Not the ‘We’ I Wanted: How Images of In-Group Moral Hypocrisy Reduce In-Group Attachment and Support for Intergroup Helping

Thomas C. O’Brien
Bernhard Leidner

University of Massachusetts Amherst

Acknowledgements: This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program awarded to the first author under Grant No. 1451512. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Correspondence regarding this article should be sent to:

Thomas C. O’Brien
Tobin Hall, 135 Hicks Way
Department of Psychological and Brain Sciences
University of Massachusetts Amherst
Amherst, MA 01003

Word count: 8,036
Abstract

Drawing on research on social identity and social identity threat, two experiments show that an image of one’s in-group as morally hypocritical is a distinct type of social identity threat that leads low glorifiers to reduce their attachment to the in-group, and, as an unintended consequence, their support for intergroup helping. As expected, when Americans learned that other groups regard the U.S. as morally hypocritical, they reduced their in-group attachment and, in turn, their support for in-group interventions in other groups’ humanitarian crises (Experiment 1, N = 435). In a pre-post design, Experiment 2 (N = 423) showed that this phenomenon occurred specifically among low glorifiers, and is specific to the image of in-group hypocrisy (rather than immorality). Identifying moral hypocrisy as a unique social identity threat that low glorifiers respond to with reduced in-group attachment, this research contributes to the literatures on social identity (threat), group identification, and de-identification.

Keywords: social identity, social identity threat, glorification, moral hypocrisy, intergroup helping
Not the ‘We’ I Wanted: How Images of In-Group Moral Hypocrisy Reduce In-Group Attachment and Support for Intergroup Helping

At the time of humanitarian crises, the world often calls on powerful countries to help. With the exception of natural disasters, this help is often expected to come in form of interventions (militarily or diplomatically) in human rights violations (Evans & Sahnoun, 2002; Finnemore, 1996). Powerful countries do indeed often heed such calls for help, attempting to intervene in human rights abuses outside their own borders. In the last two years alone, the U.S. for example has condemned Russia’s military aggression against neighboring countries (Congressional Research Service, 2014a), China’s detention of ‘dissidents’ and human rights advocates (U.S. Department of State, 2015), and Iran’s mistreatment and persecution of religious minorities (Congressional Research Service, 2014b).

At the same time, however, powerful nations that involve themselves in other nations’ human rights violations often face accusations about the contradiction between their demands for how other countries should behave and their own behavior, whether by the governments they criticize (Reuters, 2014) or non-governmental organizations. Human Rights Watch, for example, has recently pointed out the hypocrisy of the U.S.’ condemnation of China, Egypt, Indonesia, Iran, Iraq, Jordan, Libya, North Korea, Saudi Arabia, Sri Lanka, and Turkey for using interrogation and detention methods that amount to torture— even after it became public that the U.S. had been using the same methods itself in the “war on terror” (Human Rights Watch, 2014; for a more general account of U.S. hypocrisy and its repercussions see Farrell & Finnemore, 2013).

From a psychological perspective, contradictions between a group’s stance on human rights in other countries and the group’s own behavior can create an image of moral hypocrisy of
the group in the eyes of other groups. Knowledge of this image creates a meta-image or intergroup metaperception (Frey & Tropp, 2006; O’Brien, Leidner, & Tropp, under review): a perception of how members of other groups view one’s own group. This meta-image of moral hypocrisy, we argue, poses a specific threat to the (social) identity of the group’s members. The threat, in turn, can lead to a decrease in in-group attachment, especially among those group members who would otherwise support their group’s intervention in foreign humanitarian crises. As in-group attachment is usually related to greater willingness to help out-groups (Leidner, in press), a decrease in it should also decrease people’s support for intervention in a crisis on behalf of out-group members. In other words, group-level accusations of moral hypocrisy, whether legitimate or not, might erode the accused’s willingness to engage in intergroup helping. Such a phenomenon would be particularly problematic because the provision of intergroup help relies heavily on the support of the members of the providing group in contexts such as foreign policy of democratic countries, where public opinion has a strong role shaping policy.

To test this hypothesis, we conducted two experiments investigating whether meta-images of U.S. hypocrisy decrease in-group attachment among Americans and, in turn, their support for the U.S. to engage in diplomatic interventions in foreign crises. In doing so, we take research on moral hypocrisy from the individual to the group level, shifting its focus from people’s reactions to another person’s moral hypocrisy (e.g., Batson, Kobrynnowicz, Dinnerstein, Kampf, & Wilson 1997; Monin & Merritt, 2012) to people’s reactions to their own group’s (purported) moral hypocrisy. This shift in focus also brought a social identity perspective to moral hypocrisy, viewing it as a form of social identity threat. Besides integrating (meta-images of) moral hypocrisy with social identity theory, we also tie it to intergroup helping by investigating its effects on support for intergroup help.
Moral hypocrisy as a social identity threat

People view the groups they belong to as extensions of their self (Tajfel & Turner, 1979). People thus derive part of their (social) identity from groups, and are consequently motivated to view these in-groups positively (Tajfel & Turner, 1979)—be it in terms of the group’s competence or sociability/warmth (Fiske, Cuddy, Glick, & Xu, 2002) or, especially so, the group’s moral character (Leach, Ellemers, & Barreto, 2007). Hence, when the in-group’s morality is in question, this poses a threat to its members’ identity (Aquino, Reed, Thau, & Freeman, 2007; Castano & Giner-Sorolla, 2006; Leidner & Castano, 2012). While immoral or morally questionable behavior by itself is the most common source of this threat, moral hypocrisy, or accusations thereof, should magnify this threat and therefore also the reactions to it. The reason is that moral hypocrisy consists of two components: the in-group’s own immoral or morally questionable behavior, paired with the in-group’s simultaneous criticism of another group’s immoral or morally questionable behavior. This combination makes the image of moral hypocrisy especially difficult to change. Whereas immorality often is a judgment of (instances of) behavior and therefore of relatively malleable attributes, hypocrisy is by definition a judgment of character and therefore of stable attributes. Thus, accusations of moral hypocrisy go right to the heart of the group’s image and its members’ identity, and the outlook of convincing out-group members that their accusations are false is dire. These qualities of meta-images of moral hypocrisy should thus induce a social identity threat that makes it particularly challenging to respond to effectively.

Responses to social identity threat

When people encounter social identity threat, they generally respond in one of three ways (Leidner, Li, & Kardos, 2015): by engaging in defensive strategies that protect the in-group and
For Peer Review

NOT THE ‘WE’ I WANTED

its image (e.g., Bandura, 1999, 2002; Glasford, Pratto, & Dovidio, 2008; Leidner, Castano, Zaiser, & Giner-Sorolla, 2010; Roccas, Klar, & Livian, 2006); by non-defensively trying to repair the in-group’s moral image through acts that counteract or disprove those aspects of the group that have been threatened (Doosje, Branscombe, Spears, & Manstead, 1998; Iyer, Schmader, & Lickel, 2007); or by reducing identification with the group, lessening the need to view it positively (Becker & Tausch, 2014; Glasford et al., 2008). Not surprisingly, which of these ways a person will take depends largely on in-group identification. More specifically, it should depend on in-group glorification. In contrast to in-group attachment, which is characterized by a person’s commitment to the group and the importance the person assigns to the group, in-group glorification is characterized by a person’s belief that the group is superior over other groups as well as by the person’s deference to group norms and authorities (Roccas et al., 2006; Roccas, Sagiv, Schwartz, Halevy, & Eidelson, 2008).

Low glorifiers. People who are generally open to their group providing help to other groups are more likely to be found among those who do not glorify the group. Even when confronted with violence committed by the in-group, low glorifiers embrace harm and fairness principles (which lead to assessments of in-group immorality) more strongly than high glorifiers (Leidner & Castano, 2012), and report more guilt than high glorifiers (Roccas et al., 2006), which has been identified as a major predictor of reparatory motives (Doosje et al., 1998). In fact, low glorifiers are more likely than high glorifiers to demand compensation for victims of ingroup-committed violence (Leidner et al., 2010). Against the background of these findings it is not surprising, then, that low glorifiers also generally respond less defensively to social identity threat than high glorifiers (for a review see Leidner et al., 2015), engaging less in ingroup-exonerating cognitions (Roccas et al., 2006), dehumanization of out-group victims (Castano &
Giner-Sorolla, 2006; Leidner et al., 2010) or minimization of their emotions (Leidner et al., 2010), and denial of ingroup wrongdoing (Bilali, 2013). In other words, rather than defending the in-group, more often than not low glorifiers respond to social identity threat with in-group criticism and attempts to repair and improve the in-group and its image.

Yet, as explained earlier, the complicated nature of improving a morally hypocritical in-group image in the eyes of others is a daunting task, even more so than fixing an immoral image. Thus in the case of meta-images of moral hypocrisy, this strategy should be less effective to counter the social identity threat, and therefore low glorifiers should be less likely to use it. In line with this hypothesis, people are less likely to work to change negative attributes of themselves or others when they perceived the attributes in question as stable rather than malleable (e.g., Dweck, Chiu, & Hong, 1995).

Given that low glorifiers tend not to use ingroup-defensive strategies (Bilali, 2013; Leidner & Castano, 2012; Leidner et al., 2010; Roccas et al., 2006), low glorifiers must then resolve the dissonance between their motivation to be part of a group that is morally sincere and the group’s reputation as a moral hypocrite differently. We hypothesized they would do so by reducing their attachment to the in-group. This hypothesis is in line with research showing that people reduce their identification with their group when their identity is threatened by the group’s violation of moral values (Glasford, Pratto, & Dovidio, 2008). By reducing in-group attachment, the social identity threat posed by meta-images of moral hypocrisy will be dealt with effectively. After all, if a group member lowers their commitment to the group and attaches less importance to it, then the group’s behavior and character will be less consequential for the group member’s self and identity.
High glorifiers. Unlike low glorifiers, high glorifiers do not really have the option to reduce their attachment to the group to protect themselves against the identity threat of meta-images of moral hypocrisy. With their inflated image of the in-group and adherence to viewing it as morally superior, letting go of the group would be too psychologically costly for high glorifiers (Leidner et al., 2015). Instead, high glorifiers respond to social identity threat by defending the in-group and maintaining the positive image they hold of it, be it through engaging in exonerating cognitions (Roccas, Klar, & Liviatin, 2006), shifting the relative importance they place on different moral principles (Leidner & Castano, 2012), dehumanizing the out-group victims (Castano & Giner-Sorolla, 2006) or minimizing their suffering (Leidner, Castano, Zaiser, & Giner-Sorolla, 2010), or attributing more responsibility for intergroup violence to the out-group than to the in-group (Bilali, 2013). Thus, high glorifiers likely also employ psychological defense strategies in response to allegations that their group is a moral hypocrite. While both low and high glorifiers use their respective strategies to protect their psychological equanimity, unlike low glorifiers, high glorifiers should not decrease their attachment to the group. As a result of their well-established tendency to engage in in-group defense, we hypothesized that if anything, they should reaffirm their commitment to the group and therefore increase their attachment to it.

Consequences for support for intergroup help

As a side effect of the expected reduction in low glorifiers’ in-group attachment in response to meta-images of moral hypocrisy, we also expected a reduction in their support for their own group to help other groups. While low glorifiers’ support for helping other groups through individual-level means may remain unchanged, their support for helping other groups through group-level means should decrease, as motivation for changing group behavior is most
relevant for those who feel close to the group (e.g., Packer, 2007). Even if they may still engage in individual-level helping, this hypothesized decrease in group-level helping would cripple any hope for stopping human rights crises. Although individuals may have the power to voice their opposition to human rights violations through petitions and their personal choices, powerful nations such as the U.S. have the power to severely pressure other governments that violate human rights, such as by prohibiting U.S. citizens or companies to conduct any trade with that other government, or by proposing international agreements to prohibit trade on those countries (e.g., sanctions; see Baek, 2008; Martel & Leidner, under review). Thus, to the extent that such measures can be effective, citizens of a democracy have more power to advance human rights abroad by advocating for their country to take actions than by restricting their activism to what they can do as individuals.

While high glorifiers were not the focus of the studies reported here, we predicted that high (unlike low) glorifiers, as a result of their non-decrease or even increase in in-group attachment, would not respond to the social identity threat posed by meta-images of the in-group’s moral hypocrisy with a decrease in support for their own group providing help to other groups. While it was rather unclear whether the expected increase in attachment should lead to an increase in support for intergroup helping or not, we predicted the latter (i.e., no increase). This prediction was based on research showing that high glorifiers – unlike low glorifiers – prioritize values of security and conservation over those of universalism and benevolence (Roccas et al., 2010). Consequently, high glorifiers construe help for out-groups generally as a cost to, and certainly not a benefit for, the in-group. Supporting this contention, high glorifiers are rather unwilling to intervene in a humanitarian out-group crisis unless intervention is likely to also result in long-term benefits for the in-group (Leidner, in press).
In sum, then, in-group attachment should be positively related to support for intergroup helping when defined against the background of low glorifiers’ values of universalism and benevolence, but not when defined against the background of high glorifiers’ values of security and conservation. Consequently, when low glorifiers decrease their attachment to the group, they should also decrease their support for intergroup helping, whereas when high glorifiers increase their attachment to the group, they should not necessarily increase their support for intergroup helping. The full conceptual model we hypothesized is depicted in Figure 1.

![Diagram](image)

**Figure 1.** A morally hypocritical in-group image is hypothesized to reduce low (but not high) glorifiers’ support for intergroup helping by reducing their attachment to the in-group.

**Study 1**

Before testing the full (moderation) hypothesis in Study 2, Study 1 tested our main hypothesis that people can respond to meta-images of in-group hypocrisy by reducing their attachment to the group, which, in turn, reduces their support for the in-group to help other groups. To do so, we informed American participants that people from around the world viewed the U.S. as morally hypocritical (hypocrisy condition) or not (baseline condition). In a third condition, we informed participants that others view the U.S. as morally sincere (sincerity
condition), allowing us to test whether any differences between conditions were driven by hypocrisy or sincerity.

**Method**

*Participants*

We recruited 493 adult participants from Amazon’s MTurk. Sixteen participants were excluded for not indicating that they were born in the U.S. or that English was their first language. Another 16 participants were excluded for guessing the purpose of the study. Six participants were excluded for responding incorrectly to a question about the information in the manipulation. Another thirty-two participants were excluded for taking more than 35 minutes to complete the study, less than five minutes in the control condition, or less than seven minutes in the experimental conditions, as they were statistical outliers (while the mean for the remainder of the sample was $M = 12.44, SD = 4.52$). This left a remainder of 423 participants, including 192 males and 231 females, ranging in age from 18-73 ($M = 36.89, SD = 13.39$).

*Procedure*

Participants were randomly assigned to one of three conditions. In the hypocrisy and sincerity conditions, they were shown an alleged Gallup poll that asked respondents from various countries 1) Whether the U.S. violates or respects human rights more than the countries it often criticizes for doing so; and 2) Whether the U.S. is more or less aggressive than the countries it often criticizes for being aggressive. The alleged average responses to the three questions were illustrated on two graphs. In the hypocrisy condition, participants received information demonstrating that people from other countries considered the U.S. to be a moral hypocrite in the sense that the poll reported that 71% of respondents said saw the U.S. as violating human rights more than other countries it criticizes for violating human rights, while 18% disagreed and 11%
had no opinion. Further, the poll reported that 77% of respondents said the U.S. was more aggressive than other countries it criticizes for aggression, while 14% disagreed and 9% had no opinion. In the sincerity condition, participants received information demonstrating that people from other countries considered the U.S. to be morally sincere in the sense that the poll reported that 71% of respondents saw the U.S. as respecting human rights more than other countries it criticizes for violating human rights, while 18% disagreed and 11% had no opinion. Further, the poll reported that 77% of respondents said the U.S. was less aggressive than other countries it criticizes for aggression, while 14% disagreed and 9% had no opinion. The alleged poll was reported to have included respondents from England, Russia, Qatar, and South Africa, thus reducing the extent to which participants could construe the poll as coming distinctly from a high or low power country, or from one particular part of the world.

Attention Check

Following the manipulation, participants in the two experimental condition were asked on 1-9 visual analog scales about the Gallup poll they had just read, ranging from “Not true at all” to “Absolutely true”. In the hypocrisy condition, participants were asked to respond to the statement, “According to the Gallup poll, the majority of the world sees the U.S. as morally hypocritical, criticizing other countries while not behaving any better itself.” In the sincerity condition, participants were asked to respond to the statement, “According to the Gallup poll, the majority of the world sees the U.S. as morally sincere, criticizing other countries while being a role model itself.” Six participants were excluded because they responded below 5 (the midpoint) on the scale, as analyses indicated they were univariate outliers.

Measures
Unless otherwise specified, all measures were assessed with visual analog scales ranging from 1-9.

*In-group attachment* (α = .95, M = 6.61, SD = 1.89) and *glorification* (α = .89, M = 4.93, SD = 1.73) were measured with scales by Roccas et al. (2006).

*Support for diplomatic intervention in out-group human rights violations.* Participants were given the following information about human rights violations in Uzbekistan and Mexico:

_Uzbekistan has repeatedly been violating human rights by suppressing political dissidents. In at least one instance, the president of Uzbekistan ordered the military to open fire on protesters. The government has even boiled some dissidents alive._

_Mexico has been violating human rights by allowing grave abuses by security forces, such as the brutal killing of 43 students. The state prosecutor's office detained witnesses, beat them, and put plastic bags over their head to force them to say that the military was not involved._

Following the scenarios and the sentence stem “To get the government of [Uzbekistan/Mexico] to improve human rights, the U.S. would do best to…”, participants responded to two statements for each scenario, on a scale from “Strongly oppose” to “Strongly support”: ...use “tough” diplomacy (e.g., sanctions against the [out-group country] government), and ...do not get involved in the issue (reverse-scored). The altogether four items formed a reliable scale and were thus averaged together (α = .79, M = 6.09, SD = 1.98).

**Results**

All dependent variables were subjected to a one-way analysis of variance with condition as the independent variable.
Support for diplomatic intervention in out-group human rights violations. Participants differed significantly across conditions in their support for intervention, $F(2, 420) = 3.22, p = .041, \eta_p^2 = .015$. Participants in the hypocrisy condition ($M = 5.72, SD = 2.04$) expressed significantly less support than participants in the baseline condition ($M = 6.23, SD = 1.98$), $t(420) = -2.21, p = .027$, or participants in the sincerity condition ($M = 6.25, SD = 1.85$), $t(420) = -2.22, p = .027$. The baseline and sincerity conditions did not differ from each other, $t(420) = 0.10, p = .917$.

In-group attachment and glorification. Participants differed marginally across conditions in levels of attachment, $F(2, 420) = 2.57, p = .078, \eta_p^2 = .012$. As predicted, participants in the hypocrisy condition ($M = 6.30, SD = 1.90$) reported significantly lower levels of attachment compared to participants in the baseline condition ($M = 6.77, SD = 1.83$), $t(420) = -2.13, p = .034$, and marginally lower levels of attachment than participants in the sincerity condition ($M = 6.72, SD = 1.92$), $t(420) = -1.79, p = .074$. Participants in the sincerity condition did not differ from participants in the baseline condition, $t(420) = -0.25, p = .802$.

Participants differed significantly across conditions in levels of glorification, $F(2, 420) = 4.02, p = .019, \eta_p^2 = .019$. Participants in the hypocrisy condition ($M = 4.58, SD = 1.63$) glorified their in-group significantly less than participants in the baseline condition ($M = 5.11, SD = 1.79$), $t(420) = -2.64, p = .009$, or participants in the sincerity condition ($M = 5.06, SD = 1.72$), $t(420) = -2.27, p = .024$. Participants in the sincerity condition did not differ from participants in the baseline condition, $t(420) = -2.27, p = .024$.

Mediation. Since the hypocrisy condition reduced attachment, glorification, and support for intervention, we tested whether the reduction in attachment, or the reduction in glorification, or both, would mediate the effect of moral hypocrisy on support for intervention in human rights
violations. We created dummy variables for each experimental condition (hypocrisy and sincerity), assigning participants in the hypocrisy condition a value of ‘1’ and participants in the other conditions a value of ‘0’ for the hypocrisy dummy variable, and participants in the sincerity condition a value of ‘1’ and participants in the other conditions a value of ‘0’ for the sincerity dummy variable. Then, we ran Hayes’s Process Model 4 (2013) using 5,000 bootstrap samples, with the hypocrisy dummy variable predicting support for intervention via attachment as mediator in the first analysis; via glorification as mediator in the second analysis; and via attachment and glorification as mediators operating in parallel in a third analysis.

The indirect effect with attachment as the only mediator was significant, $b = -.07$, $SE = .04$, CI95% [-.1889, -.0095]; the direct effect of the hypocrisy condition on support for intervention remained significant, $b = -.45$, $SE = .20$, $t(420) = -2.21$, $p = .028$, CI95% [-.8552, -.0497]. The indirect effect with glorification as the only mediator was not significant, $b = -.05$, $SE = .04$, CI95% [-.1529, .0004]. In the parallel mediation model, the indirect effect via attachment was significant, $b = -.08$, $SE = .05$, CI95% [-.2145, -.0061], whereas the indirect effect via glorification was not, $b = .01$, $SE = .04$, CI95% [-.0675, .1025].

**Discussion**

Learning that out-groups view the in-group as morally hypocritical (rather than morally sincere) reduced participants’ support for intergroup help. Supporting our argument that moral hypocrisy poses a social identity threat, this effect was driven by moral hypocrisy, as the baseline condition differed from the hypocrisy but not from the sincerity condition, and by the reduction in in-group attachment it triggered. While moral hypocrisy also led to a decrease in in-group glorification, only the decrease in in-group attachment explained why moral hypocrisy decreased support for intervention. The effect of the hypocrisy condition on in-group attachment was
marginal, but this may be because Study 1 did not assess in-group attachment and glorification before the manipulation and therefore could not test whether the effect and mechanism observed in Study 1 occurred only among low but not high glorifying group members. If the effect is specific to low glorifiers, it is not surprising that the main effect across all participants does not reach conventional levels of significance. Further, Study 1 did not address the question of whether the observed effects were due to in-group hypocrisy in particular, as we had hypothesized, or due to in-group immorality in general (which is part of in-group hypocrisy as well). These limitations of Study 1 were thus followed up in Study 2.

**Study 2**

We designed Study 2 to replicate findings of Study 1 and extend them in three ways. First, we added a fourth condition to distinguish the effects of learning about out-groups viewing the in-group as morally hypocritical from learning about out-groups viewing the in-group as “merely” immoral. Second, we measured glorification and attachment before – as moderators – and after – as mediators – the manipulation. This way, we could test whether in-group glorification would moderate and in-group attachment mediate the effect of moral hypocrisy on support for intergroup helping. Finally, we expanded our measure of support for diplomatic intervention in human rights violations to strengthen the generalizability of our findings.

**Method**

*Participants.* Six-hundred and thirty-six participants were recruited from Amazon’s Mechanical Turk (MTurk). Seventy-three participants were deleted either because they did not belong to our target population (English was not their second language and/or they were not U.S. citizens). An additional sixty-four participants were deleted because they failed to correctly summarize the information from the manipulation materials. An additional 49 participants were
deleted for failing the attention check (described below), assessed directly after the manipulation. Another 18 participants were deleted as univariate outliers on the amount of time they took to read the manipulation text or for taking less than 10 seconds to read the manipulation, leaving a total of 435 participants (ages 19-76; \( M \) of age = 36.31, \( SD \) = 12.72), including 168 males and 267 females. The proportion of the sample we excluded fell within normal range for online studies and has been deemed necessary to ensure data quality (e.g., Chandler, Mueller, & Paolacci, 2014).

Procedure. The experiment included four conditions: three conditions similar to Study 1, and a fourth condition in which participants learned that out-groups view the in-group as immoral. This immorality condition was almost identical to the hypocrisy condition, but used the wording “immoral” rather than “morally hypocritical” when describing how outgroup members view participants’ ingroup.

Attention check. Directly after the manipulation, participants in each of the experimental conditions were asked one of three questions depending upon which experimental condition they were in. In each condition, participants were asked to report on a sliding scale from Not true at all to Absolutely true what the Gallup poll reported, either that the majority of people in other countries viewed the U.S. as morally hypocritical, that they viewed the U.S. as immoral, or that they viewed the U.S. as morally sincere. Sixty-four participants were excluded from analyses for failing this attention check, indicating that they did not pay sufficient attention to the manipulation materials.

Measures. As in Study 1, participants responded to measures on 1-9 visual analog scales. Items to assess glorification and attachment were the same across the two studies, but in Study 1 they were split between Time 1 (before the manipulation) and Time 2 (after the manipulation).
As glorification consists of two aspects, deference and superiority (Roccas et al., 2008), we ensured that of the altogether four deference and four superiority items, two of each kind were used at Time 1 and the other two at Time 2. The same is true for the two aspects of attachment, commitment and importance. We did thus not administer identical items at both time points, to avoid that participants would be motivated to answer the Time 2 items consistent with the Time 1 items rather than indicating their actual level of glorification at Time 2. Yet, since all items usually load on the same factor (Roccas et al., 2006, 2008), they can be considered to be conceptually equivalent across time points.

Time 1 Glorification ($\alpha = .78$, $M = 5.31$, $SD = 1.69$) and Time 1 Attachment ($\alpha = .91$, $M = 6.20$, $SD = 1.93$). Two items assessing deference (It is disloyal for Americans to criticize the United States and One of the important things we have to teach our children is to respect our nation’s leaders) and two items assessing beliefs in in-group superiority (The U.S. military is the best military in the world and Other nations can learn a lot from us) were combined to assess glorification before the manipulation. Two items assessing importance of the in-group (I love the United States and It is important for me that everyone sees me as an American) and two items assessing commitment to the in-group (It is important for me to contribute to my nation and Being an American is an important part of my identity) were combined to assess attachment before the manipulation.

Time 2 Glorification ($\alpha = .82$, $M = 4.32$, $SD = 1.74$) and Time 2 Attachment ($\alpha = .88$, $M = 6.45$, $SD = 1.82$). Two items assessing deference (There is generally a good reason for every rule and regulation made by our national authorities and In today’s world, the only way to know what to do is to rely on our nation’s leaders) and two items assessing beliefs in in-group superiority (Relative to other nations, we are a very moral nation, and The U.S. is better than
other nations in all respects) were used to assess glorification at Time 2. Two items assessing importance of the in-group (It is important for me to view myself as an American, and When I talk about Americans, I usually say “we”, rather than “they”) and two items assessing commitment to the in-group (I am strongly committed to the United States, and It is important for me to help my country) were used to assess attachment at Time 2.

Support for diplomatic intervention in out-group human rights violations (α = .88, M = 5.26, SD = 1.93). Participants responded to the same items as in Study 1. The text of the scenario describing the situation in Mexico was slightly altered, such that the sentence beginning with “The state prosecutor’s…” was changed to, “In this recent example, the state prosecutor’s….“ In addition to the Mexico and Uzbekistan scenarios from Study 1, we added a third scenario regarding human rights in Egypt:

Egypt has repeatedly violated human rights by torturing detainees and arresting tens of thousands of political dissidents without trials. In at least one case where there was a trial, it lasted less than an hour and the government of Egypt sentenced 529 people to death, alleging that they were involved in an attack on a police station.

Results

Even though we measured Time 1 attachment and glorification before the manipulation, we submitted them to an analysis of variance with condition as independent variable (IV) to ensure that they were indeed not affected by condition. Neither Time 1 attachment (p = .253) nor Time 1 glorification (p = .795) were, allowing us to use them as IVs in subsequent analyses, as we had intended.
**Analytical strategy.** Since our theoretical interest was in the changes in attachment independent of glorification and any effect that condition may have had on glorification, we submitted Time 2 attachment to a general linear model (GLM) with the IVs of condition, Time 1 glorification, and the critically predicted interaction between condition and Time 1 glorification. Since glorification and attachment are two aspects of the same broader construct (i.e., in-group identification) and therefore share variance, we also controlled for Time 1 attachment. Since we hypothesized a change in attachment, we had to also control for changes in glorification; thus, we also entered Time 2 glorification and the two-way interaction between Time 1 and Time 2 glorification as covariates. Last but not least, since Time 2 glorification was itself affected by condition (as reported further below), to control for it properly we also had to control for its interaction with condition (see Yzerbyt, Muller, & Judd, 2004). To test whether it is really Time 1 glorification and not Time 1 attachment that moderated the effects, we also submitted Time 2 attachment to an alternative GLM with the IVs of condition, Time 1 attachment, and the interaction between condition and Time 1 attachment, and the covariates of Time 1 glorification, Time 2 glorification, and the two-way interactions between Time 1 and Time 2 glorification, and Time 2 glorification and condition. We followed the same strategy to analyze effects on support for intervention.

**Time 2 Attachment.** The interaction between condition and Time 1 glorification was significant, $F(3, 421) = 4.90, p = .002, \eta^2_p = .034$. Low glorifiers (1 SD below the mean on Time 1 glorification), reported less Time 2 attachment in the hypocrisy condition ($M = 5.94$) than in the immorality condition ($M = 6.62$), $t(418) = -3.00, p = .003$, the sincerity condition ($M = 6.58$), $t(418) = -2.98, p = .003$, and the baseline condition ($M = 6.46$), $t(426) = -2.59, p = .010$. High glorifiers (1 SD above the mean on Time 1 glorification), on the other hand, reported more Time
attachment in the hypocrisy condition ($M = 6.86$) than in the immorality condition ($M = 6.24$), $t(426) = 2.98, p = .003$, and, marginally so, than in the sincerity condition ($M = 6.45$), $t(426) = 1.68, p = .094$, and the baseline condition ($M = 6.51$), $t(426) = 1.76, p = .077$. No other comparisons were significant. Further, Time 1 attachment significantly predicted higher Time 2 attachment, $F(1, 421) = 557.27, p < .001, \eta_p^2 = .570$. No other effects were significant, $Fs < 2.60, ps > .050, \eta_p^2s < .02$.

In the alternative GLM with Time 1 attachment as the moderator, the interaction with condition was significant, $F(3, 421) = 4.79, p = .003, \eta_p^2 = .033$. Weakly attached people (1 SD below the mean of Time 1 attachment) reported less attachment in the hypocrisy condition ($M = 4.55$) relative to the baseline ($M = 5.02$), $t(421) = -2.51, p = .012$, immorality ($M = 5.11$), $t(421) = -2.83, p = .005$, and sincerity conditions ($M = 5.05$), $t(421) = -2.83, p = .005$. Strongly attached people (1 SD above the mean of Time 1 attachment) reported more Time 2 attachment in the hypocrisy condition ($M = 8.21$) than in the immorality condition ($M = 7.68$), $t(421) = 2.71, p = .007$. No other contrasts were significant. Time 1 attachment predicted higher levels of Time 2 attachment $F(1, 421) = 541.57, p < .001, \eta_p^2 = .563$. No other main effects or interactions were significant.

**Support for diplomatic intervention in out-group human rights violations.** The main effect of condition was significant, $F(3, 421) = 4.06, p = .007, \eta_p^2 = .028$. Participants reported more support in the sincerity condition ($M = 5.90$) than in the baseline condition ($M = 5.09$), $t(421) = 3.24, p = .001$, the hypocrisy condition ($M = 5.08$), $t(421) = 2.86, p = .005$, and the immorality condition ($M = 5.05$), $t(421) = 2.56, p = .011$; other contrasts were not significant. Time 2 glorification predicted higher levels of support, $F(1, 421) = 4.55, p = .034, \eta_p^2 = .011$. Most importantly, the interaction between condition and Time 1 glorification was also

John Wiley & Sons, Ltd
significant, \(F(3, 421) = 5.06, p = .002, \eta^2_p = .035\). Low glorifiers reported less support in the hypocrisy condition (\(M = 4.66\)) compared to the immoral condition (\(M = 6.09\)), \(t(421) = -2.96, p = .003\), and sincerity condition (\(M = 5.58\)), \(t(421) = -2.03, p = .043\), but not compared to baseline (\(M = 4.96\)), \(t(421) = .71, p = .476\). In contrast, high glorifiers reported less support in the immorality condition (\(M = 4.44\)) than in the baseline (\(M = 5.34\)), \(t(421) = -2.24, p = .026\), hypocrisy (\(M = 5.74\)), \(t(421) = -2.89, p = .004\), and sincerity condition (\(M = 6.35\)), \(t(421) = -3.77, p < .001\). The sincerity condition also differed from baseline, \(t(421) = 2.10, p = .037\). No other comparison was significant.

In the alternative model with Time 1 attachment as the moderator of the condition, the interaction between condition and Time 1 attachment was not significant, \(F(3, 421) = 0.93, p = .426, \eta^2_p = .01\). The main effect of condition was significant, \(F(3, 421) = 5.21, p = .002, \eta^2_p = .04\). Participants in the sincere condition (\(M = 4.69\)) supported intervention more than participants in the baseline condition (\(M = 4.35\)), \(t(421) = 3.57, p < .001\), the immoral condition (\(M = 4.10\)), \(t(421) = 3.27, p = .001\) and the hypocrite condition (\(M = 4.16\)), \(t(421) = 2.86, p = .004\). Time 1 and Time 2 attachment interacted to predicted support, \(F(1, 421) = 4.43, p = .036, \eta^2_p = .010\). For participants low (one standard deviation below the mean) on Time 1 attachment, higher levels of Time 2 attachment significantly predicted more support, \(b = .50, SE = .20, t(421) = 2.56, p = .011\). In contrast, among participants with high levels of Time 1 attachment (one standard deviation above the mean), Time 2 attachment did not predict support, \(b = .18, SE = .23, t(421) = 0.78, p = .437\).

**Time 2 Glorification.** To test effects on Time 2 glorification, we ran a GLM parallel to the other GLMs, but replacing Time 1 and Time 2 glorification with Time 1 and Time 2 attachment. Thus, we tested the main effects of condition, Time 1 attachment (just as we used
Time 1 glorification in other analyses), Time 2 attachment (just as we controlled for Time 2 glorification in other analyses), and Time 1 glorification (just as we controlled for Time 1 attachment in other analyses), in addition to interactions between condition and Time 1 attachment (just as we tested interaction between condition and Time 1 glorification in other analyses), condition and Time 2 attachment (just as we controlled for interactions between condition and Time 2 glorification in other analyses), and Time 1 and Time 2 attachment (just as we controlled for interactions between Time 1 and Time 2 glorification in other analyses). The main effect of condition was significant, $F(3, 421) = 4.86, p = .003, \eta^2_p = .03$. Participants reported more Time 2 glorification in the sincere condition ($M = 4.69$) than in the baseline condition ($M = 4.35$), $t(421) = 2.95, p = .003$, the hypocrite condition ($M = 4.16$), $t(421) = 2.92$, $p = .004$, and the immoral condition ($M = 4.10$), $t(421) = 3.49, p = .001$; other contrasts were not significant. Time 1 glorification predicted higher levels of Time 2 glorification $F(1, 421) = 172.28, p < .001, \eta^2_p = .290$. The interaction between condition and Time 1 attachment was not significant, $F(3, 421) = 1.78, p = .150, \eta^2_p = .013$. The interaction between condition and Time 2 attachment was marginal, $F(3, 421) = 2.42, p = .065, \eta^2_p = .017$.

Mediation. We conducted a moderated mediation analysis using Hayes’s Process Model 8 (2013). We created dummy variables for each condition, with a value of ‘1’ for participants in that condition, and a value of ‘0’ for participants not in that condition. In the first analysis, we entered the dummy variable for the hypocrisy condition as the independent variable, with Time 1 glorification as a moderator, Time 2 attachment as mediator, and support for intervention as a dependent variable, controlling for Time 1 attachment, Time 2 glorification, two-way interactions between the hypocrite condition and Time 2 glorification, and between Time 1 and Time 2 glorification. Supporting our hypothesis that in-group hypocrisy decreases support for
intervention among low glorifiers by reducing their in-group attachment, the indirect effect was significant at low levels of glorification, $b = -.12, \ SE = .09, \ CI95\% \ [-.3462, -.0032]$. At high levels of glorification, the indirect effect was significant in the opposite direction, $b = .09, \ SE = .07, \ CI95\% \ [.0022, .2725]$. The index of moderated mediation was significant, $b = .11, \ SE = .07, \ CI95\% \ [.0014, .2913]$. We also tested this model with all combinations of the other dummy variables as covariates to make our analyses more closely resemble the GLMs. These mediation analyses showed similar effects regardless of whether any combination of the other dummy variables were included as covariates, indicating that our hypothesized mechanism underlies the difference in support between the hypocrisy condition and each and every one of the control conditions.

**Discussion**

As hypothesized, Study 2 found that the effect of meta-images of in-group hypocrisy on support for intergroup helping through in-group attachment is indeed specific to in-group hypocrisy (and not in-group immorality in general), as well as to low glorifiers. However, while low glorifiers’ support for intervention did differ between hypocrisy and immorality as well as sincerity, it did not differ between hypocrisy and baseline. The likely reason is that by measuring glorification before the manipulation, we may have affected participants’ baseline responses. Although analyses found partial support for a model in which Time 1 attachment interacted with condition to reduce Time 2 attachment, our hypothesized model is more consistent with past research showing that glorification is the form of identification that makes people less receptive to negative information about their ingroup (e.g., Leidner et al., 2010; Leidner & Castano, 2012; Roccas et al., 2006, 2008). Furthermore, Time 1 attachment did not significantly interact with the
condition to predict support for interventions. Thus, our hypothesized model is the one with the most theoretical support and the most support from the data.

Taken together, the two studies support the hypothesis that learning about out-groups viewing the in-group as morally hypocritical reduces support for human rights interventions by first reducing in-group attachment, particularly among those who would generally be rather supportive of human rights interventions (i.e. low glorifiers).

**General Discussion**

Together, two experiments investigated how people react when they realize that their group is seen as a moral hypocrite by other groups. In Study 1, such a meta-image of in-group hypocrisy reduced group members’ perceived importance of and commitment to the in-group. While it also reduced people’s glorification of the in-group, only the reduction in in-group attachment (but not glorification) led to a subsequent decrease in support for intergroup helping. Illuminating the boundary conditions of this phenomenon, Study 2 demonstrated that this effect and its underlying mechanism only occur among group members who did not glorify their group to begin with, and only in response to (meta-images of) in-group hypocrisy but not more general in-group immorality. In doing so, this research furthers our understanding of how people respond to different kinds of social identity threat (in-group hypocrisy vs. in-group immorality), when people turn their back on the in-group rather than constructively criticizing it (de-identification vs. in-group criticism), and why different people respond the way they do (high vs. low glorifiers). On a more applied level, this research helps explain why group members who we usually assume should advocate intervention in other groups’ humanitarian crises (i.e., low glorifiers, liberals) often seem to respond with apathy instead, failing to facilitate intervention in human rights violations.
Criticizing the in-group for immorality, leaving it for hypocrisy

Study 2 in particular demonstrates that in-group hypocrisy poses a powerful social identity threat that appears to be rather distinct from other types of social identity threat. Probably the most widely researched social identity threat, in-group immorality has been shown to be defended or condemned, depending on the strength and quality of people’s identification with the in-group: high glorifiers defend, low glorifiers condemn (e.g., Bilali, 2013; Leidner & Castano, 2012; Leidner et al., 2010; Roccas et al., 2006). In-group hypocrisy, however, does not seem to be condemned so much as it is run from. Rather than sticking with the in-group and trying to improve and redeem it, low glorifiers lower their attachment to it. This severance of ties with the group disarms the social identity threat posed by the group’s hypocrisy.

These different correlates of in-group hypocrisy and immorality are likely grounded in the different qualities of hypocrisy and immorality. Whereas immorality, at least as it has been researched and is usually encountered in the real world, is usually a judgment of behavior, hypocrisy is by definition a judgment of character. This distinction might then explain why some research on social identity threat found ingroup-critical and other ingroup-deidentifying reactions. Studies that found reactions reducing in-group identification presented social identity threat in forms that reflect judgments of character (e.g. violations of fundamental values and rights, Glasford et al., 2008), where studies that found reactions critical of the ingroup’s behavior towards outgroups presented social identity threat in forms that reflect judgments of behavior or events (e.g. the Abu Ghraib torture scandal, Leidner et al., 2010).

De-identification occurs through in-group attachment and is available through (low) in-group glorification
As alluded to above, this research also expands upon work on de-identification. Social identity theory posits that leaving the group is one option that people have when the group stops serving a positive function for their self-concept. Consistent with this position, research shows that the in-group’s perceived violations of one’s personal values (Glasford et al., 2008), engagement in radical collective action against the in-group (Becker, Tausch, Spears, & Christ, 2011), or suffering unfair treatment by the in-group (Jasinskaja-Lahti, Liebkind, & Solheim, 2009) may all lead people to reduce their identification with the group. Our work demonstrates that such de-identification occurs through a specific quality of in-group identification: the perceived importance of and commitment to the in-group (i.e., in-group attachment).

Further, de-identification appears to be contingent on people’s pre-existing levels of another quality of in-group identification: perceived in-group superiority and deference to its norms and authorities (i.e., in-group glorification). While in-group attachment is severely skewed towards the high end of the spectrum, there is greater variability in in-group glorification. It thus makes sense that people who glorify their group cannot give up on it but have to defend it at all costs – whereas people who do not glorify their group might walk away from it when group membership becomes too costly.

While past research has treated a person’s configuration in terms of in-group attachment and glorification (or other aspects of other conceptualizations of in-group identification, for that matter) as fixed, of course a human development perspective would suggest that a person’s attachment can inform changes in glorification and vice versa. In parsing out how one aspect of in-group identification (here: glorification) affects another (here: attachment), we also started to not only investigate the nuances of in-group identification, but to investigate their dynamic interplay over time. Knowledge of this dynamic can help to find ways to have people identify in
more constructive ways with their groups, attaching a lot of importance and commitment to it without believing it is superior or unconditionally deferring to its norms and authorities.

**Responses to out-group suffering: Action or apathy?**

Based on the existing body of research on in-group identification and intervention, be it in in-group wrongdoing or out-group crises, one would expect that the world’s crises should be approached actively and constructively. Strongly identified group members voice their dissent to norms they perceive as harming the ingroup (Packer, 2007). The strongly attached feel guilty (Roccas et al., 2006) and should thus be likely to repair in-group wrongs (Doosje et al., 1998). Low glorifiers reaffirm harm and fairness morals (Leidner & Castano, 2012) and should thus intervene in wrongdoing. Yet, curiously the real world fails to meet these expectations most of the time. Somehow those who we think should help instead seem lost in apathy. Our studies help explain this discrepancy between what existing research leads us to expect and what the real world teaches us really happens. If the social identity threat is too great, even those inclined to use their own group to help others will turn their backs, because doing so is the only way to maintain a healthy self.
References


