

# Threat to the Group's Image Can Motivate High Identifiers to Take Action Against In-group Transgressions

Eric Shuman<sup>1</sup> , Dan Johnson<sup>2</sup>, Tamar Saguy<sup>1</sup>, and Eran Halperin<sup>1</sup>

Personality and Social  
Psychology Bulletin  
1–22

© 2018 by the Society for Personality  
and Social Psychology, Inc  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0146167218768800  
pspb.sagepub.com



## Abstract

When transgressions are committed by a group, those highly identified with the group are often least likely to recognize the transgressions, feel collective guilt, and engage in action to address them. We hypothesized that especially among high identifiers, demonstrating that in-group transgressions threaten the group's image can induce normative conflict and thus collective guilt and action. In the first study, we demonstrate that high (vs. low) image threat increases normative conflict among high identifiers. In Study 2, we show that inducing normative conflict through image threat leads to increased collective guilt and collective action among high identifiers. In Study 3, we replicate this effect with the addition of a control condition to demonstrate increased normative conflict and collective guilt relative to both a low threat and baseline conditions. In Study 4, we again replicate these effects with a modified manipulation that more precisely manipulated image threat. Together, these studies indicate that image threat can be an effective motivator for high identifiers to address in-group transgressions.

## Keywords

image threat, group identification, normative conflict, collective guilt, collective action

Received June 24, 2017; revision accepted February 20, 2018

When atrocities have been committed by one's group against another, reestablishing positive intergroup relations is an essential, but challenging, process. Often, this process requires that group members from the offending group take action to correct these transgressions. Critical emotions toward one's group, such as collective guilt, can play an important role in motivating this kind of action (Doosje, Branscombe, Spears, & Manstead, 1998, 2004; Wohl & Branscombe, 2008). However, the experience of collective guilt is relatively rare (Iyer, Leach, & Pedersen, 2004; Wohl, Branscombe, & Klar, 2006), in part because it can conflict with group members' motivation to maintain a positive group identity (Branscombe & Miron, 2004; Tajfel & Turner, 1979). Thus, the question of how to motivate group members, particularly those who are strongly identified with their group, to feel collective guilt and take action against transgressions committed by their group is of great theoretical and applied value because high identifiers often have more authority and influence within a group and thus can be more influential on other group members, and their criticism of group actions is often perceived more positively (Hornsey, Grice, Jetten, Paulsen, & Callan, 2007; Hornsey, Trembath, & Gunthorpe, 2004; Wirtz & Doosje, 2013). Thus, high

identifiers may have potential to be more effective at correcting group transgressions.

## In-group Transgressions, Collective Guilt, and Identification

In-group transgressions are actions undertaken by the in-group that are harmful to another group (Branscombe et al., 1999). Research on psychological factors that can motivate group members to take action to correct these transgressions has focused primarily on collective guilt, although collective shame and in-group-directed anger have also been considered as motivators of action (Gunn & Wilson, 2011; Iyer, Schmader, & Lickel, 2007). Collective guilt is experienced when one perceives the in-group as having violated a moral standard and as responsible for harming another group. For

<sup>1</sup>Interdisciplinary Center Herzliya, Israel

<sup>2</sup>Washington and Lee University, Lexington, VA, USA

### Corresponding Author:

Eric Shuman, The Baruch Ivcher School of Psychology, Interdisciplinary Center Herzliya, Kanfei Nesharim Street, Herzliya 46150, Israel.  
Email: eric.shuman@idc.ac.il

example, researchers have found that collective guilt can be experienced over past misdeeds of the group such as colonialism (Doosje et al., 1998), as well as current ongoing transgressions, such as harm to innocent civilians caused by the Iraq War (Iyer et al., 2007; Wohl & Branscombe, 2008). Unlike personal guilt, collective guilt is experienced for deeds that one did not directly commit, but because one is associated with the perpetrators by group membership (Branscombe & Miron, 2004; Doosje et al., 1998; Wohl & Branscombe, 2008).

Given the vicarious nature of collective guilt, it is generally predicated on some level of identification with the group (Doosje et al., 1998; Mackie, Devos, & Smith, 2000; Roccas, Klar, & Livitain, 2006). In other words, individuals need to identify at some level as members of the perpetrating group to feel morally implicated by transgressions committed by a group or its representatives. At the same time, strong identification with the in-group also makes recognition of the in-group's negative actions *less* likely because of one's need to maintain a positive group identity, thus decreasing collective guilt (Branscombe, Slugoski, & Kappen, 2004; Tajfel & Turner, 1979).

Indeed, prior research has found that high identifiers often react defensively to information regarding in-group transgressions, both when it is stated as objective information (Branscombe, 1998) and when confronted with these transgressions by the victimized group (Peetz, Gunn, & Wilson, 2010; Sahdra & Ross, 2007). A large body of research has shown that high identifiers can often use a number of psychological mechanisms to justify, deny, or distance themselves from group wrongdoing to protect their positive identity (see Branscombe & Miron, 2004; Knowles, Lowery, Chow, & Unzueta, 2014 for reviews). Thus, this need to protect the group identity often leads high identifiers to avoid recognition of their group's wrongdoing and thus to express less willingness to correct it, highlighting the need for interventions capable of overcoming or redirecting this motivation to induce collective guilt among high identifiers.

### **When Do Highly Identified Group Members Challenge Group Actions?**

In the current research, we propose that under the right circumstances, this same motivation to maintain a positive group image can be leveraged to lead high identifiers to challenge (rather than justify) in-group transgressions. To understand how to motivate high identifiers to challenge the actions of their group, we drew from intragroup perspectives on dissent. Most research on collective guilt has focused on the intergroup nature of transgressions, that is, how it affects relations between the victim and perpetrator groups. However, transgressions also cause intragroup concerns, that is, how it might affect one's own group. For example, Iyer et al. (2007) showed that portraying transgression as reflecting the group's character can shift the focus of members'

responses to questions of collective self-definition and image management, that is, more intragroup concerns. In addition, Sullivan, Landau, Branscombe, Rothschild, and Cronin (2013) found that focusing on how in-group transgressions caused direct harm to the in-group increased collective guilt, because harm to the self was more difficult to legitimize. Although Iyer et al. (2007) and Sullivan et al. (2013) did not focus on the role of identification, we propose that this perspective and intragroup perspectives more generally might be particularly informative for understanding how to motivate high identifiers to correct transgressions.

Although research on dissent within groups does not usually address collective guilt, it does consider a key process that can lead to guilt: when group members will challenge and criticize actions of the group (see Jetten & Hornsey, 2014 for a review). Interestingly, this literature finds that high identifiers are often more willing to engage in dissent and criticism because they are more invested in the group (Packer, 2008; Packer & Chasteen, 2010). High identifiers also may feel more secure in their ability to criticize without paying a high social cost (Hornsey et al., 2004) and might also care more about correcting the group's wrongdoing (given its central part in their self-concept). These findings suggest that under the right circumstances, high identifiers could feel collective guilt, possibly be even more motivated to engage in action to correct group transgressions. However, insights from this literature have not yet been applied to work on collective guilt.

According to Packer's normative model of dissent (2008), high identified group members will take action to challenge the actions of their group when they perceive these actions to be in conflict with some normative standard of behavior, usually the group's own norms (Packer, 2008; Packer & Chasteen, 2010). When group members feel the group has taken actions that violate its central norms and values, they will be motivated to criticize these actions and attempt to change the group's behavior. From this perspective, the question then becomes how to induce normative conflict regarding in-group transgressions among high identifiers, even when high identifiers tend to deny or justify group transgressions.

Under most circumstances, high identifiers are more likely to support and abide by the actions and decisions of the group. In keeping with this, research has shown that strongly identified group members are only likely to dissent in response to a particular form of normative conflict, specifically when they see group norms or actions are harmful to the group itself (Packer & Chasteen, 2010). Because high identifiers are particularly invested in the group and care about the good of the group, this perception of harm leads them to challenge whether these actions are really in line with the groups norms and take action to correct these actions to protect the group. Thus, high identifiers may feel that the group is harming its own image or interests precisely because it is failing to live up to its own standards and values (e.g.,

Sani & Reicher, 1998, 1999; Sani & Todman, 2002) and thus take action to correct these harmful actions.

### Group Moral Image Threat

Drawing on these ideas, we propose that the potential threat of transgressions to the moral image of the group can lead highly identified group members to see transgressions as harming the group, thus inducing normative conflict and criticism. Moral image threat involves concern over wrongdoings committed by the in-group against another group that might damage the in-group's reputation as good and moral (Cottrell & Neuberg, 2005; Ellemers, Spears, & Doosje, 2002). High power group members are particularly sensitive to threats to their moral image. According to the needs-based model of reconciliation (Shanbel & Nadler, 2008), advantaged or perpetrator groups are particularly concerned about whether they are perceived as moral by other groups. This can stem both from basic needs for social acceptance and from the perception that the group's privileged status in part depends on the group being perceived as moral (Mashuri, van Leeuwen, & Hanurawan, 2016). As a result, those highly identified with high power groups are particularly motivated to maintain their moral image in the eyes of other groups because the loss of their moral image could lead to exclusion of their group or potential loss of their privileged status. In the international sphere, this could take the form of exclusion from the United Nations or other higher level international bodies or agreements or in the specific case of the U.S. loss of support for the current, U.S.-led international order.

Therefore, our perspective considers intragroup concerns that stem from a threat that has to do with how *other* third-party groups, not the transgressed group, view one's group. Because the image of the group ultimately rests on the perceptions of others, we suggest that moral image threat is less susceptible to psychological mechanisms usually used to deal with the threatening aspects of in-group transgressions. Prior research has shown that high identifiers often use psychological mechanisms such as rationalization, moral disengagement, or moral justification to reduce the threat of transgressions to their personal identity (see Leidner, Castano, Zaiser, & Giner-Sorolla, 2010; Roccas et al., 2006; Tarrant, Branscombe, Warner, & Weston, 2012). These mechanisms usually lead to a lack of guilt among high identifiers, but we suggest that these mechanisms are less capable of reducing a threat to the group's moral image, and thus, image threat can raise levels of guilt among high identifiers. For example, denying the existence of the transgression or internally rationalizing it does little to repair the image of the group in the eyes of others (Knowles et al., 2014). Thus, to restore their moral image in the eyes of other groups, group members may need to take action to change how they are perceived.

Although this hypothesis has not been directly tested, there is research that indirectly suggests that group members

will take action in an attempt to correct their external image. For example, while White Americans are usually less supportive of affirmative action when it is framed as hurting the in-group, Lowery, Chow, Knowles, and Unzueta (2012) showed that this pattern could be reversed by framing inequality as the result of White privilege. In the privilege condition, Whites were especially supportive of a policy described as hurting the White in-group, indicating that to reduce the threat of privilege to the group's positive esteem, Whites were even willing to support policies costly to their group. In addition, research from the field of meta-stereotypes supports the idea that individuals are concerned with how other groups think of their group and are even willing to take action to change these perceptions. Meta-stereotypes are a person's beliefs regarding how out-group members perceive his or her own group (Vorauer, Main, & O'Connell, 1998). Research on meta-stereotypes has found that high power group members are particularly concerned with perceptions of morality (Shanbel & Nadler, 2008; Vorauer & Sakamoto, 2008) that they will engage in costly helping behavior when a negative meta-stereotype is induced (Hopkins et al., 2007), and this effect is mediated by concern for the group's image (van Leeuwen & Täuber, 2012). This provides evidence that concerns with how one's group is perceived by others are linked to action to correct these perceptions. However, this body of research has not demonstrated that image threat-related concerns are capable of motivating collective guilt and collective action in the context of in-group transgressions when the group is under threat and there are strong motivations for group members to avoid collective guilt and corrective action.

In sum, we propose that in-group transgressions that are presented as highly threatening to the group's moral image can lead high identifiers to see the transgression as harmful to the group and thus induce normative conflict. Drawing from the normative conflict model, we suggest that normative conflict, and thus collective guilt, can be induced by focusing on harm done to the in-group and its image. Although research on collective guilt often focuses on how appraisals of responsibility for harm to the out-group induce guilt (Doosje et al., 1998; Lickel, Schmader, Curtis, Scarnier, & Ames, 2005), we suggest guilt can also be triggered by concern from harm to the in-group's image via normative conflict. As high identifiers are particularly motivated to maintain a positive group image (Shanbel & Nadler, 2008; Tajfel & Turner, 1979), they should see this threat to the group's image as harmful to the group. This should induce normative conflict because high identifiers are most likely to feel normative conflict when they see the group's actions as harmful to the group (Packer, 2008). Normative conflict in turn should help facilitate collective guilt because it is essentially an appraisal of the action as a violation of norms, specifically the group's norms. Although there are other appraisals that need to be present in order for guilt to occur, namely, responsibility and harm, norm violation is a core

appraisal of collective guilt, so activating this appraisal by inducing normative conflict should help facilitate the experience of collective guilt. Collective guilt should then drive action to correct the in-group's transgression. Altogether, this makes inducing image threat an effective way to motivate high identifiers to take action against in-group transgressions by inducing normative conflict.

This expands on the current literature in a number of important ways. First, it suggests a novel theory-based intervention for inducing collective guilt among those highly identified with the group and thus, motivating them to take action to correct in-group transgressions. Because of their greater influence and authority within a group, motivating high identifiers maybe particularly important for correcting transgressions. Second, it proposes group image threat as new mechanism to induce collective guilt and motivate action among high identifiers. Although this is based on research on meta-stereotypes, which shows that group members will take action to change their group's image, this research was not conducted in the context of intergroup transgressions, where the group members are often motivated to justify the transgression or where the transgression may have been committed in pursuit of another group goal. In addition, while Lowery et al. (2012) found that inequality framed as in-group advantage threatens group members' esteem for their group and thus motivates support for affirmative action, however, it did not investigate whether a similar concern for the esteem of the group could be triggered by a threat to the group's image and did not examine whether this could motivate collective action behavior rather than just policy support. Third, it integrates the normative conflict model of dissent and other more intragroup perspectives with the more intergroup focused literature on collective guilt.

## The Current Study

To test these ideas, we conducted four studies in the context of the American War on Terror. Specifically, we chose the issue of the Guantanamo Bay detention center, as it has been the focus of a long-standing debate about American conduct in the War on Terror. Furthermore, it could be seen as violating important American norms, such as the right to due process and a speedy trial as well as humane treatment of prisoners. Therefore, it was an ideal issue to examine whether image threat could increase perceptions of normative conflict. Initially, we manipulated external image threat based on methods modified from those used previously by Iyer et al. (2007). In the low image threat condition, transgressions that occurred at Guantanamo were presented, but a situational attribution was made for the transgressions (external attribution). In the high image threat condition, participants were told that members of an external group (Europeans) felt that these same transgressions reflected Americans' immoral character (internal attribution). According to attribution theory (Heider, 1958),

internal attributions should be more threatening because they more deeply link the transgression to the nature and intentions of the group. We chose Europeans as an external, third-party group, rather than the group that was transgressed to deliver the criticism, because we wanted participants to remain in a more intragroup mind-set, that is, focused on managing their identity and how it is perceived rather than specifically relations with the victimized group.

We hypothesized that for high identifiers, the high image threat condition will cause higher levels of normative conflict compared to the low image threat condition, thus leading to more collective guilt and collective action. Although they are not the focus of this research, we predict that low identifiers will not be highly affected by the manipulation of image threat, because they are less concerned with the group's image and overall will exhibit higher levels of normative conflict, collective guilt, and collective action and they are not as strongly identified with the group and thus exhibit less defensiveness. Therefore, merely reading about an in-group transgression (whether it is image threatening or not) should lead to normative conflict, collective guilt, and action.

In the first study, we aimed to validate that manipulation induced normative conflict among high identifiers. In the second and third studies, we aimed to replicate the effect on normative conflict, while examining its downstream effects on collective guilt and collective action aimed at correcting the in-group transgression. In the fourth study, we aimed to address some possible confounds in the manipulation and again replicate the hypothesized effects. We further assessed additional emotions (collective shame and in-group-directed anger) relevant for the image threat and for collective action.

## Study 1

This study aimed to provide an initial test of our hypothesis that a manipulation of image threat would induce normative conflict among high identifiers. This step is critical given that no prior work has examined the connection between image threat and normative conflict.

## Method

**Participants and procedure.** Through Amazon's Mechanical Turk (MTurk), 161 participants volunteered to participate in this study (this study was conducted as a part of an undergraduate research project, and thus, the sample size was determined by budgetary constraints). Participants were paid 50 cents for completing the study. We excluded 10 participants who did not take the study seriously (indicated by the fact that spent less than 30 s reading the manipulation article and answered two questions they were asked to leave blank) from analysis. This left a sample of 152 participants (84 women,  $M_{age} = 38.68$  years, 124 Caucasian, 15 African American, and 12 other). Participants first completed the scale measuring identification. They were then randomly

assigned to read either the high or low image threat article, which was followed by scales measuring group image threat, normative conflict, and demographics.

#### *Materials and measures*

**Group Identification Scale.** This scale was adapted from Roccas et al. (2006), changing “Israeli” and “Israel” to “American” and “America.” Eight items measured attachment and eight items measured glorification, because all results followed the same pattern for glorification and attachment we combined them into a single scale measuring identification ( $\alpha = .95$ ); we also report results for each component separately (in “Notes” section) although factor analysis did not support the glorification–attachment structure. On this and all other scales, participants indicated their agreement with each of the items on a 7-point Likert-type scale ranging from “strongly disagree” (1) to “strongly agree” (7). Sample items for attachment and glorification are, respectively, “Being American is an important part of my identity” and “America is better than other nations in all respects.”

**Group image threat manipulation.** Participants read an approximately 650-word article about the Guantanamo Bay Detention Center and examples of physical abuse of prisoners committed by American soldiers. The first section of the article was created from excerpts from a report published by the Center for Constitutional Rights (2006) about the abuses that occurred at Guantanamo. It described how inmates at Guantanamo have been detained indefinitely without legal rights, as well as types of abuse that have occurred at Guantanamo (physical, psychological, etc.); it also described an unnecessary beating of a prisoner as a specific example of abuse. This specific example was included because specific examples are more convincing than general descriptions (Guadagno, Rhoads, & Sagarin, 2011), and thus, we predicted that this would increase the potential of the manipulation to induce collective guilt. The second section was designed to threaten participants’ group image based on the methods of Iyer et al. (2007). This paragraph highlighted the negative responses of Europeans to Guantanamo Bay. In the high image threat condition, Europeans made stable internal attributions for the abuse, saying, “The abuse in Guantanamo and other American policies reveal the imperialistic and unjust character of Americans.” In the low image threat<sup>1</sup> condition, Europeans made situational attributions for abuse, citing the effects of the war on terror, saying, “It’s unfortunate that Americans have been forced to commit these actions because of war, but I think Americans are generally good people. In the end, it is just their actions in this war that are unjust.”

Unlike Iyer et al. (2007), we did not use the out-group that the in-group transgressed against to deliver the criticism. Rather, we used an uninvolved third-party group to deliver the image threat. We suggest that image threat from another group is more likely to be perceived as harmful to the in-group, increasing the likelihood that it would induce normative conflict and thus

collective guilt. Specifically, we chose Europeans for a number of reasons. First, the Americans are not currently in any serious conflict with Europeans and they are generally seen as allies, and thus, their opinion of the U.S. is likely to carry some weight. In addition, Americans do not generally hold significant negative stereotypes of Europeans (at least in comparison to Arabs, Africans, etc.), which could lead to discounting of their opinion. Finally, we chose an aggregate group (i.e., Europeans) rather than a specific national group (e.g., British) because we felt this would pose a more significant image threat.

**Group image threat.** Two items (“While reading the article, I felt that the violence committed against prisoners in Guantanamo makes Americans look bad” and “While reading the article, I felt that that the violence committed against prisoners in Guantanamo have led other people to have a negative view of my country”) served as a manipulation check to ensure that the two articles produced differing levels of image threat ( $r = .67$ ).

**Normative Conflict Scale.** Three questions assessed appraisals of normative conflict ( $\alpha = .85$ ): “While reading the article, I felt that the violence committed against prisoners in Guantanamo was a violation of American moral standards”; “While reading the article, I felt that the violence committed against prisoners in Guantanamo were a violation of American ethical norms”; and “While reading the article, I felt that the violence committed against prisoners in Guantanamo are compatible with American moral values” (reverse scored).

**Demographic Questionnaire.** Participants completed a brief demographic questionnaire. Items included gender, age, education, ethnicity, employment, political ideology, and political party affiliation.

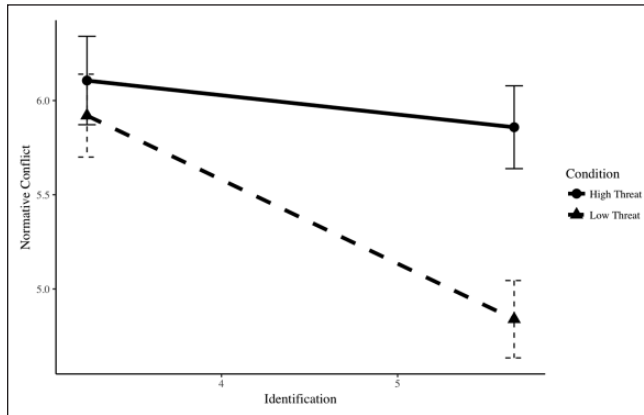
## *Results and Discussion*

Means, standard deviations, and correlations among all variables are presented in Table 1. The manipulation produced the expected differences in image threat: Participants in the high threat condition felt significantly more image threat than participants in the low image threat condition,  $t(150) = 3.71, p < .001, d = 0.61$ . We then tested our hypothesis that the high image threat condition would lead to increased normative conflict only when identification was high, using Hayes’s (2013) PROCESS command with 5,000 iterations (Model 1). Taking into account the interaction, there was a significant main effect of identification,  $b = -0.27, SE = 0.08, t = -3.31, p = .001, 95\% \text{ confidence interval (CI)} = [-0.44, -0.11]$ , and condition ( $b = 0.60, SE = 0.20, t = 2.97, p = .003, CI = [0.20, 1.00]$ ). Furthermore, the two-way interaction was significant ( $b = 0.34, SE = 0.16, t = 2.05, p = .04, CI = [0.01, 0.67]$ ). Analysis of the simple effects revealed that the image threat condition had a significant effect on normative conflict only among high

**Table 1.** Bivariate Relationships Between Study 1 Variables ( $N = 152$ ).

	$M (SD)$		1	2	3	4
1. Image threat condition	Low threat	High threat	—			
2. Political ideology	3.25 (1.89)	3.47 (1.92)	.06	—		
3. Identification	4.54 (1.19)	4.35 (1.24)	-.08	.31**	—	
4. Group image threat	5.73 (1.14)	6.35 (0.87)	.29**	-.38**	-.34**	—
5. Normative conflict	5.35 (1.40)	6.00 (1.15)	.25**	-.35**	-.27**	.67**

\* $p < .05$ . \*\* $p < .01$ .



**Figure 1.** Interaction between condition and identification on normative conflict in Study 1, points are displayed at 1  $SD$  above and below the mean.

identifiers (1  $SD$  above the mean;  $b = 1.01$ ,  $SE = 0.29$ ,  $t = 3.55$ ,  $p < .001$ ,  $CI = [0.45, 1.57]$ ; see Figure 1).<sup>2</sup> For those low on identification, normative conflict was high in both conditions indicating that they appraised the in-group transgression as a normative violation regardless of whether it was threatening to their identity. This interaction remained significant when controlling for political ideology. This supported our hypothesis that the image threat manipulation would induce a sense of normative conflict among those who highly identify with the group. Although normative conflict did increase as predicted from the low threat to high threat conditions for high identifiers, it did not exceed that of low identifiers. As we discuss in the introduction, while image threat should be particularly effective for inducing normative conflict for high identifiers, there are other forces, such as the motivation to justify the group's actions that work in the opposite direction. Thus, we did not expect normative conflict among high identifiers to be greater than among low identifiers in the high threat condition.

## Study 2

The first study supported our hypothesis that an image threat manipulation would increase levels of normative conflict,

but only when identification was high. Therefore, we moved on to test our full research model. We hypothesized that the high image threat condition would lead to increased normative conflict, which would then lead to higher levels of collective guilt and thus collective action. Importantly, we expected this process to occur particularly at high levels of identification.

## Method

**Participants and procedure.** Through Amazon's MTurk, 176 participants were recruited for this study. Participants were paid 50 cents for completing the study (this study was conducted as a part of an undergraduate research project, and thus, the sample size was determined by budgetary constraints). Fourteen participants were excluded because they did not pay sufficient attention to the study (indicated by the fact that spent less than 30 s reading the manipulation article and answered two questions they were asked to leave blank) from analysis. This left a sample of 162 participants (82 women,  $M_{age} = 37.99$  years, 86.4% White, 6.2% African American, and 7.4% other). This study was identical to Study 1, except that we added measures of collective guilt, collective shame, in-group-directed anger, and collective action. Although comparing the effects of the different group, critical emotions were not the focus of this study; we wanted to measure them in order to control for their effects nonetheless.

**Materials and measures.** The measures of identification ( $\alpha = .94$ ), image threat ( $\alpha = .85$ ), and normative conflict ( $\alpha = .92$ ) were the same as those used in Study 1. Participants indicated their agreement to the items of all scales on a 7-point Likert-type scale ranging from "strongly disagree" (1) to "strongly agree" (7), unless stated otherwise.

**Collective guilt.** A modified version of Branscombe et al.'s (2004) Collective Guilt Scale was used to measure collective guilt. This scale is phrased in general terms, so again the statements were adapted to be specific to the context under study, for example, "I feel regret for Americans' harmful actions toward prisoners of Guantanamo" ( $\alpha = .92$ ).

**Table 2.** Bivariate Relationships Between Study 2 Variables ( $N = 162$ ).

	M (SD)		1	2	3	4	5	6
1. Image threat condition	Low threat	High threat	—					
2. Political ideology	3.60 (1.95)	3.16 (1.84)	-.11	—				
3. Identification	4.01 (1.10)	4.15 (1.05)	.03	.08	—			
4. Image threat	5.70 (1.39)	5.88 (1.16)	.07	-.10	-.23**	—		
5. Normative conflict	5.60 (1.53)	5.82 (1.26)	.08	-.19*	-.13	.68**	—	
6. Collective guilt	3.96 (1.52)	4.19 (1.72)	.07	-.18*	-.17*	.57**	.53**	—
7. Collective action	2.57 (2.36)	3.05 (2.69)	.09	-.18*	-.14	.37**	.38**	.42**

\* $p < .05$ . \*\* $p < .01$ .

*In-group-directed anger.* Participants rated the degree to which they felt “hostile,” “upset,” and “angry” about the violence committed against prisoners at Guantanamo and toward Americans ( $\alpha = .93$ ).

*Collective shame.* A modified version of Branscombe et al.’s (2004) Collective Shame Scale was used. Although the Branscombe et al. (2004) scale is phrased in general terms, the statements were adapted to be specific to the context under study, for example, “I feel ashamed of the violence committed against prisoners in Guantanamo by Americans” ( $\alpha = .96$ ).

*Collective action.* Participants were given the option to engage in collective action by completing an optional part of the study. Participants were told that because of the political nature of the study, the researchers wanted to present some data to U.S. Congressmen in the hope that it would affect the policy decisions they make. They were then told that this section of the survey was completely optional and given the option to skip to the next section of the survey or to complete the optional portion. First, participants were asked to report whether they supported closing Guantanamo. Then, they were given the opportunity to write a letter to their congressional representatives expressing their views on the issue. Finally, they were given the option of providing their email to an organization (Human Rights Watch) working to close Guantanamo. Next, they were given the opportunity to complete the same three types of action regarding the issue of whether Guantanamo detainees should be granted full legal rights.

For analysis, collective action was scored by giving each participant one point for each piece of the collective action survey they completed. In other words, they were given one point for choosing to complete the optional survey, one point for providing their opinion, one point if they wrote a letter to their congressional representatives, and one point if they provided their email address. This was done for both the issues of closing Guantanamo and granting full legal rights to its detainees. There were 10 people who indicated that they wanted to participate in the optional survey, but did not report their opinion, write a letter, or provide their email.

Because they expressed some interest in engaging in action, we gave them a higher score (1) than those who skipped the survey entirely (0); however, coding these both as zero does not change the results. Thus, the scoring ranged from 0 (skipped the survey) to 1 (agreed to complete optional survey) and then 2 to 7 (depending on what other actions the participant engaged in).

## Results and Discussion

Table 2 presents means, standard deviations, and correlations among all variables. We first examined whether there was a significant interaction between the image threat condition and identification on all main study variables using Hayes’s (2013) PROCESS command with 5,000 iterations (Model 1). We report the statistics for the interactions as well as the main effects in Table 3. As indicated in Table 3, there were significant interactions between the image threat condition and identification on normative conflict, collective guilt, collective shame, and in-group-directed anger; additionally, interaction on collective action was marginally significant. A simple-effects analysis (see Table 4) revealed that all interactions followed the same pattern: The image threat condition affected levels of the dependent variables only among those high on group identification (1 *SD* above the mean; see Figures 2-5),<sup>3</sup> indicating greater critical emotions and somewhat greater collective action intentions among high identifiers who were exposed to the high (vs. low) image threat. These interactions remained significant even if we controlled for political ideology. Again, the manipulation did not have a significant impact on low identifiers.

*Path analysis.* Because there is no process model that tests serial mediated moderation, we tested our full model, including normative conflict, collective guilt, and collective action, using SPSS AMOS 6 (see Figure 5). Overall, the model fit was good,  $\chi^2 (n = 162, df = 6) = 9.28, p = .16$ , comparative fit index (CFI) = .96, root mean square error approximation (RMSEA) = .058 (see Little, 2013 for criteria for indices of model fit). The model revealed that (a) normative conflict fully mediated the relationship between the interaction of condition and identification on collective

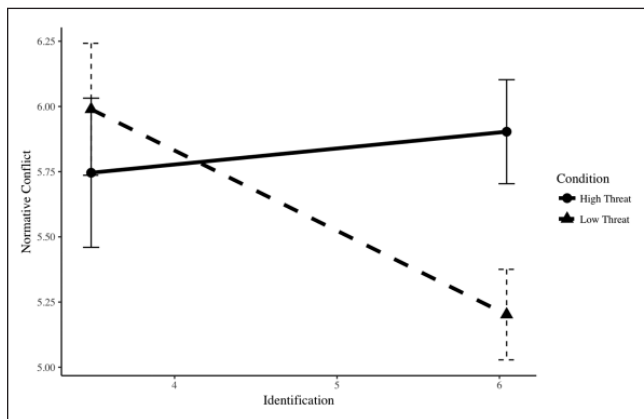
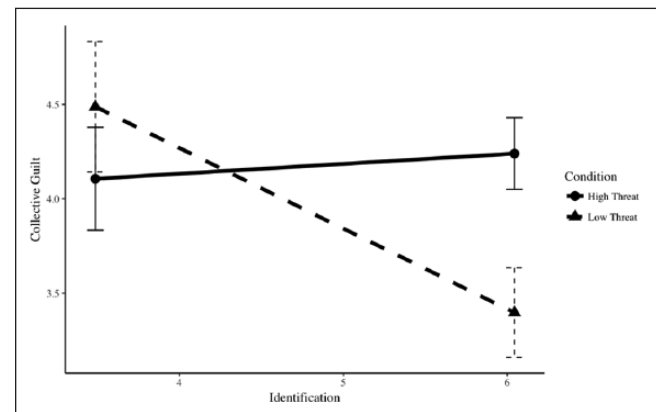
**Table 3.** Significant Effects on Dependent Variables in Study 2.

	Identification				Condition				Interaction			
	<i>b</i>	<i>SE</i>	<i>T</i>	<i>P</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Normative conflict	-0.31*	0.11	-2.72	.001	-1.53 <sup>†</sup>	0.85	-1.80	.07	0.36*	0.18	2.14	.03
Collective guilt	-0.43*	0.13	-3.29	.001	-2.05*	0.97	-2.11	.04	0.48*	0.20	2.43	.02
Collective shame	-0.54*	0.14	-3.79	.001	-2.06 <sup>†</sup>	1.07	-1.92	.06	0.50*	0.22	2.33	.02
In-group anger	-0.57*	0.11	-5.25	<.001	-1.35 <sup>†</sup>	0.81	-1.67	.10	0.33*	0.16	2.04	.04
Collective action	-0.53*	0.21	-2.58	.01	-2.18	1.54	-1.41	.16	0.55 <sup>†</sup>	0.31	1.76	.07

<sup>†</sup>*p* < .10. \**p* < .05.

**Table 4.** Simple Effects and Means for the Dependent Variables in Study 2.

	Low Identification (-1 SD)			High Identification (+1 SD)		
	Low threat	High threat		Low threat	High threat	
Normative conflict	5.99	5.75	<i>b</i> = -0.24, <i>ns</i>	5.20	5.90	<i>b</i> = 0.70, <i>SE</i> = 0.31, <i>t</i> = 2.26, <i>p</i> = .02
Collective guilt	4.49	4.11	<i>b</i> = -0.38, <i>ns</i>	3.40	4.24	<i>b</i> = 0.84, <i>SE</i> = 0.35, <i>t</i> = 2.38, <i>p</i> = .02
Collective shame	5.40	5.09	<i>b</i> = -0.30, <i>ns</i>	4.01	5.00	<i>b</i> = 0.99, <i>SE</i> = 0.39, <i>t</i> = 2.55, <i>p</i> = .01
In-group anger	5.11	4.92	<i>b</i> = -0.19, <i>ns</i>	3.66	4.33	<i>b</i> = 0.67, <i>SE</i> = 0.29, <i>t</i> = 2.28, <i>p</i> = .02
Collective action	3.22	2.97	<i>b</i> = -0.26, <i>ns</i>	1.88	3.03	<i>b</i> = 1.16, <i>SE</i> = 0.56, <i>t</i> = 2.06, <i>p</i> = .04

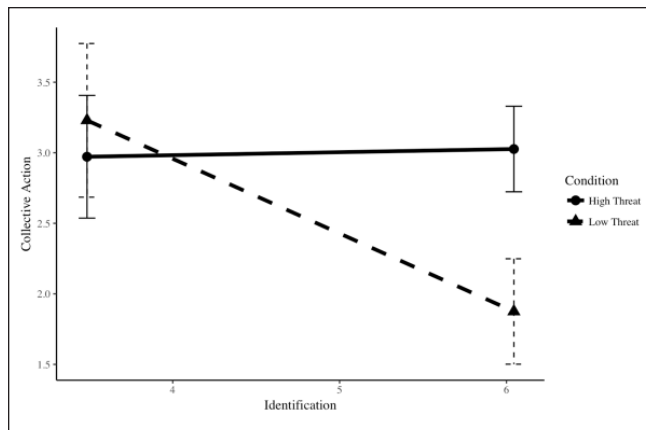
**Figure 2.** Interaction between condition and identification on normative conflict in Study 2, points are displayed at 1 SD above and below the mean.**Figure 3.** Interaction between condition and identification on collective guilt in Study 2, points are displayed at 1 SD above and below the mean.

guilt and (b) collective guilt and normative conflict both had significant effects on collective action, and collective guilt partially mediated the effect of normative conflict on collective action. We also tested an alternative model where we reversed the casual order of the dependent variables, such that collective action predicted collective guilt, which in turn predicted normative conflict. This model did not fit the data as well,  $\chi^2$  ( $df = 7$ ) = 14.22,  $p = .04$ ,  $CFI = .92$ ,  $RMSEA = .08$ , and comparison of the chi-square values revealed that the change in chi-square values between the two models was significant,  $\Delta\chi^2$  ( $df$ ) = 4.94 (1),  $p < .05$ . We also tested a model where we entered in-group-directed

anger and collective shame as additional predictors of collective action. This model also fits the data well,  $\chi^2$  ( $n = 162$ ,  $df = 10$ ) = 14.65,  $p = .15$ ,  $CFI = .99$ ,  $RMSEA = .054$ , but only collective guilt and normative conflict were significant predictors of collective action.

These results support our hypothesis that a threat to the group's image can induce collective guilt and thus motivate collective action among those highly identified with the group. In addition, it supports the mediating role of normative conflict in this process, indicating that perception of harm to the group's image leads them to challenge whether these actions are really in line with the group's standards and





**Figure 4.** Interaction between condition and identification on collective action in Study 2, points are displayed at 1 SD above and below the mean.

norms and take action to correct these actions in order to protect the group.

### Study 3

The first two studies supported our hypothesis that inducing threat to the group's positive image would increase levels of normative conflict, collective guilt, and thus collective action. Although these studies demonstrate that highlighting the image-threatening nature of a transgression increases normative conflict, and thus collective guilt and action, compared to a transgression where it is clear, the group's image is unthreatened. However, it is still unclear whether providing image-threatening information increases normative conflict and guilt compared to simply learning about a group transgression, where image threat is more ambiguous. Research on the aversive nature of collective guilt would suggest that high identifiers are unlikely to spontaneously feel high levels of guilt (see Iyer et al., 2004; Wohl et al., 2006); making it more likely that high levels of image threat increase guilt rather than the low image threat, which decreases guilt. However, it is impossible to be sure about the direction of this effect without a control condition,<sup>4</sup> where participants only received information about the in-group transgression and not about how an out-group reacted to it. Therefore, in Study 3, we set out to replicate our prior findings with the addition of a control condition, to determine whether image-threatening information increases normative conflict, collective guilt, and thus collective action compared to simply learning about an in-group transgression.

### Method

**Participants and procedure.** Through Amazon's MTurk, 270 participants were recruited for this study. Participants were paid US\$1 for completing the study. Thirty-seven

participants (13% of the original sample) were excluded from analysis because they did not pay sufficient attention to the study. This left a sample of 233 participants (123 women,  $M_{age} = 34.44$  years, 80.6% White, 8.2% African American, and 11.2% other). Although the sample sizes in the first two studies were determined primarily by budgetary constraints and thus these studies were relatively underpowered, in this study, we used G\*Power to determine an optimal sample size (Faul, Erdfelder, Buchner, & Lang, 2009) based on the effect size of the interaction on collective guilt from Study 2 and 0.80 power. This study was identical to Study 2, except we added a control condition where no information was given about the reactions of Europeans to abuse at Guantanamo Bay, and some small changes were made to the measure of collective action to maintain its timeliness (see in the following).

**Materials and measures.** The measures of identification ( $\alpha = .95$ ), image threat ( $\alpha = .94$ ), normative conflict ( $\alpha = .78$ ), in-group-directed anger ( $\alpha = .92$ ), and collective guilt ( $\alpha = .92$ ) were the same as in Study 2. However, we used a new measure of shame used in recent research (Allpress, Brown, Giner-Sorolla, Deonna, & Teroni, 2014) that showed two types of shame (moral vs. image shame) can have different effects.<sup>5</sup> Participants completed the measure of identification before the manipulation, and the others after reading the manipulation. Participants indicated their agreement to the items of all scales on a 7-point Likert-type scale ranging from "strongly disagree" (1) to "strongly agree" (7), unless stated otherwise.

**Group image threat manipulation.** Participants read the same article about abuse at the Guantanamo Bay Detention Center as they read in the other studies. However, it was modified slightly to fit the current political context. This study was conducted in the run-up to the 2016 presidential elections, so the article was framed as discussing the importance of pressuring President Obama to close Guantanamo before he left office. In addition, in this study, participants were randomly assigned to read one of the three different versions. In the control condition, participants read only the first section of the article about the abuses that occurred at Guantanamo. In the image threat conditions, participants read the same text followed by a second section that was designed to threaten participants' group image; these were the same sections used in the high and low threat conditions in the earlier studies. In the low image threat condition, Europeans made situational attributions for abuse, citing the effects of the war on terror. In the high image threat condition, Europeans made stable internal attributions for the abuse.

**Collective action.** This measure was also almost the same as the measure used in Study 2, with one minor modification to fit the current political context. Before the other action options that had been used in the first study, participants

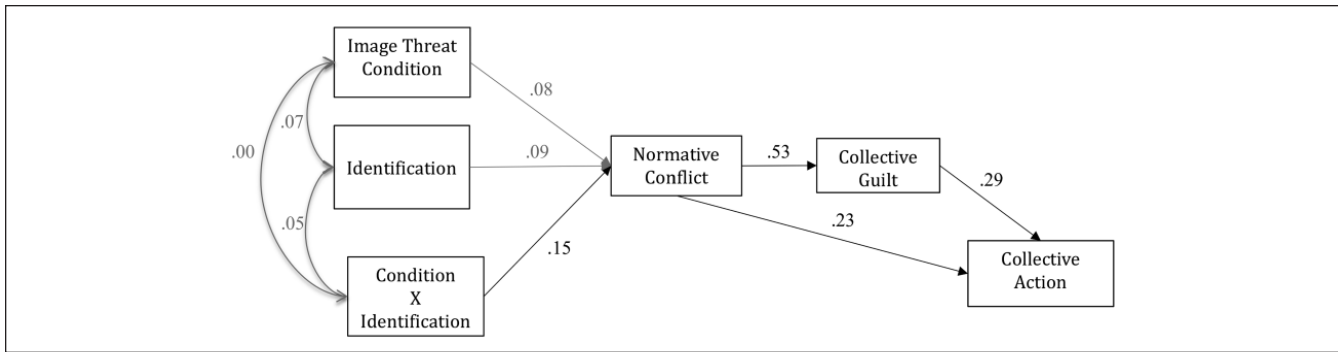


Figure 5. Full model tested using path analysis in Study 2; nonsignificant paths are displayed in gray.

Table 5. Bivariate Relationships Between Study 3 Variables (N = 233).

	M (SD)			1	2	3	4	5	6	7
1. D1	Control (1)	Low threat (0)	High threat (0)	—						
2. D2	Control (0)	Low threat (1)	High threat (0)	-.50**	—					
3. Political ideology	3.32 (1.92)	3.47 (1.97)	3.04 (1.69)	.02	.07	—				
4. Identification	4.48 (1.28)	4.45 (1.17)	4.21 (1.39)	.06	.04	.42**	—			
5. Image threat	5.55 (1.58)	5.79 (1.14)	6.06 (1.20)	-.13*	-.01	-.34**	-.26**	—		
6. Normative conflict	5.49 (1.38)	5.63 (1.26)	5.90 (1.39)	-.10	-.02	-.28**	-.17*	.64**	—	
7. Collective guilt	4.26 (1.63)	4.20 (1.57)	4.74 (1.52)	-.07	-.09	-.21**	-.08	.59**	.46**	—
8. Collective action	1.21 (2.23)	1.37 (2.35)	1.58 (2.12)	-.05	-.01	-.11†	-.16*	.28**	.14*	.27**

†p < .10. \*p < .05. \*\*p < .01.

were given the opportunity to sign a petition on the White House website. It was explained to participants that if the petition reached 100,000 signatures, the White House was required to at least issue an official response. The petition called on President Obama to close Guantanamo before he left office. After this, participants completed the same collective action measure as used in Study 2. This measure was scored according to the same method as Study 2.

Results

Table 5 presents means, standard deviations, and correlations among all variables. We first examined whether there was a significant interaction between the image threat condition and identification on all main study variables using Hayes’s (2013) PROCESS command with 5,000 iterations (Model 1). Because we now had three, rather than two conditions, we used the multicategorical IV feature in PROCESS. PROCESS created two dummy variables for condition: D1 (1 = control and 0 = low and high threat) and D2 (1 = low threat and 0 = control and high threat). The high threat condition was identified as a reference group (coded as zero in both D1 and D2). In this analysis, D1 reflects the comparison of the high threat condition with the control condition and D2 reflects the comparison of the high threat condition with the low threat condition (see Hayes & Preacher, 2014). PROCESS then includes both these variables and their interactions with the

moderator in the model, allowing a comparison of the high threat condition with both the control and the low threat conditions in the same model.

We report the statistics for the interactions as well as the direct effects of both dummy variables and identification in Table 6. As indicated in Table 6, and consistent with the first two studies, the two-way interaction between identification and D2 (high threat vs low threat) was significant on normative conflict, collective guilt, and collective shame. However, the interaction on in-group-directed anger was not significant. In addition, the two-way interaction between identification and D1 (high threat vs. control) was significant for normative conflict and collective guilt. The interaction was not significant for in-group-directed anger or collective shame, indicating that compared to a neutral control condition, the high threat condition only significantly raised levels of collective guilt and not other group critical emotions. These findings underscore the unique role of guilt (relative to anger and shame) in the tested process involving reactions to image threat and normative conflict.

A simple-effects analysis for both D1 and D2 (see Table 7) revealed that the interactions on normative conflict and collective guilt followed the same pattern: The high image threat condition affected levels of normative conflict only among those high on group identification<sup>6</sup> (1 SD above the mean); such that the high image threat condition increased normative conflict and collective guilt compared with both

**Table 6.** Significant Effects on Dependent Variables in Study 3.

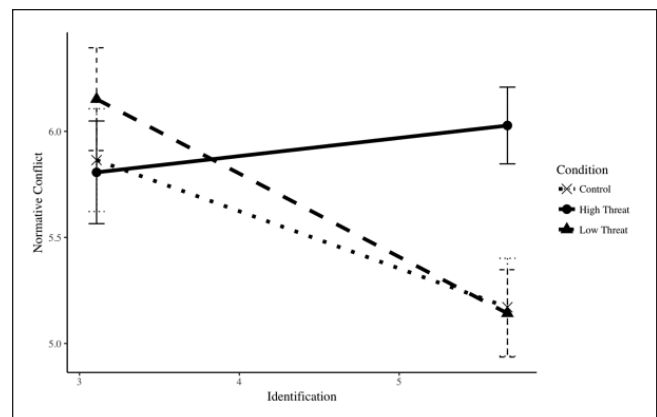
	Identification	D1	D2	D1 × Identification	D2 × Identification
Normative conflict	$b = 0.09, ns$	$b = 1.16, ns$	$b = 1.84, SE = 0.77, t = 2.40, p = .02$	$b = -0.36, SE = 0.16, t = -2.25, p = .03$	$b = -0.48, SE = 0.17, t = -2.84, p = .001$
Collective guilt	$b = 0.20, ns$	$b = 1.17, ns$	$b = 2.00, SE = 0.91, t = 2.22, p = .03$	$b = -0.38, SE = 0.19, t = -2.05, p = .04$	$b = -0.58, SE = 0.20, t = -2.92, p < .001$
Collective shame	$b = -0.04, ns$	$b = 0.03, ns$	$b = 1.41, ns$	$b = -0.10, ns$	$b = -0.42, SE = 0.21, t = -2.01, p = .046$
In-group anger	$b = -0.13, ns$	$b = 1.01, ns$	$b = 0.65, ns$	$b = -0.25, ns$	$b = -0.27, ns$
Collective action	$b = -0.45, SE = 0.18, t = -2.51, p = .06$	$b = -2.13, ns$	$b = -0.68, ns$	$b = 0.42, ns$	$b = 0.13, ns$

**Table 7.** Simple Effects and Means for the Dependent Variables in Study 3.

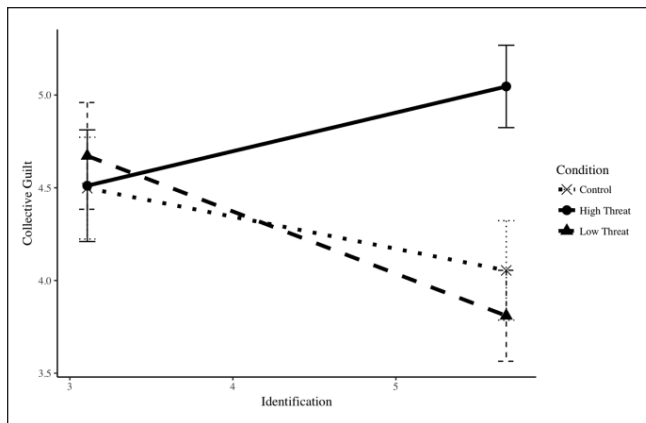
	Low Identification (-1 SD)					High Identification (+1 SD)				
	Low threat	Control	High threat	D1	D2	Low threat	Control	High threat	D1	D2
Normative conflict	6.16	5.86	5.81	$b = 0.06, ns$	$b = 0.35, ns$	5.14	5.17	6.03	$b = -0.85, SE = 0.29, t = -2.89, p < .001$	$b = -0.88, SE = 0.31, t = -2.86, p < .001$
Collective guilt	4.71	4.50	4.51	$b = -0.03, ns$	$b = 0.58, ns$	3.75	4.06	5.04	$b = -0.98, SE = 0.35, t = -2.84, p < .001$	$b = -1.29, SE = 0.36, t = -3.56, p < .001$
Collective shame	5.19	4.79	5.09	$b = -0.29, ns$	$b = 0.10, ns$	3.99	4.42	4.98	$b = -0.56, ns$	$b = -0.98, SE = 0.38, t = -2.57, p = .01$
In-group anger	4.71	5.14	4.92	$b = 0.22, ns$	$b = -0.20, ns$	3.66	4.16	4.58	$b = -0.42, ns$	$b = -0.92, SE = 0.37, t = -2.50, p = .01$
Collective action	1.81	1.26	2.08	$b = -0.82, ns$	$b = -0.28, ns$	0.92	1.18	0.98	$b = 0.26, ns$	$b = 0.06, ns$

the low threat and the control conditions (see Figures 6 and 7). Again, the manipulation did not have a significant impact on low identifiers. These interactions remained significant even if we controlled for political ideology. The interactions on collective shame and in-group-directed anger (nonsignificant) also followed a similar pattern, such that among high identifiers, these emotions gradually increased from the low threat to the control to the high threat condition, but differences between these conditions were not significant.

Unlike Study 2, in this study, there were no significant effects on collective action. We think this could be due to a floor effect on this variable: means on the measure were very low (1.1-1.8) and were substantially lower than means in the previous study (1.8-3.2). Furthermore, in Study 2, 62% of participants engaged in some form of collective action, whereas in this study, only 35% of participants did (see more in the discussion). Nevertheless, although there was no direct interactive effect of our conditions on collective action, indirect effects may exist in the absence of direct effects, particularly when they involve multiple steps or predictors (Hayes, 2009;

**Figure 6.** Interaction between condition and identification on normative conflict in Study 3, points are displayed at 1 SD above and below the mean.

Kenny & Judd, 2014); therefore, we still proceeded to test our full model including collective action using path analysis.



**Figure 7.** Interaction between condition and identification on collective guilt in Study 3, points are displayed at 1 SD above and below the mean.

**Path analysis.** Because there is no process model that tests serial mediated moderation, we tested our full model using SPSS AMOS 6. Overall, the model fit was good,  $\chi^2 (n=233, df=11) = 19.81, p = .05, CFI = .98, RMSEA = .06$ . The model revealed that (a) normative conflict fully mediated the relationship between the interaction of condition and identification and collective guilt and (b) collective guilt had a significant effect on collective action, and collective guilt fully mediated the effect of normative conflict on collective action (see Figure 8). We also tested an alternative model where we reversed the casual order of the dependent variables, such that collective action predicted collective guilt, which in turn predicted normative conflict. This model did not fit the data as well,  $\chi^2 (df=11) = 28.50, p = .002, CFI = .905, RMSEA = .09$ , and comparison of the chi-square values revealed that the change in chi-square values between the two models was significant,  $\Delta\chi^2 (df=8.70(1), p < .001$ . We also tested a model where we entered in-group-directed anger and collective shame as additional predictors of collective action. This model also fit the data well,  $\chi^2 (n=233, df=17) = 27.90, p = .05, CFI = .986, RMSEA = .053$ , but in this study, once all three emotions were included, only in-group-directed anger significantly predicted collective action.

In sum, this study indicates that highlighting the image-threatening nature of a group transgression increases normative conflict and collective guilt among high identifiers both compared to merely learning about a transgression (control condition) or when there is evidence the transgression does not affect the group's image (low threat). However, in this study, there was no direct effect of the manipulation on collective action for high identifiers, although there was an indirect effect via normative conflict and collective guilt.

## Study 4

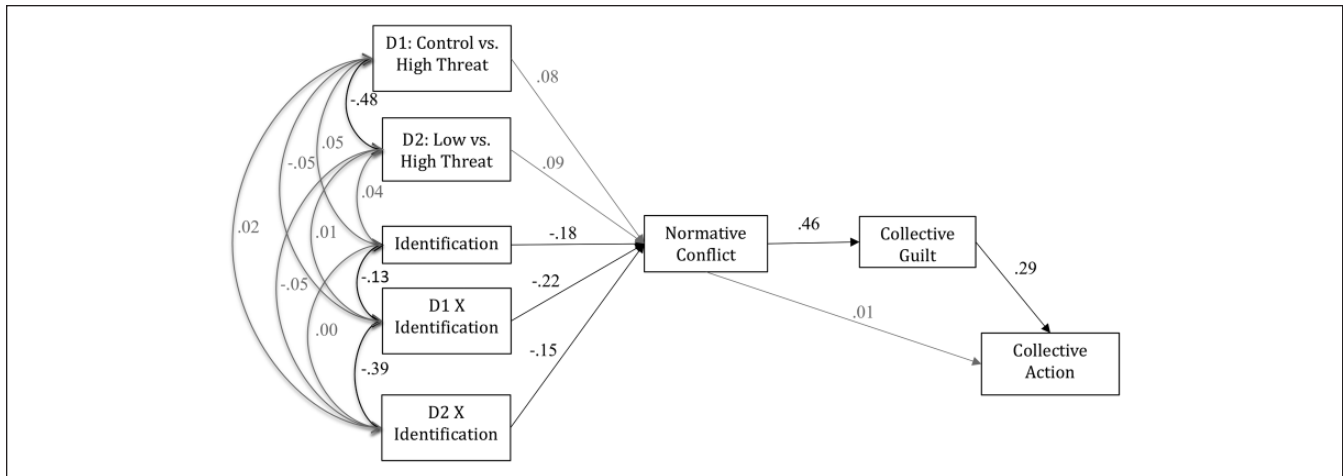
The first three studies supported our hypothesis that inducing threat to the group's positive image would increase levels of

normative conflict, collective guilt, and thus collective action. However, the manipulation of image threat was somewhat indirect, and thus, it was possible that there were factors other than image threat driving the effect. Although our manipulation of image threat was based on previously used materials (see Iyer et al., 2007), it used external versus internal attributions to manipulate image threat. Thus, it is possible that the attributions themselves are the driving factor of guilt and action and not the image threat they produce. In addition, the stable attribution (high threat) condition implicitly conveys lower expectations of Americans because of the stable attributions made. Furthermore, the quotes used to convey the image threat in the high threat condition make very specific critiques of Americans (e.g., imperialistic and unjust). Therefore, it is possible that rather than the image threat driving the effect of the manipulation, the differences observed between conditions could be due to one of these factors mentioned above. As a result, in Study 4, we aimed to replicate our previous findings with a more direct manipulation of image threat. In this study, image threat was manipulated by telling participants that a large percentage (75%) of Europeans now had either a favorable (low threat) or unfavorable (high threat) view of Americans as well as quotes explaining how abuse at Guantanamo had lead Europeans to either question and change their image of Americans (high threat) or that they maintained their positive overall image of Americans (low threat).

In addition, we sought to address a number of other gaps left by our previous studies. First, our theoretical reasoning is based in part on the assumption that group members who are highly identified with the group are particularly concerned with maintaining the group's positive image. Therefore, we added a measure of concern for the group's image to this study, to allow us to test this theoretical assumption. Second, because in Study 3 there was no direct interaction on collective action, in this study we included a more traditional measure of collective action intentions, in addition to our behavioral measure of action.

## Method

**Participants and procedure.** Through Amazon's MTurk, 270 participants were recruited for this study. Participants were paid US\$1 for completing the study. Thirty-three participants (12% of the original sample) were excluded from analysis because they did not pay sufficient attention to the study. This left a sample of 237 participants (125 women,  $M_{age} = 35.48$  years, 76.7% White, 9.3% African American, and 14% other). The sample size for this study was based on the same analysis used to establish the sample size for Study 3. This study was identical to Study 3, except for changes made to the manipulation, added measures of collective action intentions and concern for the group's image, and some minor changes that were made to the general text of the manipulation and the measure of collective action to maintain their



**Figure 8.** Full model tested using path analysis in Study 3; nonsignificant paths are displayed in gray.

timeliness (see more details below).

**Materials and measures.** The measures of identification ( $\alpha = .94$ ), image threat ( $\alpha = .90$ ), normative conflict ( $\alpha = .84$ ), in-group-directed anger ( $\alpha = .96$ ), shame ( $\alpha = .96$ ), and collective guilt ( $\alpha = .95$ ) were the same as in Study 3. Participants completed the measure of identification before the manipulation and the others after reading the manipulation. Participants indicated their agreement to the items of all scales on a 7-point Likert-type scale ranging from “strongly disagree” (1) to “strongly agree” (7), unless stated otherwise.

**Group image threat manipulation.** Participants read a similar article about abuse at the Guantanamo Bay Detention Center to the other studies. However, it was modified slightly to fit the current political context. At the time the study was run, Donald Trump had already become president, and thus, there was less discussion about closing Guantanamo. However, a number of prisoners were engaged in a hunger strike to demand for their rights to a fair trial. At the time of the study, the Department of Defense had just changed their policy for managing hunger-striking prisoners, and this had received attention in the media. Thus, the new article was framed around this hunger strike. In addition, we made changes to the manipulation mentioned above. In the control condition, participants read the same general description of abuses in Guantanamo simply with an introduction referencing the ongoing hunger strike. In the image threat conditions, participants read the same text followed by a second section that was designed to threaten participants’ group image. In both conditions, participants read a short paragraph describing a large Gallup poll recently conducted across Europe regarding attitudes toward the United States. In both conditions, participants read that 75% of Europeans “found the treatment of prisoners at Guantanamo unjust or very unjust.” Then, in the high/low threat conditions, participants were informed

that “85% of Europeans (still) held a unfavorable/favorable or very unfavorable/favorable image of Americans.” In addition, quotes were given from survey respondents that highlighted the connection between Guantanamo and their image of Americans; these quotes were also used to ensure that expectations of Americans remained constant across conditions. For example, one quote was “this isn’t what I’ve come to expect of Americans, and/but it does/doesn’t make me question my view of Americans as good and moral people” in the high/low threat conditions.

**Collective action intentions.** This measure was based on previous measures of collective action intentions (see Shuman, Cohen-Chen, Hirsch-Hoefler, & Halperin, 2016; Tausch et al., 2011) and asked participants to rate their willingness to engage in a variety of actions. This measure consisted of five items including “I would sign a petition calling for fair treatment and the right to a trial of prisoners at Guantanamo” and “I would participate in a demonstration against the mistreatment of prisoners at Guantanamo” ( $\alpha = .90$ ).

**Collective action.** This measure was also similar to the measure used in Study 2, with some minor modifications to fit the current political context. The petition was changed from a petition to close Guantanamo to one calling on the Secretary of Defense to ensure that the prisoners at Guantanamo were treated humanely and received a fair trial and to work toward eventually closing the prison. In addition, participants were told that they would be given a 50-cent bonus for completing the survey and were then given the opportunity to donate some of this money to organizations working on this issue. Other than this, the measure was the same as in Studies 2 and 3.

**Concern for the group’s image.** Participants rated their agreement with a scale of seven statements designed to mea-

**Table 8.** Bivariate Relationships Between Study 4 Variables ( $N = 237$ ).

	M (SD)			1	2	3	4	5	6	7	8	9
1. D1	Control (1)	Low Threat (0)	High Threat (0)	—								
2. D2	Control (0)	Low Threat (1)	High Threat (0)	-.52**	—							
3. Political ideology	3.64 (1.91)	3.39 (1.92)	3.36 (1.84)	.06	-.03	—						
4. Identification	4.82 (1.11)	4.37 (1.31)	4.21 (1.26)	.14*	-.13	.31**	—					
5. Image concern	5.25 (1.21)	5.10 (1.01)	5.09 (1.15)	.06	-.03	-.12	.33**	—				
6. Image threat	5.73 (1.41)	5.18 (1.42)	6.07 (1.00)	.04	-.25**	-.33**	-.14*	.33**	—			
7. Normative conflict	5.62 (1.61)	5.57 (1.38)	5.76 (1.24)	-.01	-.04	-.25**	-.20**	.41**	.70**	—		
8. Collective guilt	5.06 (1.75)	4.92 (1.66)	4.98 (1.71)	.03	-.03	-.41**	-.10	.50**	.66**	.71**	—	
9. Collective action intentions	3.61 (1.81)	3.42 (1.50)	3.49 (1.75)	.04	-.04	-.45**	-.09	.38**	.39**	.41**	.62**	—
10. Collective action	1.24 (2.16)	1.49 (2.23)	1.19 (1.91)	-.03	.06	-.22**	-.13*	.24**	.18**	.30**	.34**	.40**

† $p < .10$ . \* $p < .05$ . \*\* $p < .01$ .

sure their concern about the group's (in this case, America's) image that was developed for this study. Items included "It is important to be that America be viewed positively" and "I think that it is critical that America maintain its positive moral image in the world" ( $\alpha = .87$ ).

### Results and Discussion

Table 8 presents means, standard deviations, and correlations among all variables. We first test our hypothesis that group identification would be positively related to concern for the groups image. Indeed, group identification and group image concern were positively correlated and this relationship did not change between conditions.

Next, we examined whether there was a significant interaction between the image threat condition and identification on all main study variables using Hayes's (2013) PROCESS command with 5,000 iterations (Model 1). We again used the multicategorical IV feature in PROCESS. PROCESS created two dummy variables for condition: D1 (1 = control and 0 = low and high threat) and D2 (1 = low threat; 0 = control and high threat). The high threat condition was identified as a reference group (coded as zero in both D1 and D2). In this analysis, D1 reflects the comparison of the high threat condition with the control condition and D2 reflects the comparison of the high threat condition with the low threat condition (see Hayes & Preacher, 2014). PROCESS then includes both these variables and their interactions with the moderator in the model, allowing a comparison of the high threat condition with both the control and the low threat conditions in the same model.

We report the statistics for the interactions as well as the direct effects of both dummy variables and identification in Table 9. As indicated in Table 9, and consistent with the first two studies, the two-way interaction between identification and D2 (high threat vs low threat) was marginally significant on normative conflict and significant for collective guilt, in-group-directed anger, collective shame, and collective action intentions. However, in this study, the two-way interaction

between identification and D1 (high threat vs control) was not significant for any of the study variables. A one-way ANOVA of the image threat measure (manipulation check) revealed that while there were significant differences between the conditions,  $F(2, 237) = 9.35, p < .001$ , post hoc contrasts revealed that levels of image threat were not significantly higher in the high threat condition compared to the control ( $p > .10$ ), but the low threat condition was significantly lower than both the high threat and control conditions on image threat ( $p$ 's  $< .01$ ). Whereas in Study 3, the control was more similar to the low threat condition in the amount of image threat. It is possible that contextual changes between Study 3 and Study 4 affected the baseline amount of image threat in the control condition between the two studies. A major contextual change that occurred between Study 3 and Study 4 was the election of Donald Trump, and a recent Pew Research survey shows that this significantly worsened America's image abroad (Wike, Stokes, Poushter, & Fetterolf, 2017). It may be that Americans (especially the slightly more liberal Mturk population) are aware of this trend and thus at default more concerned with America's image. As a result, we focused more on the comparison between low and high threat in our further analyses, and this was also the comparison central to our hypotheses.

A simple-effects analysis for both D1 and D2 (see Table 10) revealed that the interactions on normative conflict and collective guilt followed a similar pattern: The high image threat condition affected levels of normative conflict only among those high on group identification (see Note 6; 1  $SD$  above the mean); such that the high image threat condition significantly increased normative conflict, collective guilt, and collective action intentions (marginally) compared to the low threat condition, but not the control condition (see Figures 9-12). For low identifiers, there was usually no effect of condition. However, for collective guilt, the opposite pattern was found: the high threat condition decreased collective guilt compared to the low threat and control conditions. If political ideology was added as covariate, the

**Table 9.** Significant Effects on Dependent Variables in Study 4.

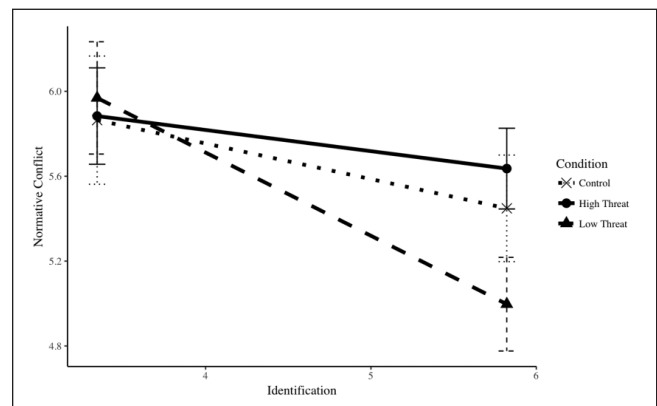
	Identification	D1	D2	D1 × Identification	D2 × Identification
Normative conflict	$b = 0.10, ns$	$b = 0.21, ns$	$b = 1.06, ns$	$b = -0.07, ns$	$b = -0.29, SE = 0.17, t = -1.68, p = .09$
Collective guilt	$b = 0.18, ns$	$b = 0.91, ns$	$b = 3.00, SE = 0.97, t = 3.08, p = .002$	$b = -0.18, ns$	$b = -0.69, SE = .21, t = -3.31, p = .001$
Collective shame	$b = 0.03, ns$	$b = -0.47, ns$	$b = 1.62, ns$	$b = 0.09, ns$	$b = -0.47, SE = 0.22, t = -2.18, p = .03$
In-group anger	$b = -0.08, ns$	$b = 0.18, ns$	$b = 1.67, SE = .92, t = 1.81, p = .07$	$b = -0.02, ns$	$b = -0.45, SE = 0.20, t = -2.30, p = .02$
Collective action intentions	$b = -0.04, ns$	$b = 0.13, ns$	$b = 1.87, SE = 0.97, t = 1.92, p = .06$	$b = -0.004, ns$	$b = -0.44, SE = 0.21, t = -2.11, p = .04$
Collective action	$b = 0.07, ns$	$b = 1.85, ns$	$b = 2.37, SE = 1.22, t = 1.95, p = .05$	$b = -0.38, ns$	$b = -0.47, ns$

**Table 10.** Simple Effects and Means for the Dependent Variables in Study 4.

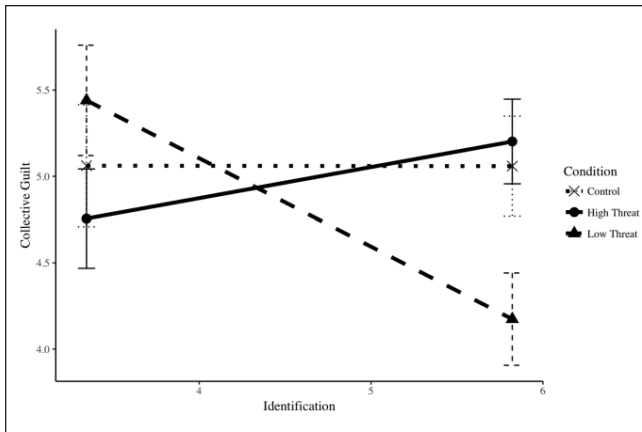
	Low Identification (-1 SD)					High Identification (+1 SD)				
	Low threat	Control	High threat	D1	D2	Low threat	Control	High threat	D1	D2
Normative conflict	5.86	5.96	5.88	$b = -0.02, ns$	$b = 0.09, ns$	5.00	5.45	5.64	$b = -0.19, ns$	$b = -0.64, SE = 0.32, t = -1.97, p = .05$
Collective guilt	5.44	5.06	4.76	$b = 0.30, ns$	$b = 0.65, SE = 0.36, t = 1.92, p = .06$	4.17	5.05	5.20	$b = -0.14, ns$	$b = -1.03, SE = .38, t = -2.65, p < .001$
Collective shame	4.73	4.51	4.67	$b = -0.16, ns$	$b = 0.05, ns$	3.62	4.81	4.74	$b = 0.07, ns$	$b = -1.11, SE = 0.40, t = -2.77, p < .001$
In-group anger	4.71	4.67	4.56	$b = 0.11, ns$	$b = -0.15, ns$	3.38	4.41	4.36	$b = -0.05, ns$	$b = -0.98, SE = .36, t = -2.66, p < .001$
Collective action intentions	3.84	3.56	3.45	$b = 0.11, ns$	$b = 0.40, ns$	2.84	3.64	3.54	$b = 0.11, ns$	$b = -0.70, SE = 0.38, t = -1.79, p = .07$
Collective action	1.90	1.69	1.10	$b = 0.57, ns$	$b = 0.65, ns$	0.91	0.93	1.28	$b = -0.34, ns$	$b = -0.36, ns$

interactions on normative conflict and collective action were no longer significant, but the interaction on collective guilt remained significant; in all cases, the pattern of effects remained the same. The interactions on collective shame and in-group-directed anger also followed a similar pattern, such that among high identifiers, these emotions increased from the low threat to the high threat condition, but not from the control to the high threat.

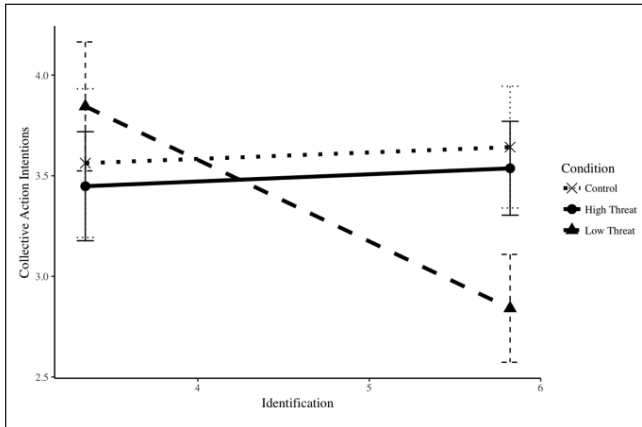
Unlike Study 2 but similarly to Study 3, in this study there were no significant effects on collective action. We think this could be due to a floor effect on this variable: means on the measure were very low (approx. 1.5) and were substantially lower than means in Study 2. Further in Study 2, 62% of participants engaged in some form of collective action, whereas in this study only 32% of participants did (see more in the discussion). Nevertheless, although there was no direct interactive effect of our conditions on collective action, indirect effects may exist in the absence of direct effects, particularly when they involve multiple steps or predictors (Hayes, 2009; Kenny & Judd, 2014); therefore, we still proceeded to

**Figure 9.** Interaction between condition and identification on normative conflict in Study 4, points are displayed at 1 SD above and below the mean.

test our full model including collective action using path analysis.



**Figure 10.** Interaction between condition and identification on collective guilt in Study 4, points are displayed at 1 SD above and below the mean.



**Figure 11.** Interaction between condition and identification on collective action intentions in Study 4, points are displayed at 1 SD above and below the mean.

**Path analysis.** Because there is no Process model that tests serial mediated moderation, we tested our full model using SPSS AMOS 6. Overall, the model fit was good,  $\chi^2 (n = 237, df = 11) = 18.04, p = .08, CFI = .99, RMSEA = .05$ . The model revealed that (a) normative conflict fully mediated the relationship between the interaction of condition (high threat versus low threat) and identification and collective guilt and (b) collective guilt had a significant effect on collective action, and collective guilt fully mediated the effect of normative conflict on collective action (see Figure 11). We also tested an alternative model where we reversed the casual order of the dependent variables, such that collective action predicted collective guilt, which in turn predicted normative conflict. This model did not fit the data as well,  $\chi^2 (df = 11) = 23.62, p = .01, CFI = .99, RMSEA = .07$ , and comparison of the chi-square values revealed that the change in chi-square values between the two models was significant,  $\Delta\chi^2 (df) = 5.58 (0), p < .001$ . We also tested a model where we

entered in-group-directed anger and collective shame as additional predictors of collective action. This model also fit the data well,  $\chi^2 (n = 237, df = 18) = 25.56, p = .11, CFI = .99, RMSEA = .04$ , and in this study, once all three emotions were included, only collective guilt significantly predicted collective action.

Overall, although they are more nuanced than the results of the first three studies, the current results still provide important support for our main hypothesis that framing in-group transgressions as posing a high rather than low threat to the group's image increases normative conflict, collective guilt, and thus collective action. Thus, this study aligns with the other studies presented and supports our theoretical argument that highlighting the image-threatening nature of in-group transgressions is an effective means for promoting reparatory action. However, looking across all studies, there are three limitations that become apparent. First, the direction of the effect of image threat is somewhat unclear because in Study 4, unlike Study 3, the high threat condition did not increase the main dependent variables compared to the control condition. Second, the effects the image threat condition on collective action among high identifiers are weak, either marginal (Study 2) or nonsignificant (Studies 3 and 4). This may be partially due to the fact that we are using a behavioral measure of action, and thus, effects are small and difficult to detect. Finally, while our hypotheses rest on normative action leading to collective guilt as the key process leading to action, in Study 3, anger was the stronger predictor of guilt. Although we address these limitations conceptually in the general discussion, we aimed to first address them empirically by using a mini meta-analysis.

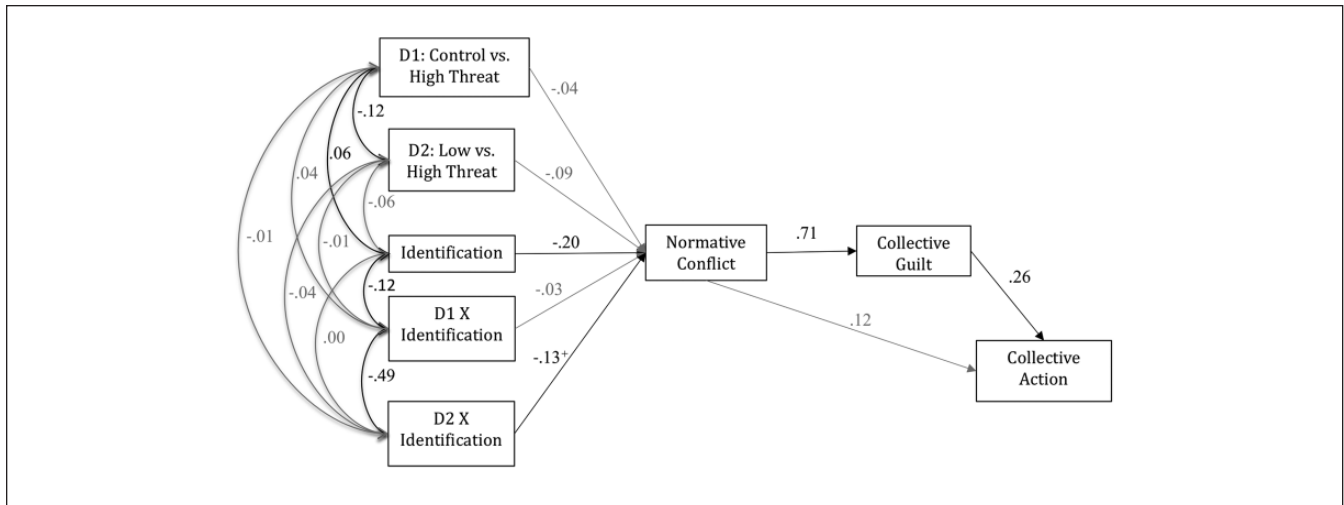
### Internal Meta-Analysis

To address these remaining limitations and examine the robustness of our main hypothesized effects, we conducted an internal meta-analysis, or a mini meta-analysis, on the studies presented here. In a recent paper, Goh, Hall, and Rosenthal (2016) outline the many benefits of conducting such a mini meta-analysis and outline a detailed method for doing so.

In our mini meta-analysis, we had four main aims: (a) to generally aggregate our main hypothesized effects, (b) to provide evidence that the high threat did increase our key dependent variables compared to a control condition, (c) to aggregate marginal effects and trends on our behavioral measure of collective action to determine whether there was an effect on actual behavior, and (d) to aggregate the effects of various emotions across studies to demonstrate that collective guilt was overall the strongest predictor of action.

**Main hypotheses, control condition, and collective action.** Our main hypothesis was that a group transgression that threatened the group's image (high image threat) would increase normative conflict, collective guilt, and thus collective action





**Figure 12.** Full model tested using path analysis in Study 4; nonsignificant paths are displayed in gray.  
<sup>†</sup> $p < .10$ .

**Table 11.** Effect Sizes of Condition Differences Among High Identifiers (+1 SD).

	Study 2 (N = 162)		Study 3 (N = 233)		Study 4 (N = 237)	
	High threat versus low threat	High threat versus low threat	High threat versus control	High threat versus low threat	High threat versus control	
Normative conflict	.25	.31	.30	.22	.06	
Collective guilt	.26	.38	.29	.29	.04	
Collective action	.23	.01	.05	.08	.08	

Note. All effect sizes are reported in terms of Pearson's *r*.

compared to a transgression that did not threaten the group's image (low threat condition) or a transgression that was not linked to the group's image (control condition). We meta-analyzed only the three studies that included all study variables (Studies 2-4), using fixed effects in which the key effect size was weighted by sample size. In this case, the key effect size was the difference between the high image threat condition and low image threat (or control condition) on normative conflict, collective guilt, and collective action for high identifiers—in statistical terms, the coefficient of the simple effect of condition at one *SD* above the mean of identification. As this simple effect essentially represents a difference between conditions, it can also be expressed in terms of Pearson's *r* (or Cohen's *d*). We first converted these simple effect coefficients into Pearson's *r* for ease of analysis (see Table 11) and proceeded according to the methods outlined in Goh et al. (2016). All correlations were then Fisher's *z* transformed for analyses and converted back to Pearson correlations for presentation of general effect sizes. Overall among high identifiers, the high threat significantly increased normative conflict ( $M r = .26, Z = 6.45, p < .001$ , two-tailed), collective guilt ( $M r = .32, Z = 7.87, p < .001$ , two-tailed), and collective action ( $M r = .10, Z = 2.52, p = .01$ , two-tailed)

compared to the low threat condition. In addition, the high threat significantly increased normative conflict ( $M r = .18, Z = 3.88, p < .001$ , two-tailed) and collective guilt ( $M r = .17, Z = 3.57, p < .001$ , two-tailed), but not collective action ( $M r = .07, Z = 1.42, p = .15$ , two-tailed) compared to the control condition (based on only Studies 3 and 4). Overall, this supports our hypotheses that high versus low threat increases normative conflict and collective guilt and has a small but significant effect actual action behavior. In addition, it indicates that overall, the high threat condition does increase normative conflict and collective guilt compared to a control condition.

**Emotions and collective action.** Across studies, the emotion that best predicted collective action when controlling for all emotions varied between guilt and anger, though we hypothesized normative conflict leading to guilt as the key process for inducing action. To help resolve this inconsistency, we conducted a mini meta-analysis of the effects of each emotion on emotions (controlling for the other emotions). Becker and Wu (2007) show that it is possible to synthesize slopes from a multiple regression in a meta-analysis provided that the following conditions are met: (a) the outcome variable is

measured similarly across studies, (b) the predictor variables are measured similarly across studies, and (c) exactly the same predictors are included in the model across studies. Because our predictor and outcome variables were measured almost identically across studies, the first two conditions are met. To meet the third condition, we took the effects (standardized betas) for each emotion from a multiple regression containing only the three emotions predicting collective action. We then aggregated these effects following the generalized least squares approach outlined in Becker and Wu (2007). Overall, collective guilt had the only significant unique effect on collective action ( $M\beta = .22, Z = 2.41, p = .02$ , two-tailed), and the unique effects of in-group-directed anger ( $M\beta = .15, Z = 1.82, p = .07$ , two-tailed) and collective shame ( $M\beta = .03, Z = 0.81, p = .42$ , two-tailed) were not significant.

## General Discussion

Understanding how to motivate highly identified group members to take action to stop and correct in-group transgressions is an important step in promoting more positive intergroup relations. Highly identified group members often have greater influence on the group and greater freedom to criticize it without harsh penalties, thus making their actions particularly effective for correcting in-group transgressions. Highly identified group members' need to maintain a positive group identity often prevents them from recognizing and addressing in-group transgressions. However, the results of the current research indicate that the threat in-group transgressions pose to the group's moral image can motivate high identifiers to take action to address the harm it is doing to their group's image. Specifically, we found that when high identifiers were presented with evidence that actions committed by their group were harming the group's moral image in the eyes of others, they were more willing to engage in collective action to address the in-group transgression.

In addition, we found that this process was mediated by normative conflict and collective guilt. In the high image threat condition, high identifiers saw the in-group transgression as significantly more in conflict with American norms and values. In turn, this increased normative conflict also led to increased collective guilt for the transgression, which also drove collective action, mediating the effect of normative conflict. In sum, demonstrating the harmful effects of transgressions on the group's external image can in turn cause high identifiers to see those actions as a violation of the group's own norms, which in turn promotes collective guilt and collective action to address the transgression.

## Theoretical and Practical Implications

These findings support and extend the normative conflict model of dissent (Packer, 2008); according to this model, when group members see a group's action as harmful to the group,

they are likely to appraise this action as conflicting with the group's own norms for behavior, and this normative conflict in turn leads them to criticize the group's action. Although this model has been empirically tested (e.g., Packer & Chasteen, 2010), it has not been examined in the context of intergroup conflict. These findings indicate that normative conflict can also motivate highly identified group members to take action against transgressions committed against an out-group. Furthermore, this research indicates that image threat is an effective way to induce normative conflict in this context. Because high identifiers are particularly concerned with the group's positive identity, including its moral image, actions which threaten this image can be appraised as harmful to the group and thus induce normative conflict. In addition, it supports and extends the needs-based model of reconciliation (Shnabel & Nadler, 2008), which suggests that high power group members are particularly concerned with being morally accepted. These data indicate that the need for moral acceptance can even drive those highly identified with a high power group to correct transgressions as a result of the threat they might pose to the group's moral image. Furthermore, this research suggests that this motivation can sometimes overcome the usual reticence of highly identified group members to recognize group transgressions and experience collective guilt.

Despite the fact that collective guilt theoretically depends on some identification with the group, researchers have often found that identification with a group decreases collective guilt for its actions because group members are motivated to maintain the group's positive identity (see Branscombe et al., 2004; Doosje et al., 1998). Although Doosje et al. (1998) were able to induce collective guilt in high identifiers when transgressions were unambiguous, other research indicates that high identifiers can often justify or deny even clear transgressions (see Branscombe & Miron, 2004; Roccas et al., 2006). Similarly, in this study, high identifiers felt significantly less collective guilt than low identifiers low threat and sometimes the control conditions, even though the transgressions were described unambiguously in all conditions. It was only the high image threat condition that increased guilt among high identifiers. These results indicate that moral image threat can overcome the usual barriers to the experience of collective guilt among high identifiers and thus motivate them to support action in order to restore their group's positive image.

Taken together, these implications can also inform practical interventions to generate criticism and action against in-group transgressions. Because in-group transgression often occur during war, and other times when criticism of one's in-group is especially unlikely, activists often face a particular challenge in motivating action to address in-group transgressions. These results indicate that demonstrating how the transgressions committed by the group threaten the group's moral image in the eyes of other groups can be an effective way to mobilize even those highly identified with the group. This in turn could lead to the increased likelihood of the ces-

sation of ongoing transgressions and/or corrective or compensatory policies for those harmed.

However, considering the applied implications of this research also raises potential contradictions between recommendations based on the collective guilt literature. Most other interventions to increase collective guilt have focused on self-affirmation as means to remove the defenses that often prevent the experience of collective guilt (see Čehajić-Clancy et al., 2011; Gunn & Wilson, 2011; Miron, Branscombe, & Biernat, 2010). This literature relies on the assumption that affirming the self can allow people to cope with information that threatens the group's positive identity without becoming. However, our perspective, and that of Lowery et al. (2012) and others, highlights the potential of the motivation for a positive group identity (in this case specifically as positive image) that sometimes leads to defensiveness to also motivate guilt and action under the right circumstances. And Lowery et al. (2012) actually show that affirmation can eliminate this constructive effect of the motivation to preserve a positive image of one's group, so it may be difficult to combine these two types of interventions. Thus, future research may be needed to identify under what circumstances each of these types of interventions is effective.

### *Limitations and Future Directions*

However, this research does suffer from some limitations that should be addressed in future research. First, all studies were conducted in the context of the War on Terror and abuse that has occurred at Guantanamo Bay. It is important for future research to examine these processes in other contexts because it is possible that image threat does not always lead to corrective action. Stated differently, boundary conditions to the obtained effects may still exist, groups could try to explain or justify their past actions as a way to reduce image threat, especially in contexts where corrective action might be seen as impossible or extremely costly. For example, Schmitt, Miller, Branscombe, and Brehm (2008) found that collective guilt was reduced when reparative action was highly costly. Another possible boundary condition might be that the manipulation is especially effective for participants who believe groups can change (i.e., hold and incremental theory about groups, see Dweck, Chiu, & Hong, 1995; Rydell, Hugenberg, Ray, & Mackie, 2007). Participants may need to believe that groups can change, to take action to change their group's image in the eyes of another group.

Second, the high threat condition did not produce consistently higher levels of collective guilt among high identifiers in comparison to a control condition. In Study 4, there were little differences between the high threat and control condition. It appears that in Study 4, the control condition was interpreted as highly image threatening even without additional information specifically about image threat. Although the low and high threat conditions produced similar levels of

image threat across these studies, the control condition produced higher levels of image threat in Study 4 compared to Study 3, which may point to a contextual change that led to an increase in baseline levels of image threat. A major contextual change that occurred between Study 3 and Study 4 was the election of Donald Trump, and this significantly worsened America's image abroad (Wike et al., 2017). It may be that Americans (especially the slightly more liberal Mturk population) are aware of this trend and thus are already primed to be considering actions in light of a negative image of America. Despite this limitation of Study 4, the mini meta-analysis did indicate that overall when the results of Studies 3 and 4 were combined, the high threat condition did raise collective guilt relative to a control condition among high identifiers.

Third, we only found a direct effect of the manipulation on collective action among high identifiers in Study 2; however, in Studies 3 and 4, the manipulation did have an indirect effect on collective action through collective guilt and normative conflict. We think this was due to a floor effect of the later studies, where only a third of the sample engaged in action at all, as well as the fact that this variable measured actual behavior and thus effect sizes are likely to be smaller. For this reason, we conducted a mini meta-analysis, which can help detect a small effect across studies that may not rise to significance in any one study (Goh et al., 2016) and indeed found the hypothesized effect that the high threat condition increased collective action compared to the low threat condition.

In addition, while this research does argue that collective guilt is an important predictor of collective action, we do not make the claim that it is the only or the strongest emotional predictor of action. Although collective guilt was a significant predictor of collective action in all studies and the meta-analysis of these studies showed it was the strongest predictor of action, in Study 3, anger was the most proximal predictor of action when all emotions were included in the same model. This reflects an inconsistency in the literature on this issue, with some research suggesting guilt is more strongly linked to action and other research arguing that in-group-directed anger is a better predictor of action. This highlights the need for additional research on which emotion is most strongly linked to action as well as situations in which certain emotions maybe more effective than others at motivating action.

Another possibility is that all group critical emotions need to be present to drive action. Although we showed that across studies guilt was the strongest predictor of action, participants generally felt similar levels of all three group critical emotions. Thus, it is possible that guilt drives action when these other emotions are present. Similarly, there may be additional mechanisms for the effects found here than the ones directly measured and discussed. For example, Mackie et al. (2000) found that high in-group support for a certain group-based emotion increases the likelihood that group members will feel that emotion. It may be that when a

transgression is perceived as violating a group norm (i.e., normative conflict), people may assume that other group members will also feel guilt and take action against the transgression. This implied in-group support in normative conflict could also potentially explain the increase in collective guilt and collective action among high identifiers.

Finally, this research only examined image threat in the eyes of a fairly uninvolved out-group (i.e., Europeans), as this was the best test of our hypothesis that an external image threat could be most effective on high identifiers. However, it would be interesting to determine the boundary conditions of this image threat. In other words, what are the characteristics necessary for an out-group opinion to generate an image threat, or an image threat capable of inducing collective guilt. We chose Europeans because they are an out-group that is viewed relatively positively; however, it would be interesting to examine whether this is in fact a prerequisite. It is possible that the effectiveness of image threat is determined by more structural factors. For example, the current American-led world order depends in part on European support (e.g., in the United Nations or through North Atlantic Treaty Organization (NATO) or other international organizations). Thus, image threat from this group may seem to pose a threat to the legitimacy and/or stability of American dominance. This is in line with work by Chow, Lowery, and Hogan (2013), which showed that information that minorities hold Whites in low regard can be perceived as a threat to the group's status and thus motivate compensatory action.

In addition, it would be interesting to examine whether criticism from the in-group can have similar effects to those found here. For example, there is research indicating that criticism delivered from within one's group is received more favorably (see Hornsey et al., 2004), which could suggest that image threat in the eyes of one's own group may be even more effective; however, this effect can disappear in contexts of intergroup conflict (Ariyanto, Hornsey, & Gallois, 2006). However, image threat in the eyes of one's group may no longer be perceived as a threat to the group's image as it is internal rather than external criticism, thus the manipulation might lose its effectiveness. Future research should explore this and other possibilities.

This research helps to address the important question of how to motivate strongly identified group members to take action against transgressions committed by their group. Although often the motivation to protect their group's positive identity prevents high identifiers from recognizing or addressing in-group transgressions, this research indicates that it may be possible to harness that same motivation to drive action to correct these transgressions. By demonstrating how transgressions threaten the positive image of the groups in the eyes of others, this motivation to protect the group's identity can be redirected to correct in-group transgressions rather than deny or ignore them.

## 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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: During the writing of this article, the first author was supported by a Fulbright Student Fellowship through the U.S. Department of State and United States-Israel Educational Fund. This work was also supported by a European Research Council grant (335607) to the last author.

## Notes

1. We refer to this condition as the "low threat" condition, because while it provides positive information about the group's image, it still presents an in-group transgression, which is likely to be perceived as threatening in and of itself.
2. When attachment and glorification were analyzed separately as moderators, the results followed the same pattern. However, the interaction with attachment was marginal ( $p = .08$ ) and the interaction with glorification was significant ( $p = .03$ ).
3. When attachment and glorification were analyzed separately as moderators, the results followed the same pattern. For normative conflict, the interaction with attachment was significant ( $p = .03$ ) and the interaction with glorification was marginal ( $p = .07$ ). For collective guilt, both attachment and glorification interactions with condition were significant at the  $p = .01$  level. Glorification and attachment also did not load onto separate factors in this study.
4. We refer to this as a control condition in relation to the two image threat conditions, because it does not provide either positive or negative direct information about the group's image, even though it still provides information about a group transgression and thus is not a true empty control.
5. However, we did not find different effects for these two types and all items loaded on one factor in a factor analysis, so we simply combined them into a general measure of collective shame.
6. When attachment and glorification were analyzed separately as moderators, the results followed the same pattern, and the interactions were significant.

## ORCID iD

Eric Shuman  <https://orcid.org/0000-0003-3830-4926>

## Supplemental Material

Supplementary material is available online with this article.

## References

- Allpress, J. A., Brown, R., Giner-Sorolla, R., Deonna, J. A., & Teroni, F. (2014). Two faces of group-based shame: Moral shame and image shame differentially predict positive and negative orientations to ingroup wrongdoing. *Personality and Social Psychology Bulletin*, *40*, 1270-1284. doi:10.1177/0146167214540724

- Ariyanto, A., Hornsey, M. J., & Gallois, C. (2006). Group-directed criticism in Indonesia: Role of message source and audience. *Asian Journal of Social Psychology, 9*, 96-102. doi:10.1111/j.1467-839X.2006.00186.x
- Becker, B. J., & Wu, M. J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science, 22*, 414-429.
- Branscombe, N. R. (1998). Thinking about one's gender group's privileges or disadvantages: Consequences for well-being in women and men. *British Journal of Social Psychology, 37*, 167-184.
- Branscombe, N. R., & Miron, A. (2004). Interpreting the ingroup negative actions towards another group: Emotional reactions to appraised harm. In L. Z. Tiedens & C. W. Leach (Eds.), *The social life of emotions* (pp. 314-345). Cambridge, UK: Cambridge University Press.
- Branscombe, N. R., Slugoski, B., & Kappen, D. M. (2004). The measurement of collective guilt. In N. R. Branscombe & B. J. Doosje (Eds.), *Collective guilt: International perspectives* (pp. 95-111). New York, NY: Cambridge University Press.
- 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.
- Center for Constitutional Rights. (2006). *Report on torture and cruel, inhuman, and degrading treatment of prisoners at Guantánamo Bay, Cuba*. New York, NY: Published online.
- Chow, R. M., Lowery, B. S., & Hogan, C. M. (2013). Appeasement: Whites' strategic support for affirmative action. *Personality and Social Psychology Bulletin, 39*, 332-345.
- Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to "prejudice." *Journal of Personality and Social Psychology, 88*, 770-789. doi:10.1037/0022-3514.88.5.770
- Doosje, B., Branscombe, N. R., Spears, R., & Manstead, A. S. R. (1998). Guilty by association: When one's group has a negative history. *Journal of Personality and Social Psychology, 75*, 872-886. doi:10.1037/0022-3514.75.4.872
- Doosje, B. J., Branscombe, N. R., Spears, R., & Manstead, A. S. R. (2004). Consequences of national ingroup identification for responses to immoral historical events. In N. R. Branscombe & B. J. Doosje (Eds.), *Collective guilt: International perspectives* (pp. 95-111). New York, NY: Cambridge University Press.
- Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgment and reactions: A world from two perspectives. *Psychological Inquiry, 6*, 267-285. doi:10.1207/s15327965pli0604\_1
- Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual Review of Psychology, 53*, 161-186. doi:10.1146/annurev.psych.53.100901.135228
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social & Personality Psychology Compass, 10*, 535-549.
- Guadagno, R. E., Rhoads, K. V. L., & Sagarin, B. J. (2011). Figural vividness and persuasion: Capturing the "elusive" vividness effect. *Personality and Social Psychology Bulletin, 37*, 626-638. doi:10.1177/0146167211399585
- Gunn, G. R., & Wilson, A. E. (2011). Acknowledging the skeletons in our closet: The effect of group affirmation on collective guilt, collective shame, and reparatory attitudes. *Personality and Social Psychology Bulletin, 37*, 1474-1487. doi:10.1177/0146167211413607
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*, 408-420. doi:10.1080/03637750903310360
- Hayes, A. F. (2013). *An introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.
- Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology, 67*, 451-470. doi:10.1111/bmsp.12028
- Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: Wiley.
- Hopkins, N., Reicher, S., Harrison, K., Cassidy, C., Bull, R., & Levine, M. (2007). Helping to improve the group stereotype: On the strategic dimension of prosocial behavior. *Personality and Social Psychology Bulletin, 33*, 776-788. doi:10.1177/0146167207301023
- Hornsey, M. J., Grice, T., Jetten, J., Paulsen, N., & Callan, V. (2007). Group-directed criticisms and recommendations for change: Why newcomers arouse more resistance than old-timers. *Personality and Social Psychology Bulletin, 33*, 1036-1048.
- Hornsey, M. J., Trembath, M., & Gunthorpe, S. (2004). "You can criticize because you care": Identity attachment, constructiveness, and the intergroup sensitivity effect. *European Journal of Social Psychology, 34*, 499-518. doi:10.1002/ejsp.212
- Iyer, A., Leach, C. W., & Pedersen, A. (2004). Racial wrongs and restitutions: The role of guilt and other group-based emotions. In *Off white: Readings on power, privilege, and resistance* (pp. 345-361). Cambridge, UK: Cambridge University Press.
- Iyer, A., Schmader, T., & Lickel, B. (2007). Why individuals protest the perceived transgressions of their country: The role of anger, shame, and guilt. *Personality and Social Psychology Bulletin, 33*, 572-587. doi:10.1177/0146167206297402
- Jetten, J., & Hornsey, M. J. (2014). Deviance and dissent in groups. *Annual Review of Psychology, 65*, 461-485. doi:10.1146/annurev-psych-010213-115151
- Kenny, D. A., & Judd, C. M. (2014). Power anomalies in testing mediation. *Psychological Science, 25*, 334-339.
- Knowles, E. D., Lowery, B. S., Chow, R. M., & Unzueta, M. M. (2014). Deny, distance, or dismantle? How white Americans manage a privileged identity. *Perspectives on Psychological Science, 9*, 594-609. doi:10.1177/1745691614554658
- Leidner, B., Castano, E., Zaiser, E., & Giner-Sorolla, R. (2010). Ingroup glorification, moral disengagement, and justice in the context of collective violence. *Personality and Social Psychology Bulletin, 36*, 1115-1129. doi:10.1177/0146167210376391
- Lickel, B., Schmader, T., Curtis, M., Scarnier, M., & Ames, D. R. (2005). Vicarious shame and guilt. *Group Processes & Intergroup Relations, 8*, 145-157. doi:10.1177/1368430205051064

- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.
- Lowery, B. S., Chow, R. M., Knowles, E. D., & Unzueta, M. M. (2012). Paying for positive group esteem: How inequity frames affect whites' responses to redistributive policies. *Journal of Personality and Social Psychology, 102*, 323-336. doi:10.1037/a0024598
- Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup emotions: Explaining offensive action tendencies in an intergroup context. *Journal of Personality and Social Psychology, 79*, 602-616. doi:10.1037/0022-3514.79.4.602
- Mashuri, A., van Leeuwen, E., & Hanurawan, F. (2016). How morality threat promotes reconciliation in separatist conflict: A majority group perspective. *Group Processes & Intergroup Relations*. Advance online publication. doi:10.1177/1368430216677302
- Miron, A. M., Branscombe, N. R., & Biernat, M. (2010). Motivated shifting of justice standards. *Personality and Social Psychology Bulletin, 36*, 768-779.
- Packer, D. J. (2008). On being both with us and against us: A normative conflict model of dissent in social groups. *Personality and Social Psychology Review, 12*, 50-73.
- Packer, D. J., & Chasteen, A. L. (2010). Loyal deviance: Testing the normative conflict model of dissent in social groups. *Personality and Social Psychology Bulletin, 36*, 5-18. doi:10.1177/1088868307309606
- Petz, J., Gunn, G. R., & Wilson, A. E. (2010). Crimes of the past: Defensive temporal distancing in the face of past in-group wrongdoing. *Personality and Social Psychology Bulletin, 36*, 598-611. doi:10.1177/0146167210364850
- Roccas, S., Klar, Y., & Livitain, I. (2006). The paradox of group-based guilt: Modes of national identification, conflict vehemence, and reactions to the in-group's moral violations. *Journal of Personality and Social Psychology, 91*, 698-711. doi:10.1037/0022-3514.91.4.698
- Rydell, R. J., Hugenberg, K., Ray, D., & Mackie, D. M. (2007). Implicit theories about groups and stereotyping: The role of group entitativity. *Personality and Social Psychology Bulletin, 33*, 549-558.
- Sahdra, B., & Ross, M. (2007). Group identification and historical memory. *Personality and Social Psychology Bulletin, 33*, 384-395. doi:10.1177/0146167206296103
- Sani, F., & Reicher, S. (1998). When consensus fails: An analysis of the schism within the Italian Communist Party (1991). *European Journal of Social Psychology, 28*, 623-645.
- Sani, F., & Reicher, S. (1999). Identity, argument and schism: Two longitudinal studies of the split in the Church of England over the ordination of women to the priesthood. *Group Processes & Intergroup Relations, 2*, 279-300. doi:10.1177/1368430299023005
- Sani, F., & Todman, J. (2002). Should we stay or should we go? A social psychological model of schisms in groups. *Personality and Social Psychology Bulletin, 28*, 1647-1655. doi:10.1177/014616702237646
- Schmitt, M. T., Miller, D. A., Branscombe, N. R., & Brehm, J. W. (2008). The difficulty of making reparations affects the intensity of collective guilt. *Group Processes & Intergroup Relations, 11*, 267-279.
- Shnabel, N., & Nadler, A. (2008). A needs-based model of reconciliation: Satisfying the differential needs of victim and perpetrator. *Journal of Personality and Social Psychology, 94*, 116-132. doi:10.1037/12061-021
- Shuman, E., Cohen-Chen, S., Hirsch-Hoefler, S., & Halperin, E. (2016). Explaining normative versus nonnormative action: The role of implicit theories. *Political Psychology, 37*, 835-852.
- Sullivan, D., Landau, M. J., Branscombe, N. R., Rothschild, Z. K., & Cronin, T. J. (2013). Self-harm focus leads to greater collective guilt: The case of the US-Iraq conflict. *Political Psychology, 34*, 573-587.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of group conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 7-24). Chicago, IL: Nelson-Hall.
- Tarrant, M., Branscombe, N. R., Warner, R. H., & Weston, D. (2012). Social identity and perceptions of torture: It's moral when we do it. *Journal of Experimental Social Psychology, 48*, 513-518.
- Tausch, N., Becker, J. C., Spears, R., Christ, O., Saab, R., Singh, P., & Siddiqui, R. N. (2011). Explaining radical group behavior: Developing emotion and efficacy routes to normative and non-normative collective action. *Journal of Personality and Social Psychology, 101*, 129-148.
- van Leeuwen, E., & Täuber, S. (2012). Outgroup helping as a tool to communicate in-group warmth. *Personality and Social Psychology Bulletin, 38*, 772-783. doi:10.1177/0146167211436253
- Vorauer, J. D., Main, K. J., & O'Connell, G. B. (1998). How do individuals expect to be viewed by members of lower status groups? Content and implications of meta-stereotypes. *Journal of Personality and Social Psychology, 75*, 917-937. doi:10.1037/0022-3514.75.4.917
- Vorauer, J. D., & Sakamoto, Y. (2008). Who cares what the outgroup thinks? Testing an information search model of the importance individuals accord to an outgroup member's view of them during intergroup interaction. *Journal of Personality and Social Psychology, 95*, 1467-1480. doi:10.1037/a0012631
- Wike, R., Stokes, B., Poushter, J., & Fetterolf, J. (2017, June 26). U.S. image suffers as publics around world question Trump's leadership. *Pew Research Center*. Retrieved from <http://www.pewglobal.org/2017/06/26/u-s-image-suffers-as-publics-around-world-question-trumps-leadership/>
- Wirtz, C., & Doosje, B. (2013). Reactions to threatening critical messages from minority group members with shared or distinct group identities. *European Journal of Social Psychology, 43*, 50-61.
- Wohl, M. J. A., & Branscombe, N. R. (2008). Remembering historical victimization: Collective guilt for current ingroup transgressions. *Journal of Personality and Social Psychology, 94*, 988-1006. doi:10.1037/0022-3514.94.6.988
- Wohl, M. J. A., Branscombe, N. R., & Klar, Y. (2006). Collective guilt: Emotional reactions when one's group has done wrong or been wronged. *European Review of Social Psychology, 17*, 1-37. doi:10.1080/10463280600574815