.g., Crandall, Eshleman, & O'Brien, 2002). Although we decided to continue here by focusing on the Collective Blame Hypocrisy intervention because it had the numerically largest effects, future research could explore the potential for normative intergroup interventions to shift intergroup attitudes and behavior.

It is also noteworthy that the remaining five interventions failed to significantly reduce collective blame (or other anti-Muslim sentiments). That said, because we "crowdsourced" the videos, rather than developing them ourselves to test specific theories, they often contained multiple potential psychological elements. These elements were also packaged in ways that could have enhanced or detracted from their efficacy. These results should therefore not be used as evidence *against* specific psychological theories. Instead, taking a *rule-in* (vs. *rule-out*) approach, we zoomed in on the Collective Blame Hypocrisy video, which, among other things, highlighted individuals' hypocrisy in collectively blaming some groups but not others for the actions of a few. We focused on replicating the effects of this video (Study 2b) and then verifying our theoretical supposition that highlighting hypocrisy was central to its effects (Studies 3a and 3b).

Study 2b

Study 2a established that revealing the hypocrisy of collectively blaming Muslims but not White people/Christians for individual acts of violence significantly decreased collective blame of Muslims and marginally reduced prejudice; it also revealed that a video framing Muslims as collectively responsible for violence (i.e., Negative control—"Muslims Responsible") increased collective blame, prejudice, and dehumanization. However, the number of conditions (and thus comparisons) increases the probability that our results reflected a false positive. In Study 2b, we therefore sought to replicate the results of Study 2a, focusing on the critical conditions: Video 1—Collective Blame Hypocrisy, Negative Control—Muslims Responsible, and no-video control.

Participants. To obtain similar sample sizes for each of the three conditions as were obtained in Study 2a, we recruited 600 participants on Mechanical Turk. Three people did not finish the survey, and 15 people failed the check question, leaving 582 participants (329 female, M_{age} = 34.69, SD = 11.70). The final sample was 52.9% Christian, 1.9% Jewish, 1.9% Buddhist, 0.2% Hindu, 37.1% atheist/agnostic, and 6.0% "Other." Ethnically, the sample was 77.3% White, 5.2% Asian, 6.4% Hispanic, 7.7% Black, 0.7% Native American, 0.2% Arab, 1.9% biracial, and 0.7% "Other."

Procedures and Stimuli. The procedure was identical to Study 2a, except that participants were randomly assigned to one of only three conditions: Video 1—Collective Blame Hypocrisy, Negative Control—Muslims Responsible, or no-video control.

Collective blame was assessed as in Study 1.

Blatant Dehumanization was assessed with the multi-item trait measure from Study 1 (α = .95) and the single-item Ascent dehumanization measure from Study 2a. We combined the measures to form a single scale by standardizing each and averaging them together (see, for example, Kteily & Bruneau, 2017).

Prejudice was assessed as in Studies 1 and 2a, by averaging the z scored feeling thermometer and Islamoprejudice (α = .88) ratings, r = .66, p < .001, as was *Anti-Muslim Policy Support* (α = .91).[4]

Results

For descriptive statistics and variable intercorrelations for the control condition, see Table S4. Mean results for each condition, ANOVAs, and t tests are presented in Table 4.

As predicted, a univariate ANOVA performed on collective blame revealed a main effect of condition, $F(2, 578) = 19.0$, p < .001, η^2 = .062. Follow-up independent-samples t tests replicated the results of Study 2a: Collective blame was approximately 10 points higher for those in the Muslims Responsible condition (M = 40.50, SD = 34.90) versus no-video controls, M = 29.89, SD = 33.03; $t(390) = 3.09$, p = .002, d = .31, and approximately 10 points lower for those in the Collective Blame Hypocrisy condition (M = 20.47, SD = 26.57) versus no-video controls, $t(386) = 3.09$, p = .002, d = .31; Table 4, Figure S4.

Demonstrating a similar pattern as Study 2a, but with stronger effects, ANOVAs for anti-Muslim attitudes and beliefs (prejudice, blatant dehumanization) were also significant (Fs > 9.9, ps < .001, η^2 > .030). Follow-up planned t tests showed that dehumanization was significantly higher for those in the Muslims Responsible condition (M = .24, SD = 1.02) than the no-video controls, M = –.030,

Table 4. Study 2b Results: Means for All Measures Across Conditions, Omnibus ANOVAs, and Independent *t* Tests Across Conditions.

	Condition	Collective blame	Blatant dehumanization	Prejudice	Anti-Muslim policies
	Scale	*0-100*	*z score*	*z score*	*1-7*
Means (*SD*)	Muslims responsible (*N* = 193)	40.50 (34.90)	.237 (1.02)	.241 (1.03)	3.26 (1.85)
	No-video controls (*N* = 199)	29.89 (33.03)	−.030 (1.01)	.00 (1.00)	3.05 (1.71)
	Collective blame hypocrisy (*N* = 190)	20.47 (26.57)	−.210 (0.92)	−.244 (0.90)	2.78 (1.67)
	ANOVA *F*(2, 578)	*19.0**** η^2 = .062	*9.97**** η^2 = .033	*11.67**** η^2 = .039	*3.59** η^2 = .012
Independent *t* tests: *t* value (Cohen's *d*)	Control vs. hypocrisy *t*(387)	3.09*** (*d* = .31)	1.84† (*d* = .19)	2.50* (*d* = .26)	1.57
	Muslims resp vs. control *t*(390)	3.09*** (*d* = .31)	2.59* (*d* = .26)	2.36* (*d* = .24)	1.16
	Muslims resp vs. hypocrisy *t*(381)	6.30*** (*d* = .65)	4.50*** (*d* = .46)	4.89*** (*d* = .50)	2.64*** (*d* = .27)

†*p* < .10. **p* < .05. ***p* < .01. ****p* < .001.

SD = 1.01; *t*(390) = 2.59, *p* = .010, *d* = .26; dehumanization was marginally lower for those in the Collective Blame Hypocrisy condition (*M* = –.21, *SD* = 0.92) than the no-video controls, *t*(387) = 1.84, *p* = .067. Results for prejudice were very similar: Prejudice among those in the Muslims Responsible condition (*M* = .24, *SD* = 1.03) was significantly greater than for those in the no-video controls, *M* = .00, *SD* = 1.00; *t*(390) = 2.36, *p* = .019, *d* = .24, and prejudice was significantly lower for those in the Collective Blame Hypocrisy condition (*M* = –.24, *SD* = 0.90) than for no-video controls, *t*(387) = 2.50, *p* = .013, *d* = .26.

There was also a main effect of condition for the outcome measure, anti-Muslim policy support, *F*(2, 579) = 3.59, *p* = .028, η^2 = .012, driven by a significant difference between the Muslims Responsible video and the Collective Blame Hypocrisy video, *t*(381) = 2.64, *p* = .009, *d* = .27. The difference between each video and the no-video control condition did not reach significance (*t*s < 1.6, *p*s > .10). See Table 4 for a summary of results.

As with Study 2a, we tested the indirect effects of the intervention (Collective Blame Hypocrisy video vs. no-video control) on anti-Muslim policy support. Consistent with Study 2a, there was a significant indirect effect from the condition to anti-Muslim policies via collective blame (condition → CB → anti-Muslim policies). Beyond the role collective blame played independent of prejudice and dehumanization, we also observed a significant sequential indirect effect from the intervention condition on anti-Muslim policies through collective blame's relationship with prejudice (condition → CB → prejudice → anti-Muslim policies) and (unlike Study 2a) dehumanization (condition → CB → dehumanization → anti-Muslim policies). Also consistent with Study 2a, once the indirect effects via collective blame were taken into account, there were no significant indirect effects of condition on outcomes through prejudice or dehumanization (i.e., condition → prejudice/dehumanization → anti-Muslim policy support; all indirect effects included 0 in the 95% confidence intervals). See Figure S5 and Table S5. Study 2b therefore replicated the main results from Study 2a.[5]

Study 3a

Studies 2a and 2b provided evidence that a 2-min video interview with a Muslim American woman was sufficient to change how much people collectively blamed Muslims in general for individual acts of violence. A portion of the interview included a comment by the Muslim American guest that many people blame all Muslims for actions committed by individual Muslims, but do not blame all Christians for the actions of Christian extremist groups (i.e., the "Westboro Baptist Church" and the Ku Klux Klan [KKK], which are characterized as Hate Groups by the Southern Poverty Law Center; "*Extremist Files: Westboro Baptist Church*," 2016). We hypothesized that this video effectively reduced collective blame because it helped to reveal to viewers the (potentially unconscious) hypocrisy of holding some groups (i.e., Muslims) more responsible for the actions of individual group members than other groups (i.e., White Americans, Christians). As holding inconsistent views is generally aversive (Festinger, 1962), we reasoned that the specter of hypocrisy was enough to cause people to reduce their attributions of collective blame so that they could avoid the inconsistency. However, video stimuli are inherently complex, and it is possible that other aspects of the video were in fact responsible for the observed effects. If revealing the hypocritical nature of collective blame was in fact driving our effects as we assumed, then revealing it in

other ways (i.e., beyond the specific video used in Studies 2a and 2b) should yield similar effects.

In Study 3a, we therefore sought to specifically test the effects of revealing the intergroup bias in collective blame using a different and more controlled method. Rather than exposing participants to the hypocrisy of collective blame through a didactic argument, we illuminated the hypocrisy to participants through a targeted interactive activity that employed a Socratic approach. In the activity, participants first reported how much they blamed themselves and White Americans for acts of mass violence committed by highly self-identified White men. Next, using the same slider scale, participants reported how much they blamed individual Muslims for a terror attack. Finally, they reported how much they blamed Muslims in general for an act of mass violence committed by Muslims (i.e., collective blame). We reasoned that people would be very unlikely to blame themselves or White Americans for acts of mass violence by ingroup members, and that they would subsequently hold Muslims minimally responsible for acts of terrorism to avoid cognitive dissonance. In line with the results from Studies 2a and 2b, we further predicted that the hypothesized reductions in collective blame would mediate reductions in anti-Muslim policy support and anti-Muslim behavior, both directly and by reducing anti-Muslim attitudes and beliefs (i.e., prejudice and dehumanization).

Method

Participants and design. We recruited 605 participants from Mechanical Turk for a five-condition study. Sample sizes were slightly smaller than obtained in Studies 2a and 2b, but still large enough to provide 80% power to detect a small to medium effect size ($d = .35$). Twelve people failed the attention check question, leaving 593 participants (314 female, $M_{age} = 35.56$, $SD = 11.78$). The final sample was 79.6% White, 5.7% Asian, 4.6% Hispanic, 7.4% Black, 1.0% Native American, 0.5% Middle Eastern, 1.0% biracial, and 0.2% "Other." Due to a coding error, religious affiliation was not collected.

Participants were randomly placed into one of five conditions: A Collective Blame Hypocrisy activity, a no-activity control condition, or one of three alternative activities (described below) that were inspired by psychological theory and represented in arguments that were widely circulated through social media (in an attempt to reduce anti-Muslim sentiments) in the wake of terror attacks by Muslim extremists.

Procedure and stimuli. The Collective Blame Hypocrisy activity was composed of two parts. First, participants reported how responsible they held White Americans and themselves for three different individual acts of violence committed by White people: Dylann Roof (who killed nine Black parishioners at a church in 2015), Anders Breivik (who killed 77 Norwegians, mostly children, in 2011), and Wage Page (who killed six Sikhs at a temple, believing they were Muslims, in 2012). To foreshadow a comparison with violence committed by "Muslim extremists," we noted that each perpetrator was motivated by his White identity. For example, "On June 17, 2015, Dylann Roof entered the Emanuel African Methodist Episcopal Church, and during a prayer service killed nine African American parishioners. Roof cited his White identity as a motivation for the attacks." Participants then responded to the following: "How responsible do you think you are for the acts of Dylann Roof?" and "How responsible do you think White Americans are for the acts of Dylann Roof?" Responses to each question were made using unmarked sliders anchored at 0 (*not at all responsible*) and 100 (*completely responsible*). We then asked how responsible participants felt White Americans were for hate crimes by White supremacists in the United States, and White supremacists in Europe. We predicted that participants would attribute very little responsibility to themselves and White Americans for the specific actions of mass violence, and for hate crimes committed by White supremacist groups.

Next, we asked participants to report, using the same scales, how responsible they felt individual Muslims were for an act of violence committed by Muslim extremists (e.g., "Ahmad works as a bank teller in Jordan. How responsible do you think Ahmad is for the Brussels Airport attacks?"). Finally, we asked how responsible they thought Muslims were, in general, for the Paris terror attacks. Overall, we hypothesized that reporting low levels of collective blame for oneself and White Americans would precipitate lower levels of collective blame of Muslims, in general, for terror attacks, which would have downstream effects on anti-Muslim attitudes and policy support.

Similar to Study 2a, we examined the impact of the Hypocrisy activity relative to other popular approaches that mapped broadly onto psychological theories suggesting their plausibility in reducing collective blame. The first ("Ingroup Guilt") exposed participants to historical opinion polling prior to and during World War II showing that Americans were opposed to accepting Jewish refugees. We hypothesized that this could elicit collective ingroup guilt for American rejection of Jews during the Holocaust—a moral emotion that has been shown to facilitate support for reparations (Brown, González, Zagefka, Manzi, & Čehajić, 2008; Čehajić-Clancy, Effron, Halperin, Liberman, & Ross, 2011; Lickel, Schmader, & Barquissau, 2004). We reasoned that individuals who were exposed to this information might soften their attitudes toward Muslims and Muslim refugees to assuage their guilt. This strategy was used widely in social media to evoke sympathy for Muslim refugees, and was reported on by major news outlets (e.g., *The Washington Post*; Tharoor, 2015). A second version of the intervention additionally presented photos directly drawing the link between interned Jewish children and interned Muslim refugee children, and provided a statement by the Holocaust

Memorial admonishing governments for their refusal to accept Muslim refugees. Because this version had an additional component that, at least in theory, strengthened the basis for feeling guilt, we labeled this intervention "Ingroup Guilt+." Although we thought it plausible that these two interventions (Ingroup Guilt and Ingroup Guilt+) could also reduce collective blame of Muslims, we thought it most likely that this intervention would change policy support and behaviors toward Muslims via reducing prejudice.

The final intervention was designed to challenge stereotypes about Muslim aggression by highlighting participants' incorrect assumptions. As with the stereotype reduction videos, we predicted that challenging the stereotype of Muslims as violent may reduce the tendency to blame all Muslims for the violent actions of individual group members. In the activity ("Counterstereotyping"), participants were first asked to guess statistics related to aggression by Muslims and refugees (e.g., the percent of European terror attacks in the past 10 years that had been perpetrated by Muslims). After guessing, participants were shown the true answer, which was consistently less in line with prevailing stereotypes than their estimates. Specifically, the mean estimate for the percent of European terror attacks committed by Muslims over a 5-year period was 38.75% (SD = 31.92), and the correct response, subsequently revealed, is less than 2% (more than 97% of the sample overestimated the statistic). Similarly, of the 190,000 murders committed in the United Sates since 9/11, participants guessed that on average 5,042 (SD = 18,742) were committed by Muslim extremists, whereas the correct answer is 37 (more than 65% of participants overestimated); and of the 194,000 refugees granted shelter in the United States since 9/11, participants guessed on average that 899 (SD = 5,580) had committed murder, whereas the correct answer is 0 (more than 70% of participants overestimated).[6] As part of collectively blaming Muslims for violence likely involves the stereotype that Muslims as a group are violent, we predicted that challenging this perception could potentially reduce collective blame and anti-Muslim sentiments.

After completing one of the activities (or no activity in the control condition), participants completed a survey that included the key measure of collective blame, as well as blatant dehumanization, prejudice, and two downstream outcome measures: support for anti-Muslim policies and signing anti-Muslim petitions.

Collective Blame was assessed as in Studies 1 and 2b (i.e., toward the Paris terror attacks).

Dehumanization was assessed as in Study 2b, by standardizing and then combining the trait measure ($\alpha = .91$) and the Ascent dehumanization measure ($r = .55$, $p < .001$).

Prejudice was assessed with feeling thermometers, and expressed as the difference between warmth toward Americans versus Muslims.

Anti-Muslim Policy Support ($\alpha = .94$) was assessed as in Studies 1, 2a, and 2b; *Signing Anti-Muslim Petitions* ($\alpha = .87$) was assessed as in Study 1.[7]

Results

For descriptive statistics and variable intercorrelations for the control condition, see Table S6. Mean results for each condition, ANOVAs, and t tests are presented in Table 5. None of the measures differed across the Ingroup Guilt versus Ingroup Guilt+ interventions (ts < 1.3, ps > .20), so results were collapsed across the two.

First, we assessed levels of blame attributed to oneself, White Americans, and individual Muslims for those who engaged in the Collective Blame Hypocrisy activity. We found that self-blame for the three specific events was near floor on the 100-point scale (M = 9.73, SD = 19.46). Collective blame was similarly low for White Americans across the three events (M = 9.73, SD = 18.64), for White supremacists in the United States (M = 12.40, SD = 24.88) and White supremacists in Europe (M = 9.59, SD = 19.86). After assessing blame for themselves and White people, Americans attributed very little blame to individual Muslims (M = 8.65, SD = 22.28).

Next, we turned to the primary measure of interest: collective blame of Muslims. Overall, there was a main effect of condition, $F(4, 582) = 5.37$, $p < .001$, $\eta^2 = .036$, and planned t tests revealed the predicted outcome: For participants who participated in the Collective Blame Hypocrisy activity, collective blame of Muslims (M = 17.78, SD = 29.07) was half what it was for controls, M = 35.46, SD = 37.59; $t(302) = 4.12$, $p < .001$, $d = .47$. Collective blame was also significantly lower for those who engaged in the Collective Blame Hypocrisy activity than for those who took part in each of the other activities (Ms = 35.01-36.93; ts > 3.7, ps < .001; see Figure S6). There were also main effects for blatant dehumanization, anti-Muslim refugee policy support, and anti-Muslim petition signing (see Table 5) that was driven by the Collective Blame Hypocrisy activity: Compared with the control condition, those who completed the Collective Blame Hypocrisy activity reported less blatant dehumanization, $t(303) = 2.38$, $p = .016$, $d = .27$, showed less anti-Muslim refugee policy support, $t(303) = 2.04$, $p = .042$, $d = .23$, and were less likely to sign anti-Muslim petitions, $t(303) = 2.24$, $p = .026$, $d = .26$. Only prejudice was not significantly different for those in the Collective Blame Hypocrisy condition versus no-activity controls, $t(303) = 1.6$, $p = .11$.

Aside from Hypocrisy, none of the other activities were significantly different from the control condition for any of the measures (ts < 1.6, ps > .10). See Table 5 for full results across all conditions.

As with Studies 2a and 2b, we tested the sequential indirect effect of the intervention (Collective Blame Hypocrisy activity) versus no-activity control on outcomes, with collective blame as a first mediator and dehumanization or prejudice (controlling for the other) as a second mediator. Consistent with Studies 2a and 2b, there were significant indirect effects of condition on the outcome measures via collective blame (condition → CB → anti-Muslim policies; condition → CB → anti-Muslim behavior).

Table 5. Study 3a Results: Means for All Measures Across Conditions, Omnibus ANOVAs, and Independent *t* tests Across Conditions.

	Condition	Collective blame	Blatant dehumanization	Prejudice	Support anti-Muslim policies	Sign anti-Muslim petitions
	Scale	*0-100*	*z score*	*−100-+100*	*1-7*	*−1-+1*
M (*SD*)	No-activity controls (*N* = 206)	35.45 (37.59)	0.117 (.928)	30.74 (36.39)	3.87 (1.80)	−0.062 (.441)
	Hypocrisy Activity (*N* = 99)	17.78 (29.07)	−0.148 (.875)	23.42 (39.10)	3.43 (1.68)	−0.184 (.458)
	Ingroup guilt (*N* = 190)	35.29 (35.59)	−0.023 (.830)	27.34 (36.54)	3.78 (1.74)	−0.070 (.477)
	Counterstereotype (*N* = 98)	37.40 (36.23)	−0.052 (.785)	24.13 (35.16)	3.57 (1.62)	−0.080 (.470)
	ANOVA *F*(4, 582)	5.37*** η^2 = .036	1.82	0.931	1.41	1.73
Independent *t* tests: *t* value (Cohen's *d*)	Hypocrisy vs. control *t*(303)	4.12*** (*d* = .47)	2.38* (*d* = .27)	1.60	2.04* (*d* = .23)	2.24* (*d* = .26)
	Hypocrisy vs. guilt *t*(287)	4.21*** (*d* = .50)	1.19	0.84	1.65	1.94† (*d* = .23)
	Hypocrisy vs. counterstereo *t*(195)	4.18*** (*d* = .60)	0.81	0.13	.59	1.57
	Guilt vs. control *t*(394)	0.04	1.58	0.93	.50	0.19
	Challenge vs. control *t*(302)	0.42	1.56	1.50	1.41	0.33

Note. Counterstereo = Counterstereotype intervention.
†*p* < .10. *p < .05. **p < .01. ***p < .001.

Also consistent with Studies 2a and 2b, there were significant sequential indirect effects from the condition to each of the outcomes via collective blame's link to prejudice (condition → CB → prejudice → anti-Muslim policies; condition → CB → prejudice → anti-Muslim behavior) and dehumanization (condition → CB → dehumanization → anti-Muslim policies; condition → CB → dehumanization → anti-Muslim behavior). Again consistent with Studies 2a and 2b, all the indirect effects were mediated through collective blame: The indirect effects of condition on outcomes through prejudice or dehumanization controlling for collective blame (i.e., condition → prejudice/dehumanization → outcomes) were all nonsignificant (all 95% confidential intervals [CIs] included 0). For results regarding anti-Muslim behavior, see Figure 1 and Table 6. For results regarding anti-Muslim policy support, see Figure S7 and Table S7.

Discussion

In sum, we followed up on the successful video intervention from Studies 2a and 2b by targeting the specific mechanism that we posited had been crucial to the video's success (i.e., revealing hypocrisy in collective blame). In the interactive activity, participants were required to first reflect on their own (and White people's) *lack of* collective responsibility for the actions of individual group members. Subsequently, they reported levels of collective blame of Muslims that were half those reported in the control condition, thus avoiding potential cognitive dissonance. Consistent with our theorizing about the causal impact of collective blame, those who completed the Collective Blame Hypocrisy activity reported lower levels of prejudice and dehumanization, and were less likely to endorse anti-Muslim policies, and sign anti-Muslim petitions, relative to no-activity controls. All changes in outcomes were mediated by the activity's effect on collective blame.

Study 3b

As with the video interventions tournament, we conducted a follow-up study to replicate the effects from the activities tournament. In Study 3b, we also examined a variant of the Collective Blame Hypocrisy activity ("Hypocrisy+"). This activity was identical to the Collective Blame Hypocrisy activity, with one exception: After completing the activity, participants in this condition were asked to report on (a) a time when they thought they were responsible for the negative actions of an ingroup member, (b) a time when they thought an outgroup member was responsible for the actions of another outgroup member, and, finally, (c) whether they found it easier to generate an example in which others (vs. they themselves) were responsible for their group member's actions. We hypothesized that the additional reflection about internal biases induced by completing these questions could potentially accentuate the effects of the activity.

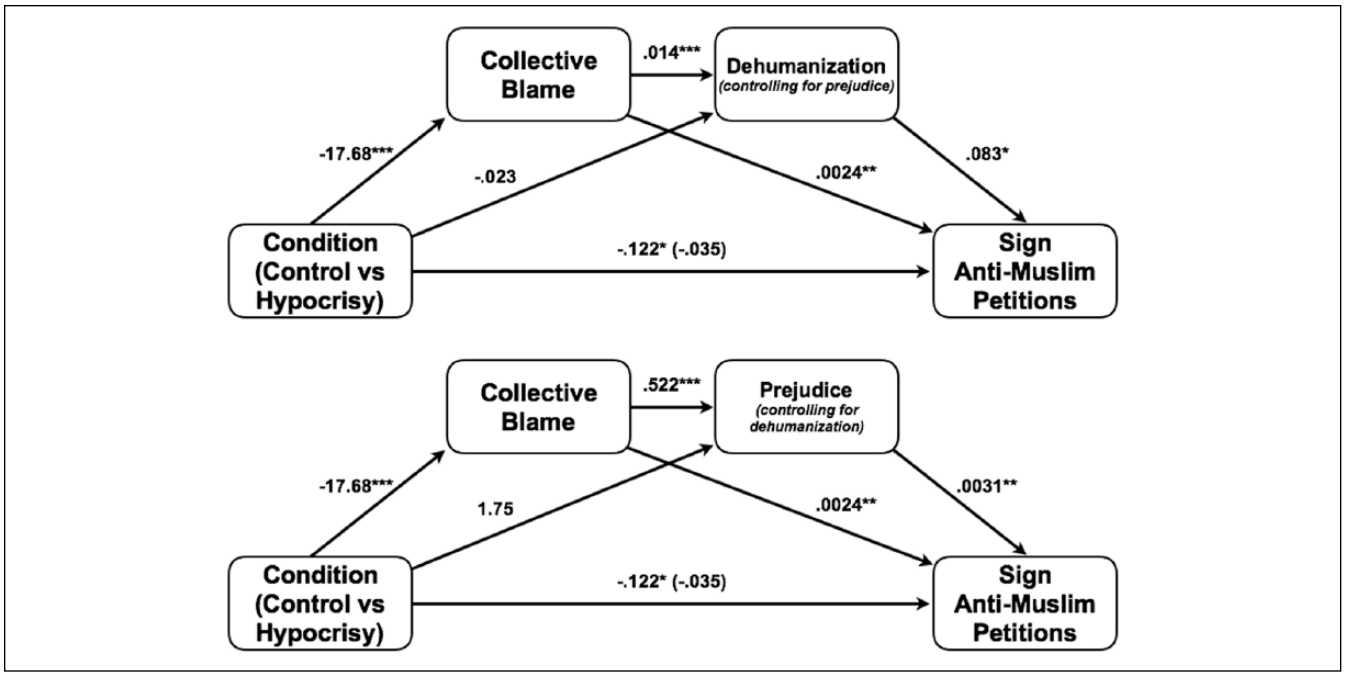

Figure 1. Model testing the effect of condition (Control vs. Collective Blame Hypocrisy activity) on anti-Muslim behavior (signing anti-Muslim petitions) through collective blame and either dehumanization or prejudice (while controlling for the other) for Study 3a.
Note. Unstandardized coefficients displayed.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 6. Study 3a.

Effects	Anti-Muslim petitions
Cond → CB → outcome	**–.042 [–.079, –.017]**
Cond → Dehum → outcome	−.002 [–.020, .013]
Cond → Prejudice → outcome	.005 [–.019, .033]
Cond → CB → Dehum → outcome	**–.020 [–.043, –.004]**
Cond → CB → Prejudice → outcome	**–.029 [–.053, –.013]**
Total indirect (CB + Dehum)	**–.064 [–.109, –.026]**
Total indirect (CB + Prejudice)	**–.065 [–.119, –.018]**
Total direct	−.035 [–.130, .061]
Total effect (CB + Dehum)	−.081 [–.178, .016]
Total effect (CB + Prejudice)	−.065 [–.164, .033]

Note. Unstandardized indirect, direct, and total effects of condition (Cond) on behavior (signing anti-Muslim petitions) through sequential mediators: Mediator 1 = collective blame (CB) and Mediator 2 = dehumanization (Dehum; controlling for prejudice) or prejudice (controlling for dehumanization). Results reported as point estimate with 95% confidence interval in brackets. Results in bold are significant: 95% CI does not include 0. CI = confidential interval.

Method

Participants and design. Consistent with the previous studies, we recruited 200 participants per each of three conditions. Of the 600 people recruited, 15 failed an embedded check question, leaving 585 participants (346 female, M_{age} = 34.54, SD = 11.50). The final sample was 48.4% Christian, 2.2% Jewish, 1.7% Buddhist, 0.3% Hindu, 40.5% atheist/agnostic, and 6.8% "Other." Ethnically, the sample was 77.3% White, 5.1% Asian, 5.6% Hispanic, 6.2% Black, 0.9% Native American, 0.5% Middle Eastern, 4.1% biracial, and 0.3% "Other."

Procedure and stimuli. The Collective Blame Hypocrisy activity was identical to the activity in Study 3a, with the following exception: Instead of asking how responsible White Americans were for the actions of American and European White extremists, we asked how responsible participants thought mainstream Christians were for the actions of the KKK.

Collective Blame was measured as in Study 2b (collective blame of Muslims for the Brussels Airport attacks), *Blatant Dehumanization* was measured with the trait measure of dehumanization used in Studies 1, 2b, and 3a, and *Islamoprejudice* was measured as in Studies 1, 2a, and 2b.[8]

Results

For descriptive statistics and variable intercorrelations, see Table S8. Mean results for each condition, ANOVAs, and t tests are presented in Table 7.

Among those who engaged in the Hypocrisy activities, self-blame was near floor on the 100-point scale (M = 4.74, SD = 18.03), and collective blame of White Americans and Christians was similarly very low (Whites: M = 9.78, SD = 17.84; Christians: M = 11.38, SD = 20.86). After assessing blame for themselves, White people and Christians,

Table 7. Study 3b Results: Means for All Measures Across Conditions, Omnibus ANOVAs, and Independent *t* Tests Across Conditions.

	Condition	Collective blame	Blatant dehumanization	Prejudice
	Scale	*0-100*	*1-7*	*0-100*
Means (*SD*)	No-activity controls (*N* = 193)	34.10 (34.52)	3.55 (1.40)	38.40 (27.12)
	Hypocrisy activity (*N* = 199)	11.70 (22.77)	3.27 (1.37)	33.73 (24.48)
	Hypocrisy + Activity (*N* = 193)	9.29 (16.92)	3.20 (1.21)	32.13 (24.46)
	ANOVA *F*(2, 582)	54.38*** η^2 = .158	3.88* η^2 = .013	3.19* η^2 = .011
Independent *t* tests: *t* value (Cohen's *d*)	Control vs. hypocrisy *t*(390)	7.61*** (*d* = .78)	2.04* (*d* = .20)	1.79† (*d* = .18)
	Control vs. hypocrisy+ *t*(581)	8.93*** (*d* = .97)	2.67** (*d* = .27)	2.39* (*d* = .24)
	Control vs. hypocrisy (combined) *t*(581)	10.39*** (*d* = .86)	2.74** (*d* = .23)	2.45* (*d* = .20)

†*p* < .10. *p < .05. **p < .01. ***p < .001.

Americans attributed very little blame to individual Muslims (*M* = 2.65, *SD* = 10.53).

For the key collective blame measure, we found the predicted main effect of condition, $F(2, 580) = 54.38$, $p < .001$, $\eta^2 = .16$, and planned *t* tests confirmed that collective blame of Muslims among those in the control condition (*M* = 34.10, *SD* = 34.52) was significantly greater than for those who engaged either in the Collective Blame Hypocrisy activity, *M* = 11.70, *SD* = 22.77; $t(390) = 7.61$, $p < .001$, $d = .78$, or the Collective Blame Hypocrisy+ activity, *M* = 9.29, *SD* = 16.92; $t(382) = 8.93$, $p < .001$, $d = .97$; see Table 7 and Figure S8.

Although we hypothesized that the Collective Blame Hypocrisy+ intervention might be significantly stronger than the Hypocrisy intervention, the two conditions did not differ from each other on any of the measures (*t*s < 1.2, *p*s > .23). As collective blame of Muslims among those in the Hypocrisy activity were just as low as collective blame of Christians and Whites (and lower than observed in Study 3a), it is possible that the lack of difference between Collective Blame Hypocrisy activities was due to a floor effect (i.e., the intergroup bias in collective blame was eliminated).

Also consistent with Study 3a, blatant dehumanization of Muslims was significantly lower for those in the Collective Blame Hypocrisy conditions (*M* = 3.23, *SD* = 1.29) versus controls, *M* = 3.55, *SD* = 1.40; $t(583) = 2.74$, $p = .006$, $d = .23$, and those who engaged in a Collective Blame Hypocrisy activity also expressed significantly less prejudice (*M* = 32.94, *SD* = 24.45) than controls, *M* = 38.40, *SD* = 27.12; $t(583) = 2.45$, $p = .015$, $d = .20$. See Table 7 and Figure S8. Anti-Muslim policy attitudes and behavior were not assessed in Study 3b.

Therefore, as with Study 3a, collective blame (and dehumanization) of Muslims was significantly reduced among participants who engaged in one of two variants of the Hypocrisy activity versus no-activity controls.

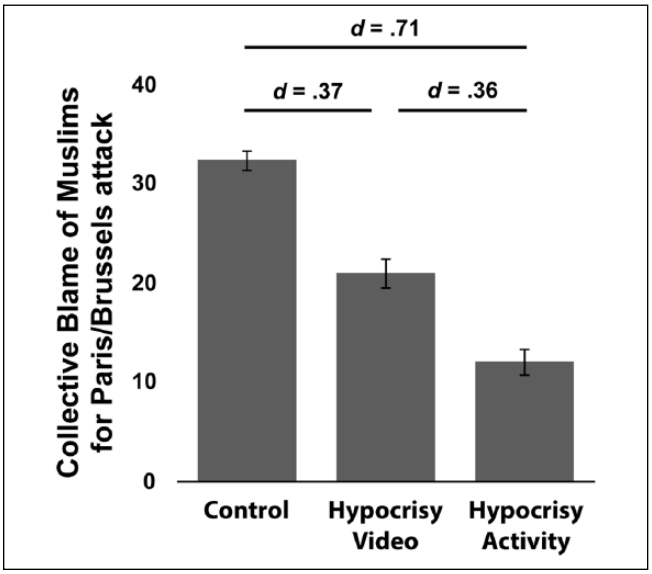

Figure 2. Results of meta-analysis across all participants who were in either the hypocrisy intervention or control condition for Studies 2a, 2b, 3a, and 3b.
Note. Error bars represent ± standard error of the mean. Differences between all groups *p* < .001. Reported are Cohen's *d* effect sizes.

Meta-Analysis

Finally, to compare results across the Hypocrisy video and Hypocrisy activities, we combined the results from the Hypocrisy video and controls (Studies 2a and 2b) and the Hypocrisy activities and controls (Studies 3a and 3b), and performed a 2 modality (video, activity) × 2 condition (intervention, control) ANOVA with collective blame as the outcome. We found that there was a strong main effect of condition, $F(1, 1659) = 114.89$, $p < .001$, $\eta^2 = .065$, illustrating that the hypocrisy approach reliably reduces collective

blame. We also found a significant Modality × Condition interaction, $F(1, 1659) = 19.12$, $p < .001$, $\eta^2 = .011$, such that collective blame was reduced more by engaging in the Collective Blame Hypocrisy activity ($M = 11.99$, $SD = 22.22$) versus watching the Collective Blame Hypocrisy video ($M = 20.92$, $SD = 27.75$). See Figure 2.

General Discussion

Intergroup violence is a major cause of death and suffering around the world. In 2014 alone, over 180,000 people died in violent conflicts, and an estimated 50 million were displaced (Armed Conflict Survey 2015, 2015). Understanding the psychological processes that feed conflict, and how to short-circuit these processes, may be critical to reducing human suffering (see Kteily & Bruneau, 2017). Here, we examine one process particularly relevant to conflict escalation—collective blame—which can induce vicarious retribution against uninvolved outgroup members following an act of violence (Lickel et al., 2006), and which we suggest can stimulate a conflict spiral in intergroup contexts. Our work advances previous research on the consequences of collective blame—which has been examined largely in organizational settings—by showing that collective blame of Muslims for terror attacks is associated with anti-Muslim attitudes and behavior.

Most importantly, we provide evidence that revealing the hypocrisy of collectively blaming Muslims for acts of terrorism, but not collectively blaming White people or Christians for individual acts of violence by members of those groups, reliably reduces collective blame of Muslims, and thereby decreases anti-Muslim attitudes and behavior associated with vicarious retribution. The results were robust whether the collective blame hypocrisy was revealed didactically in a brief video or when revealed through a Socratic activity. In both cases, the approaches highlighting hypocrisy in collective blame were numerically more effective than all other alternative videos and activities and significantly more effective than most. The effects of the intervention were also not obvious, as an independent sample of forecasters predicted that the Collective Blame Hypocrisy video used in Studies 2a and 2b would be completely ineffective—extending the literature on individuals' poor forecasting of effective interventions (Cialdini, 2003) by showing that this also applies to interventions aimed at reducing intergroup hostility.

Our findings contribute to a long tradition of research demonstrating that highlighting hypocrisy can drive behavior change. The hypocrisy paradigm has been effective in several contexts, increasing condom use (Aronson et al., 1991; Stone, Aronson, Crain, Winslow, & Fried, 1994), the recycling of household waste (Fried & Aronson, 1995), and respect of traffic laws (Fointiat, 2004). Our data extend this prior research, showing that similar results can also be obtained when it comes to reducing collective blame and improving intergroup attitudes and behaviors.

At the same time, the cognitive dissonance literature has also demonstrated that arguments highlighting hypocrisy and other threats to self-worth can sometimes backfire. For example, people have been shown to resolve the dissonance between their support for environmental policies and recalling times when they wasted water in some cases by reducing water usage (Dickerson et al., 1992), but in others by derogating the importance of water rationing policies (Liégeois, Yserbyt, & Corneille, 2005). If a hypocrisy intervention were to induce defensiveness—for example, by publicly shaming the participants—it seems likely that it could exacerbate, rather than ease, the desired attitude or behavior. Inductions of defensiveness may help to explain why some of the interventions from the current research failed to reduce collective blame or anti-Muslim attitudes and behavior, even though they might have led to cognitive dissonance by highlighting hypocrisy or revealing incorrect beliefs. For example, Video 3 called out the hypocrisy of reporters viewing all Muslim countries as the same (and not doing this with Christian countries), and Video 8 implied hypocrisy by having people respond to passages that were purportedly from the *Quran*, but were in fact revealed to be from the *Bible*. These approaches could not only invoke hypocrisy but also involve an element of public shaming or combativeness. It is possible that the combativeness reduces their efficacy.

By contrast, the didactic approach in the Collective Blame Hypocrisy video does not call out the hypocrisy of the interviewer, but reveals it as a view held by unnamed others. The Socratic approach in the activity is even more gentle, allowing participants to proactively *avoid* reporting any hypocrisy themselves. It would be interesting to see if adjusting the other potentially threatening interventions could render them more successful. For example, the "Counter-Stereotype" activity intervention from Study 3a could allow participants to discover the statistics about refugee violence on their own, rather than correcting them after their erroneous predictions. Because these interventions were chosen from popular approaches, were not explicitly designed to test particular theories, and might have been effective had they been presented in different ways, we think it important to emphasize that the absence of effects here for some of the interventions should not be held as evidence against a psychological approach to which they were mapped (e.g., counterstereotyping).

Although the principal focus of this research was on interventions aimed at reducing collective blame and other anti-Muslim sentiments, it is also worth highlighting the "effectiveness" of the negative control video at increasing collective blame attributions (and anti-Muslim attitudes and policy support). From a theoretical perspective, the fact that this intervention increased downstream hostility via collective blame provides further confidence in our proposed model: Just as reducing collective blame can reduce dehumanization and prejudice with consequences for outcomes like support for anti-Muslim policies, *increasing* collective blame can do the

reverse. From a practical perspective, the fact that a brief video can dramatically and reliably increase anti-Muslim sentiments highlights concerns about the anti-Muslim rhetoric increasingly prevalent on certain media platforms and frequently promulgated by the current U.S. president (Kteily & Bruneau, 2017). At the very least, the effectiveness of the negative control at increasing anti-Muslim hostility suggests that a great need for countermessaging exists. Determining what types of messages (including collective blame hypocrisy and beyond) might protect against or mitigate the negative effects of a compelling narrative that includes anti-Muslim speech represents an important avenue of future research.

Limitations and Future Directions

Despite the consistent and robust results from this research, it is important to note some limitations. First, it is important to acknowledge that the results were obtained exclusively from samples obtained through Mechanical Turk. Although mTurk samples are reliable and diverse (Buhrmester, Kwang, & Gosling, 2011), it is possible that other populations may respond differently. For example, very liberal samples (e.g., psychology undergraduates) may be near floor in collective blame to begin with, which would minimize or eliminate the efficacy of the intervention.

Second, still more work is required to better understand the psychological mechanisms behind the collective blame hypocrisy approach. Although the approach is similar to cognitive dissonance hypocrisy paradigms, there are some key differences that may reveal significant disparities between the effectiveness of these approaches in the intergroup context. For example, the strength of cognitive dissonance hypocrisy paradigms is enhanced when cognitive dissonance (and therefore internal discomfort) is maximized (Fried & Aronson, 1995). By contrast, the collective blame hypocrisy activity allows people to preemptively avoid cognitive dissonance. Allowing cognitive dissonance to establish itself first, for example, by having people declare their collective blame of Muslims before being exposed to an argument about hypocrisy in collective blame, might reduce the efficacy of the intervention by making people feel defensive and/or "trapped." These predictions should be tested in future research.

Third, we demonstrate here effects immediately following the interventions; future research should establish how long the effects of the intervention last.

Fourth, the results reported here refer to a single target group. Although collective blame of Muslims is an important phenomenon that could be driving some of the most pressing contemporary violence in the United States today (mass violence by Muslim extremists and anti-Muslim hate crimes, which are mutually reciprocal), it will be important in the future to extend these results to other groups. For example, it will be important to see if the degree to which marginalized groups like African Americans in the United States and the Roma in Europe are collectively blamed for violence and drug abuse/trafficking, and whether this perception can similarly be reduced through the collective blame hypocrisy approach.

Finally, it is worth acknowledging the variable strength of the evidence for the direct effects of the interventions on the outcome measures, and discussing the causal ordering implied by our model. The Collective Blame Hypocrisy video and interactive activity very consistently influenced levels of collective blame across all studies. The interventions also reduced prejudice in Studies 2b and 3b (but marginally in Study 2a and not in Study 3a) and dehumanization in Studies 3a and 3b (but marginally in Study 2b and not in Study 2a). On the contrary, despite the significant indirect effects of the hypocrisy intervention on anti-Muslim policy support and anti-Muslim behavior via collective blame and prejudice/dehumanization in all studies, only the Socratic activity (and not the video) exerted direct effects on these outcomes. The fact that our manipulation of collective blame had stronger direct effects on prejudice and dehumanization than on anti-Muslim policies and behavior is consistent with our model of attitudes and beliefs about Muslims as more proximal than policy support and behavior. At the same time, although we provided causal evidence that manipulating hypocrisy produced downstream effects on prejudice, dehumanization, and policy attitudes and behavior via collective blame, we did not explicitly test the causal relationship between prejudice/dehumanization and the policy attitudes and behavior. That part of our model therefore remains more tentative, requiring confirmatory research.

Conclusion

In sum, we established causal relationships between revealing the hypocrisy of collectively blaming some groups more than others for the acts of individuals and reducing collective blame of Muslims for individual terror attacks. In turn, we found across all studies that changes in collective blame mediated the relationship between intervention exposure and downstream policy support and behaviors associated with vicarious retribution, both directly and through collective blame's link to anti-Muslim prejudice and dehumanization. Taken together, our results highlight the importance of collective blame in intergroup conflict and show that making people aware of the hypocrisy inherent in blaming some groups more than others for the actions of individual outgroup members can mitigate collective blame and diminish downstream consequences associated with vicarious retribution.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Due to the high number of experimental conditions, we did not correct for multiple comparisons in Study 2a (which would have resulted in a highly conservative threshold). Instead, we opted to confirm any significant results with a replication study (Study 2b).
2. Included in the survey for other purposes, but not analyzed here, were measures of socioeconomic status, social dominance orientation (SDO), conservatism, media consumption, and support for 2016 American presidential candidates.
3. Examining the negative control versus controls yielded similar results, in the opposite direction.
4. Included in the survey for other purposes, but not analyzed further here, were measures of socioeconomic status, educational attainment, media consumption, and support for 2016 American presidential candidates.
5. Examining the negative control versus controls yielded similar results, in the opposite direction.
6. All subsequent analyses for this condition are similar for those who overestimated Muslim violence on one or all items, and those who did not overestimate on any.
7. Included in the survey for exploratory purposes, but not analyzed further here, were measures of socioeconomic status, educational attainment, employment, political conservatism, SDO, right-wing authoritarianism (RWA), and need for cognition (NFC).
8. Included in the survey for exploratory purposes, but not analyzed here, were measures of socioeconomic status, educational attainment, employment, conservatism, SDO, RWA, NFC, and support for 2016 American presidential candidates.

Supplemental Material

Supplementary material is available online with this article.

References

Allpress, J. A., Barlow, F. K., Brown, R., & Louis, W. R. (2010). Atoning for colonial injustices: Group-based shame and guilt motivate support for reparation. *International Journal of Conflict and Violence*, *4*(1), 75-88.

Armed Conflict Survey 2015. (2015). International Institute for Strategic Studies. Retrieved from https://www.iiss.org/en/publications/acs/by%20year/armed-conflict-survey-2015-46e5

Aronson, E., Fried, C., & Stone, J. (1991). Overcoming denial and increasing the intention to use condoms through the induction of hypocrisy. *American Journal of Public Health*, *81*, 1636-1638.

Bastian, B., & Haslam, N. (2010). Excluded from humanity: The dehumanizing effects of social ostracism. *Journal of Experimental Social Psychology*, *46*, 107-113.

Brown, R., González, R., Zagefka, H., Manzi, J., & Čehajić, S. (2008). Nuestra culpa: Collective guilt and shame as predictors of reparation for historical wrongdoing. *Journal of Personality and Social Psychology*, *94*, 75-90.

Bruneau, E. G., & Saxe, R. (2012). The power of being heard: The benefits of "perspective-giving" in the context of intergroup conflict. *Journal of Experimental Social Psychology*, *48*, 855-866.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3-5.

Čehajić-Clancy, S., Effron, D. A., Halperin, E., Liberman, V., & Ross, L. D. (2011). Affirmation, acknowledgment of in-group responsibility, group-based guilt, and support for reparative measures. *Journal of Personality and Social Psychology*, *101*, 256-270.

Chiu, C. Y., Morris, M. W., Hong, Y. Y., & Menon, T. (2000). Motivated cultural cognition: The impact of implicit cultural theories on dispositional attribution varies as a function of need for closure. *Journal of Personality and Social Psychology*, *78*, 247-259.

Cialdini, R. B. (2003). Crafting normative messages to protect the environment. *Current Directions in Psychological Science*, *12*, 105-109.

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*, 359-378.

Denson, T. F., Lickel, B., Curtis, M., Stenstrom, D. M., & Ames, D. R. (2006). The roles of entitativity and essentiality in judgments of collective responsibility. *Group Processes & Intergroup Relations*, *9*, 43-61.

Dickerson, C. A., Thibodeau, R., Aronson, E., & Miller, D. (1992). Using cognitive dissonance to encourage water conservation. *Journal of Applied Social Psychology*, *22*, 841-854.

Extremist Files: Westboro Baptist Church. (2016). Southern Poverty Law Center. Retrieved from https://www.splcenter.org/fighting-hate/extremist-files/group/westboro-baptist-church

Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford, CA: Stanford University Press.

Fointiat, V. (2004). "I know what I have to do, but . . ." when hypocrisy leads to behavioral change. *Social Behavior and Personality: An International Journal*, *32*, 741-746.

Fried, C. B., & Aronson, E. (1995). Hypocrisy, misattribution, and dissonance reduction. *Personality and Social Psychology Bulletin*, *21*, 925-933.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, *78*, 708-724.

Haddock, G., Zanna, M. P., & Esses, V. M. (1993). Assessing the structure of prejudicial attitudes: The case of attitudes toward homosexuals. *Journal of Personality and Social Psychology*, *65*, 1105-1118.

Hayes, A. F. (2012). *PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling* [White paper]. Retrieved from http://www.afhayes.com/public/process2012.pdf

Imhoff, R., & Recker, J. (2012). Differentiating Islamophobia: Introducing a new scale to measure Islamoprejudice and secular Islam critique. *Political Psychology*, *33*, 811-824.

Ingraham, C. (2015, February 11). Anti-Muslim hate crimes are still five times more common today than before 9/11. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/wonk/wp/2015/02/11/anti-muslim-hate-crimes-are-still-five-times-more-common-today-than-before-911/?utm_term=.5d4968a61871

J-PAL Policy Bulletin. (2012). *Deworming: A best buy for development*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab.

Kteily, N., & Bruneau, E. (2017). Backlash: The politics and real-world consequences of minority group dehumanization. *Personality and Social Psychology Bulletin*, *43*, 87-104.

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.

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.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., . . . Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*, 1765-1785.

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., . . . Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*, 1001-1016.

Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues*, *2*(4), 34-46.

Lickel, B., Miller, N., Stenstrom, D. M., Denson, T. F., & Schmader, T. (2006). Vicarious retribution: The role of collective blame in intergroup aggression. *Personality and Social Psychology Review*, *10*, 372-390.

Lickel, B., Schmader, T., & Barquissau, M. (2004). The distinction between collective guilt and collective shame. In N. R. Branscombe & B. Doosje (Eds.), *Collective guilt: International perspectives* (pp. 35-55.). Cambridge, UK: Cambridge University Press.

Lickel, B., Schmader, T., & Hamilton, D. L. (2003). A case of collective responsibility: Who else was to blame for the Columbine High School shootings? *Personality and Social Psychology Bulletin*, *29*, 194-204.

Liégeois, A., Yserbyt, V., & Corneille, O. (2005). *I'm dirty as anyone else . . . so what? When attempts at inducing hypocrisy backfire*. Annual meeting of the Belgian Association for Psychological Sciences (BAPS), *Ghent University, Belgium*.

Manchi Chao, M., Zhang, Z. X., & Chiu, C. Y. (2008). Personal and collective culpability judgment: A functional analysis of East Asian–North American differences. *Journal of Cross-cultural Psychology*, *39*, 730-744.

McDonald, R. I., & Crandall, C. S. (2015). Social norms and social influence. *Current Opinion in Behavioral Sciences*, *3*, 147-151.

Menon, T., Morris, M. W., Chiu, C. Y., & Hong, Y. Y. (1999). Culture and the construal of agency: Attribution to individual versus group dispositions. *Journal of Personality and Social Psychology*, *76*, 701-717.

Noar, S. M. (2006). A 10-year retrospective of research in health mass media campaigns: Where do we go from here? *Journal of Health Communication*, *11*(1), 21-42.

Singh, R., Simons, J. J., Self, W. T., Tetlock, P. E., Zemba, Y., Yamaguchi, S., . . . Kaur, S. (2012). Association, culture, and collective imprisonment: Tests of a two-route causal-moral model. *Basic and Applied Social Psychology*, *34*, 269-277.

Stenstrom, D. M., Lickel, B., Denson, T. F., & Miller, N. (2008). The roles of ingroup identification and outgroup entitativity in intergroup retribution. *Personality and Social Psychology Bulletin*, *34*, 1570-1582.

Stone, J., Aronson, E., Crain, A. L., Winslow, M. P., & Fried, C. B. (1994). Inducing hypocrisy as a means of encouraging young adults to use condoms. *Personality and Social Psychology Bulletin*, *20*, 116-128.

Stone, J., & Fernandez, N. C. (2008). To practice what we preach: The use of hypocrisy and cognitive dissonance to motivate behavior change. *Social and Personality Psychology Compass*, *2*, 1024-1051.

Tharoor, I. (2015, November 17). What Americans thought of Jewish refugees on the eve of World War II. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/worldviews/wp/2015/11/17/what-americans-thought-of-jewish-refugees-on-the-eve-of-world-war-ii/?utm_term=.1331a0c693d4

Wilson, S., Miller, G., & Horwitz, S. (2013, April 23). Boston bombing suspect cites US wars as motivation, officials say. *The Washington Post*. Retrieved from https://www.washingtonpost.com/national/boston-bombing-suspect-cites-us-wars-as-motivation-officials-say/2013/04/23/324b9cea-ac29-11e2-b6fd-ba6f5f26d70e_story.html?utm_term=.c5ccd6c80888

Zemba, Y., Young, M. J., & Morris, M. W. (2006). Blaming leaders for organizational accidents: Proxy logic in collective- versus individual-agency cultures. *Organizational Behavior and Human Decision Processes*, *101*, 36-51.