Our Flaws Are More Human Than Yours : Ingroup Bias in Humanizing Negative Characteristics

Peter Koval, Simon M. Laham, Nick Haslam, Brock Bastian and Jennifer A. Whelan

Additional services and information for *Personality and Social Psychology Bulletin* can be found at:

Email Alerts: http://psp.sagepub.com/cgi/alerts

Subscriptions: http://psp.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Feb 22, 2012

OnlineFirst Version of Record - Sep 22, 2011

What is This?

Our Flaws Are More Human Than Yours: Ingroup Bias in Humanizing Negative Characteristics

Peter Koval[1,2], Simon M. Laham[1], Nick Haslam[1], Brock Bastian[3], and Jennifer A. Whelan[1]

Abstract

Four studies investigated whether people tend to see ingroup flaws as part of *human nature* (HN) to a greater degree than outgroup flaws. In Study 1, people preferentially ascribed high HN flaws to their ingroup relative to two outgroups. Study 2 demonstrated that flaws were rated higher on HN when attributed to the ingroup than when attributed to an outgroup, and no such difference occurred for positive traits. Study 3 replicated this *humanizing ingroup flaws* (HIF) effect and showed that it was (a) independent of desirability and (b) specific to the HN sense of humanness. Study 4 replicated the results of Study 3 and demonstrated that the HIF effect is amplified under ingroup identity threat. Together, these findings show that people humanize ingroup flaws and preferentially ascribe high HN flaws to the ingroup. These ingroup humanizing biases may serve a group-protective function by mitigating ingroup flaws as "only human."

Keywords

People routinely reserve characteristics that they judge to be most typically human for themselves, thereby subtly denying some aspects of humanness to others (Haslam, Loughnan, Kashima, & Bain, 2008; Leyens, Demoulin, Vaes, Gaunt, & Paladino, 2007). Although studies have shown that these biases have important psychological (e.g., Loughnan, Haslam, & Kashima, 2009) and behavioral consequences (e.g., Vaes, Paladino, Castelli, Leyens, & Giovanazzi, 2003), little is known about the motivations underlying the tendency to claim greater humanness for oneself and one's ingroup. We argue that people may be motivated to humanize their ingroup's negative traits to excuse these characteristics as "only human."

Most research on the attribution and denial of humanness in everyday social perception has defined humanness as consisting of the characteristics that distinguish humans from other animals. *Infrahumanization* involves the subtle, yet troubling, tendency for people to ascribe more such uniquely human attributes to their ingroups than to outgroups (e.g., Leyens et al., 2000; Leyens et al., 2001; Leyens et al., 2003). Recent studies have begun to shed light on the motivations underlying the infrahumanization bias. On one hand, perceiving an outgroup as lacking uniquely human qualities may help people to justify atrocities committed by their ingroup against that outgroup (Castano & Giner-Sorolla, 2006). On the other hand, imbuing the ingroup with

greater *human uniqueness* (HU) may help people to defend against the existential threat of their own mortality by distancing them from their animal nature (Vaes, Heflick, & Goldenberg, 2010).

Complementing research on infrahumanization is another line of work by Haslam and colleagues, which has defined a different, noncomparative, dimension of humanness labeled *human nature* (HN; Haslam, Bain, Douge, Lee, & Bastian, 2005). HN consists of characteristics that are seen as typically, fundamentally, and essentially human (e.g., warmth, emotionality, depth, curiosity; Haslam et al., 2008), without necessarily being unique to humans. Laypeople reliably distinguish between HN and HU, ratings of personality traits on these dimensions are generally uncorrelated (Haslam, Bastian, & Bissett, 2004; Haslam et al., 2005), and each sense of humanness is associated with different social categories (Loughnan & Haslam, 2007). The role of HN in interpersonal comparisons is well established: Multiple studies

[1]University of Melbourne, Melbourne, Australia
[2]University of Leuven, Leuven, Belgium
[3]University of Queensland, Brisbane, Australia

Corresponding Author:
Peter Koval, Department of Psychology, University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium
Email: peter.koval@psy.kuleuven.be

have demonstrated that people tend to ascribe more HN traits to themselves than to others (Haslam & Bain, 2007; Haslam et al., 2005), and this *self-humanizing* bias appears to be stable across cultures (Loughnan et al., 2010).

Recent research suggests that a bias similar to self-humanizing may exist in intergroup perception, such that people attribute HN preferentially to their ingroup relative to outgroups (Bain, Park, Kwok, & Haslam, 2009). However, no research to date has investigated the possible motivations underlying this ingroup bias in HN judgments. The current studies aimed to address this gap in the literature by investigating whether humanizing the ingroup on HN may serve a group-protective function. Self-humanizing has repeatedly been found to be stronger for negative traits, suggesting that it might be motivated by a desire to humanize personal weaknesses (Haslam & Bain, 2007; Haslam et al., 2005). Building on these previous findings, the current studies explored whether HN may serve a similar function in the intergroup context. We argue that people may mitigate or excuse their ingroup's flaws by judging them to be higher in HN than the negative traits of outgroups.

People derive important aspects of their identities from memberships in social groups (e.g., Brewer & Gardner, 1996; Sedikides & Brewer, 2001; Tajfel & Turner, 1986). Thus, acknowledging an ingroup's flaws is likely to pose some identity threat to its members (Branscombe, Ellemers, Spears, & Doosje, 1999; Ellemers, Spears, & Doosje, 2002), who are consequently likely to respond defensively (Hornsey, 2005). We argue that one such defensive maneuver is the motivated humanizing of ingroup flaws. Although there have been no studies that directly address this claim, recent research on the effects of *superordinate humanity salience* is relevant. This work shows that emphasizing the shared humanity of victims and perpetrators in the context of historical atrocities may allow people to excuse their ingroup's wrongdoing by seeing such behavior as common to all humans (Greenaway & Louis, 2010; Morton & Postmes, 2011). These studies provide insight into how people can justify acts of extreme violence or harm committed by their ingroups in the past but leave open the question of whether people may use humanness to mitigate flaws presently associated with their ingroups. Furthermore, research on the functions of superordinate humanity salience has not distinguished between different senses of humanness. In the current studies we investigated whether people are motivated to humanize ingroup flaws on both the HN and HU senses of humanness.

In light of the important distinctions between HN and HU identified in previous research (see Haslam et al., 2008, for a review), these two senses of humanness may be expected to operate differently in people's judgments of ingroup flaws. Specifically, we argue that there are a number of reasons why the HN dimension should be better suited to mitigating flaws than HU. First, HN traits are essentialized (i.e., seen as deep, inherent, and immutable), whereas HU characteristics are not (Haslam et al., 2005). Judging an ingroup's flaws as high on HN thus implies they are inborn and uncontrollable, which may help exonerate the group of blame for possessing them. Second, HN traits are perceived as being more prevalent and universal than HU characteristics (Haslam et al., 2005). As a consequence, seeing an ingroup's flaws as part of HN implies that they are shared by all of humanity and cannot be considered deviant. Finally, although the definition of HU is constrained to those attributes not shared by other species, HN can, in principle, include any characteristic that is believed to be essentially human. HN is therefore a relatively ambiguous and unconstrained sense of humanness, making it easier to judge an ingroup's flaws as high on HN regardless of their specific content.

Overview of the Current Studies

Four studies examined whether people tend to see their ingroup's flaws as higher in HN than outgroup flaws. We argue that humanizing ingroup flaws allows people to excuse these negative characteristics and thus serves a group-protective function. If such a bias exists, it should be observable in two distinct ways. First, flaws considered to be high in HN should be preferentially ascribed to the ingroup relative to outgroups. We tested this prediction in Study 1. Second, the degree to which flaws are judged to be aspects of HN should increase when they are attributed to an ingroup, relative to when the same flaws are ascribed to an outgroup. Study 2 tested this prediction and also examined people's judgments of positive traits. We reasoned that if humanizing ingroup flaws (relative to outgroup flaws) serves to mitigate these negative characteristics, no such bias should appear for the ingroup's positive traits because they do not require excusing. Study 3 sought to replicate the findings of Study 2 using different personality traits and examined whether the tendency to increase humanness ratings of ingroup flaws extends to the HU dimension as well as to judgments of desirability. Study 4 provided a more direct test of the prediction that humanizing ingroup flaws serves a group-protective function by examining whether the bias is amplified when ingroup identity is threatened.

Study 1

Study 1 examined whether people preferentially ascribe HN flaws to their ingroup relative to outgroups. Participants were asked to rate a number of flaws on HN and then to indicate how well each flaw described their ingroup and two outgroups. We predicted a positive association between ratings of flaws on HN and the degree to which flaws were endorsed as more descriptive of the ingroup than of outgroups. Participants were also asked to rate flaws on HU. Consistent with previous research on infrahumanization (e.g., Leyens

et al., 2001), we expected that people would preferentially attribute flaws rated high on HU to their ingroup. Finally, following previous studies (e.g., Haslam et al., 2005), we examined whether the predicted ingroup-humanizing bias was independent of flaw desirability. Although the exclusively negative traits used in this study were all relatively undesirable, it was nevertheless important to control for desirability to dissociate any bias in the ascription of HN traits from a basic evaluative tendency to ascribe less undesirable flaws to the ingroup.

Method

Participants. In return for course credit, 77 undergraduates (75% female), aged from 17 to 31 years ($M = 19.47$, $SD = 2.08$), participated. Most identified as Australian (64.9%), with the remainder identifying as Singaporean (11.7%), Malaysian (6.5%), Sri Lankan (3.9%), and other (13.0%).

Materials and procedure. Participants arrived at the laboratory in small groups of approximately 5 to 10, provided informed consent, and then individually completed a questionnaire composing two sections.

Section 1. Participants were presented with a list of 60 negatively valenced traits (flaws) drawn from McCrae and Costa (1985). These flaws, consisting of 12 traits sampled from the negative pole of each Big Five factor, were presented in a fixed random order. Participants rated the 60 flaws on desirability ("Is this characteristic desirable?"), HN ("Is this characteristic an aspect of 'human nature'?"), and HU ("Is this characteristic exclusively or uniquely human; i.e., does not apply to other species?"). Responses were made on a scale from 1 (*not at all*) to 7 (*very much so*). Participants were randomly assigned to complete these three items in one of three rating orders (counterbalanced according to a Latin square design).

Section 2. The second section of the questionnaire asked participants to specify their national identity. Next, participants rated the 60 flaws on ingroup descriptiveness ("How much do each of these flaws describe people in your country?") and two separate items measuring outgroup descriptiveness for Nigerians and Estonians ("How much do each of these flaws describe Nigerians/Estonians?"). Responses were made on a scale from 1 (*not at all*) to 7 (*very much so*). Thus, participants rated the 60 flaws on three items measuring how typical flaws were of their national ingroup and two outgroups. Participants were randomly assigned to complete the ingroup- and outgroup-descriptiveness items in one of three rating orders (counterbalanced according to a Latin square design). We selected Nigerians and Estonians as outgroups because the two nations differ in geographic location and ethnic/cultural composition. Consistency of effects across these two dissimilar groups would support the generalizability of our findings. After completing the questionnaires, participants were thanked and debriefed.

Results and Discussion

Each participant rated 60 flaws on six dimensions, resulting in a two-level nested data structure with flaws (at Level 1) nested within participants (at Level 2). To account for this nested data structure we conducted multilevel analyses, allowing us to simultaneously model variance at both levels of the data and obtain unbiased hypothesis tests (Nezlek, 2001). For a similar data analytic approach, see Vaes and Paladino (2010).

To test our main hypothesis, that people preferentially ascribe HN flaws to their ingroup relative to outgroups, we first calculated an intergroup-comparison index. Given that participants' ratings of the flaws for the two outgroups were highly correlated, we calculated an average outgroup rating for each flaw (separately for each participant) and then subtracted this average outgroup-descriptiveness rating from the ingroup-descriptiveness rating for each flaw to obtain an average intergroup-comparison score.[1] Similar comparative indices are commonly employed in intergroup research (e.g., Mummendey, Otten, Berger, & Kessler, 2000; Reynolds, Turner, & Haslam, 2000).

A two-level random coefficient model tested HN, HU, and desirability ratings as predictors of the intergroup-comparison index at Level 1. To allow for an interpretation of within-person effects, all Level 1 predictors were group-mean centered, thus removing between-person differences from the Level 1 parameter estimates (Enders & Tofighi, 2007; Nezlek, 2001). At Level 2 of the model, the intercept and slope values were allowed to vary randomly across participants.

As predicted, HN was a significant predictor of the intergroup-comparison score ($B = 0.11$, $SE = 0.02$, $p < .001$), indicating that participants preferentially ascribed HN flaws to their ingroup relative to the average outgroup. Moreover, this effect was obtained while controlling for HU and desirability ratings, indicating that it is independent of infrahumanization and evaluative ingroup favoritism. HU was also a significant predictor of the intergroup-comparison index ($B = 0.06$, $SE = 0.01$, $p < .001$). Desirability ratings did not significantly predict intergroup-comparison scores ($B = -0.03$, $SE = 0.03$, $p = .39$). This finding is consistent with previous studies showing that evaluative ingroup-favoritism effects tend to be attenuated on negative dimensions (e.g., Mummendey & Otten, 1998).

In sum, the findings of Study 1 support the hypothesized ingroup bias in the ascription of HN flaws. Moreover, this tendency to preferentially ascribe HN flaws to the ingroup relative to outgroups was independent of HU and desirability.

Study 2

Study 1 demonstrated that high HN flaws tend to be seen as more characteristic of the ingroup than of outgroups. However,

if judging ingroup flaws as part of HN serves to excuse them, people should also be more likely to judge flaws as high on HN when those flaws are associated with their ingroup than when associated with an outgroup. In Study 2, we investigated this "humanizing of ingroup flaws" (HIF) effect by providing mixed feedback (comprising positive and negative traits) to participants about their ingroup and an outgroup. Participants subsequently rated the traits attributed to their ingroup and outgroup on HN. We predicted that participants would rate flaws higher on HN when they were attributed to their ingroup than when those same flaws were attributed to an outgroup.

If HIF serves to excuse ingroup flaws, no corresponding ingroup bias should be observed for positive traits. This would be generally consistent with the notion that "people are more strongly motivated to avoid bad views of self than to claim good ones" (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001, p. 351). Thus, the HIF effect was conceptualized as an asymmetric ingroup bias in HN judgments whereby participants were expected to rate negative ingroup traits, but not positive ingroup traits, higher on HN (relative to the outgroup).

Method

Participants. In return for course credit, 51 undergraduates (69% female), aged from 17 to 50 years ($M = 20.67$, $SD = 6.53$), participated. Participants were Australians who had lived in Australia for at least 15 years. Most (88%) were Australian born.

Materials and procedure. Participants arrived at the laboratory in small groups of approximately 5 to 10, provided informed consent, and completed a questionnaire. Participants first read an introductory paragraph about cross-cultural psychology. They then read the abstract of "a recent scientific article presenting research on cross-cultural differences in personality." This mock abstract, created for the experiment, was titled "Cross-Cultural Differences in Personality Traits and Social Behavior Across 11 Nations" and summarized a research program that ostensibly used "a highly reliable and valid measure of personality" and "expert behavioral observations" to investigate cross-cultural differences in personality. The abstract stated that researchers had found personality to vary significantly across cultures and that "the strongest cultural differences in personality" had emerged between Australians (ingroup) and South Africans (outgroup). The abstract concluded by attributing two traits (1 positive, 1 negative) to each group. Thus, trait valence (positive vs. negative) and target (ingroup vs. outgroup) were manipulated within participants. As a counterbalancing measure, two versions of the article were constructed in which the specific trait pairs attributed to participants' ingroup and outgroup varied. Participants were randomly assigned to read an abstract claiming either that Australians were "calm" (positive trait) and "narrow-minded" (negative trait) whereas South Africans were "helpful" (positive trait)

and "stubborn" (negative trait; Article Version 1; $n = 25$), or vice versa (Article Version; $n = 26$).

Ratings of the traits from a previous study (Loughnan et al., 2010) indicated that the two trait pairs were balanced on HN (Ms = 4.20 & 4.18, SDs = 1.46 & 1.37) and desirability (Ms = 3.75 & 3.85, SDs = 1.23 & 1.28), ts < 0.26, ps > .80.

After reading the mock article, participants completed an open-response item and a multiple-choice question measuring their comprehension of the mock article.

Following the comprehension items, participants rated the four traits attributed to their ingroup and outgroup (embedded among six distracter traits) on HN using the same item and response scale as the one used in Study 1. Participants then rated how positively or negatively Australians had been portrayed in the mock article on a scale from 1 (*very negatively*) to 7 (*very positively*) and rated their agreement with each trait ascribed to the ingroup (e.g., "Do you agree that Australians tend to be narrow-minded?"). Finally, participants completed three items measuring agreement with the overall conclusions of the article (e.g., "Do you believe that the conclusions made in the article are plausible?") using a scale from 1 (*not at all*) to 7 *(very much so)*. After a thorough debriefing, no participants expressed suspicion of the study's hypotheses or of the mock article's authenticity.

Results and Discussion

Responses to the comprehension items indicated that all participants had read and correctly understood the mock article. The seven items measuring participants' agreement with the mock article were combined into a single measure ($\alpha = .63$). Participants assigned to read the two versions of the article did not differ in their ratings of agreement with the article ($t = 0.90$, $p > .36$), and agreement ratings did not moderate any of the effects reported below.

Participants rated the personality feedback attributed to their ingroup as neutral in valence ($M = 4.20$, $SD = 0.83$). However, participants who were assigned to read Article Version 2 ($M = 4.50$, $SD = 0.71$) rated the description of their ingroup as significantly more positive than participants who read Article Version 1 ($M = 3.88$, $SD = 0.83$), $t(49) = 2.87$, $p = .01$. This difference had no implications for the hypothesized HIF effect because valence ratings did not moderate any of the effects in the analysis of HN ratings (reported below).

Participants' HN ratings of the four traits attributed to their ingroup and outgroup (calm, narrow-minded, helpful, and stubborn) were the dependent measures. These HN ratings were analyzed using a 2 (target: ingroup vs. outgroup) × 2 (trait valence: positive vs. negative) × 2 (article version: Version 1 vs. Version 2) mixed ANOVA. Target and trait valence were within-participants factors, and article version was a between-participants factor. The hypothesized HIF effect was represented by the two-way interaction between target and trait valence, which was statistically significant,

$F(1, 49) = 4.96$, $p = .03$, $\eta_p^2 = .09$. Simple effects analyses revealed that, as predicted, participants rated flaws attributed to their ingroup ($M = 4.14$, $SD = 1.39$) significantly higher on HN than flaws attributed to the outgroup ($M = 3.61$, $SD = 1.55$), $F(1, 50) = 6.58$, $p = .01$, $\eta_p^2 = .12$. However, there was no significant difference in participants' HN ratings of positive traits attributed to their ingroup ($M = 4.51$, $SD = 1.45$) compared to the outgroup ($M = 4.61$, $SD = 1.20$), $F(1, 50) = 0.25$, $p = .62$, $\eta_p^2 = .01$.

In addition to the predicted target by trait valence interaction, the main effect of trait valence was also statistically significant, indicating that the two positive traits ($M = 4.56$, $SD = 1.13$) were rated higher on HN than the two flaws ($M = 3.87$, $SD = 1.27$), $F(1, 49) = 13.57$, $p < .001$, $\eta_p^2 = .22$, despite being balanced on the preratings. It should be noted that participants in the prerating sample rated the traits individually and in a context-free manner, whereas in the current study traits were attributed, in pairs, to participants' ingroup and outgroup. These differences may have caused participants to process trait-relevant information differently in the two studies.

The two-way interaction between target and article version was also significant, $F(1, 49) = 14.37$, $p < .001$, $\eta_p^2 = .23$. Analysis of simple effects revealed that participants who were assigned to read Article Version 2 rated the traits attributed to Australians significantly higher on HN ($M = 4.54$, $SD = 0.97$) than the traits ascribed to South Africans ($M = 3.85$, $SD = 1.02$), $F(1, 25) = 13.24$, $p = .001$, $\eta_p^2 = .35$. However, for participants who read Article Version 1, HN ratings of the traits attributed to Australians ($M = 4.10$, $SD = 1.18$) and South Africans ($M = 4.38$, $SD = 1.26$) trended in the opposite direction, although this difference did not reach statistical significance, $F(1, 25) = 2.68$, $p = .12$, $\eta_p^2 = .10$. Although unexpected, this result is not of theoretical interest to the current research as it represents an effect of the specific traits themselves on participants' HN ratings. No other effects in the model were statistically significant.

In sum, the findings of Study 2 supported the hypothesized HIF effect. When flaws were attributed to Australians (participants' national ingroup) they were rated higher on HN than when the same flaws were attributed to South Africans (an outgroup). No such difference emerged for positive traits.

Study 3

The results of Study 2 support the hypothesized HIF effect; however, it is unclear from these results whether HIF is exclusive to HN or if the HU dimension may show similar effects. Paladino and Vaes (2009) demonstrated that characteristics are judged to be more "uniquely human" when attributed to an ingroup compared to when attributed to an outgroup. This effect was obtained for both positive and negative characteristics, suggesting that it may be distinct from the HIF effect observed in Study 2. Paladino and Vaes (2009) did not examine the HN dimension, however, so it

remains to be seen how each dimension of humanness functions when participants have the opportunity to make judgments on both. Thus, Study 3 examined whether the HIF effect occurs exclusively on the HN dimension or whether ingroup flaws are also humanized on HU.

In Study 3, we also explored whether the tendency to humanize ingroup flaws is qualified by level of ingroup identification. The degree to which people identify as ingroup members has been shown to moderate a variety of intra- and intergroup processes (Ellemers et al., 2002), including how people respond to ingroup-threatening information such as negative feedback about the ingroup (Branscombe et al., 1999). Thus, it is possible that people who identify more strongly with their ingroup may display a greater HIF effect compared to low identifiers.

Finally, we examined whether HIF may simply be a way of reinterpreting ingroup flaws in a more positive light. If this were the case, humanizing might be a subtle form of evaluative ingroup favoritism. We tested this alternative explanation in Study 3 by measuring participants' desirability ratings of the traits attributed to their ingroup and outgroup.

Method

Participants. In return for course credit, 79 undergraduates (75% female), aged from 17 to 33 years ($M = 20.13$, $SD = 2.75$), participated. All participants were Australian and had lived in Australia for at least 15 years, and most (87%) were Australian born.

Materials and procedure. This experiment closely replicated Study 2 but differed in three respects. First, in addition to HN ratings, participants also rated the traits attributed to their ingroup and outgroup on desirability and HU using the same items and response scales as those used in Study 1. Participants were randomly assigned to complete these items in one of three rating orders, counterbalanced according to a Latin square design (Order 1 = HN, HU, desirability; Order 2 = desirability, HN, HU; Order 3 = HU, desirability, HN).

Second, after completing these trait ratings, participants completed Leach et al.'s (2008) 14-item ingroup identification scale, which measures *solidarity* (e.g., "I feel solidarity with Australians"), *satisfaction* (e.g., "I am glad to be Australian"), *centrality* (e.g., "The fact that I am Australian is an important part of my identity"), *individual self-stereotyping* (e.g., "I am similar to the average Australian"), and *ingroup homogeneity* (e.g., "Australian people are very similar to each other"). Responses to these items were made on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*).

Third, two new trait pairs were used in the mock article abstract that described Australians as "honest" and "impatient" and South Africans as "generous" and "nervous" (Article Version 1; $n = 41$), or vice versa (Article Version 2; $n = 38$). Pilot ratings of the traits indicated that the two trait pairs were balanced on HN (Ms = 4.27 & 4.22, SDs = 1.05 & 1.23),

desirability (Ms = 4.24 & 4.27, SDs = 0.77 & 0.83), and HU (Ms = 4.00 & 3.76, SDs = 1.26 & 1.58), ts < 0.71, ps > .48.

The trait pairs were used to describe Australians or South Africans as in Study 2, using the same mock article abstract.

Results and Discussion

Responses to comprehension items indicated that all participants had read and understood the mock article. Ratings of agreement with the conclusions of the mock article on a composite measure (α = .81) did not differ between the two versions of the article (p = .88). Importantly, agreement ratings did not significantly moderate any of the effects reported below.

Participants rated the personality feedback attributed to their ingroup as neutral in valence (M = 4.64, SD = 0.87), and valence ratings did not differ as a function of article version (p = .56).

HN ratings of the four experimental traits were subjected to a 2 (target: ingroup vs. outgroup) × 2 (trait valence: positive vs. negative) × 2 (article version: Version 1 vs. Version 2) × 3 (rating order: Order 1 vs. Order 2 vs. Order 3) mixed ANOVA with article version and rating order as between-participants factors, and target and trait valence as within-participants factors.

As predicted, the two-way interaction between target and trait valence was statistically significant, $F(1, 73) = 5.16$, $p = .026$, $\eta^2_p = .07$, replicating the HIF effect found in Study 2. Simple effects analyses indicated that participants rated flaws attributed to their ingroup (M = 5.04, SD = 1.20) significantly higher on HN than flaws attributed to the outgroup (M = 4.66, SD = 1.41), $F(1, 78) = 4.98$, $p = .03$, $\eta^2_p = .06$, whereas there was no significant difference for positive traits attributed to the ingroup (M = 4.57, SD = 1.41) versus the outgroup (M = 4.65, SD = 1.44), $F(1, 78) = 0.28$, $p = .60$, $\eta^2_p = .004$ (see Figure 1). Thus, the hypothesized HIF effect was found using a different set of personality traits to those used in Study 2.

There was also a main effect of rating order, $F(2, 73) = 3.63$, $p = .03$, $\eta^2_p = .09$. Post hoc analyses with Tukey's honestly significant difference test revealed that HN ratings were higher when completed last (Order 3) compared to when completed first (Order 1), $p = .05$. However, no other pairwise comparisons were significant. Crucially, rating order did not moderate the predicted target by trait valence interaction (representing the HIF effect) nor any other effects in the model. There were no other significant effects in the model.

The 14 items measuring ingroup identification were averaged to form a composite score (α = .89; M = 4.54, SD = 0.97). To examine whether ingroup identification moderated the HIF effect, ingroup identification scores were centered and included as a continuous covariate in the mixed ANOVA predicting HN ratings. Product terms were built using ingroup identification and each factor in the model to test for interactions. The main effect of the covariate was not significant,
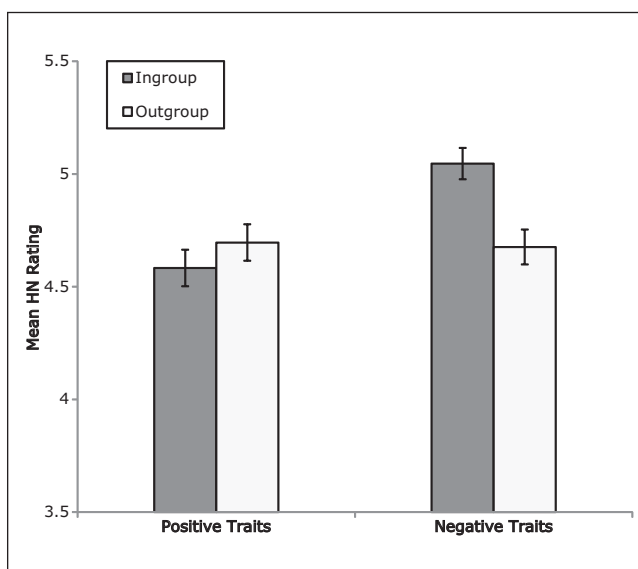
Figure 1. Mean human nature (HN) ratings of positive and negative traits attributed to the ingroup and outgroup in Study 3
Error bars represent one standard error of the mean.

$F(1, 67) = 0.41$, $p = .53$, $\eta^2_p = .01$. More importantly, ingroup identification did not moderate the predicted target by trait valence interaction, $F(1, 67) = 0.12$, $p = .74$, $\eta^2_p = .002$, or any other effects in the model. Furthermore, the predicted two-way interaction between target and trait valence remained statistically significant when controlling for ingroup identification, $F(1, 67) = 6.65$, $p = .01$, $\eta^2_p = .09$. Ingroup identification was also examined as a continuous moderator in the analyses of HU and desirability (below) and was found not to significantly moderate any of the effects in these models.[2]

HU ratings were also analyzed with a 2 (target) × 2 (trait valence) × 2 (article version) × 3 (rating order) mixed ANOVA. A significant main effect of target was found, $F(1, 73) = 6.85$, $p = .01$, $\eta^2_p = .09$, indicating that ingroup traits received higher HU ratings than outgroup traits, replicating the findings of Paladino and Vaes (2009). The main effect of trait valence was also significant, $F(1, 73) = 67.44$, $p < .001$, $\eta^2_p = .48$, indicating that positive traits received higher HU ratings than negative traits. However, the crucial two-way interaction between target and trait valence (representing the HIF effect) was not significant, $F(1, 73) = 0.90$, $p = .35$, $\eta^2_p = .01$. The main effects of target and trait valence were qualified by higher order interactions with article version and rating order, although these were of little theoretical significance and are therefore not discussed further.

Desirability ratings of the two positive and two negative traits were analyzed separately, with 2 (target) × 2 (article version) × 3 (rating order) mixed ANOVAs. For positive traits, no significant effects emerged (all Fs < 2.4, all ps > .12), indicating that participants' desirability ratings of the positive traits (honest: M = 6.49, SD = 0.83; generous: M = 6.35, SD = 0.75) did not differ from each other in any of

the conditions. Analysis of the two negative traits in the same model revealed only a significant target × article version interaction, $F(1, 73) = 15.97$, $p < .001$, $\eta_p^2 = .18$. This interaction reflects an effect of trait content and is not of any theoretical interest to the current research. Importantly, neither positive nor negative traits were judged to be more desirable when attributed to the ingroup compared to the outgroup, suggesting that the HIF effect is not reducible to an increase in the perceived desirability of flaws when they are attributed to the ingroup.

Study 4

Studies 2 and 3 demonstrated that people humanize flaws on the HN dimension when they are attributed to their ingroup, relative to when they are ascribed to an outgroup. We argue that this HIF effect may be driven by a motivation to mitigate ingroup flaws by casting them as only human. Although the findings of Studies 2 and 3 are consistent with this explanation by showing that humanizing occurs only for the undesirable traits that threaten to tarnish ingroup identity (and not for positive traits), they do not directly demonstrate that a group-protective motivation underlies the HIF effect. Study 4 was designed to test this hypothesis more directly by examining whether the activation of ingroup identity threat amplifies the HIF effect.

Previous research has shown that self-protective biases are stronger under conditions of self-threat (Campbell & Sedikides, 1999; Green, Sedikides, & Gregg, 2008). Similarly, people become more defensive of the ingroup when their ingroup identities are threatened (Cadinu & Cerchioni, 2001; Ellemers et al., 2002). Following this reasoning, we argue that people's motivation to humanize their ingroup's flaws should increase when their ingroup identity is threatened. In other words, if HIF serves a group-protective function, then increasing the need to protect the ingroup (by threatening the value of ingroup identity) should amplify the need to humanize flaws attributed to the ingroup. To test this prediction, we replicated Study 3 with the addition of a between-participants manipulation of ingroup identity threat. Although ingroup identity threat may take various forms (Branscombe et al., 1999), we focused on threatening the value of the ingroup. This was manipulated by giving participants feedback that their ingroup compares unfavorably with an outgroup. We argue that judging ingroup flaws as higher in HN than outgroup flaws serves to protect the ingroup by excusing the ingroup's negative traits as more innate, essential, and common to all humans. As a consequence, we predicted that ingroup identity threat would magnify the HIF effect. In contrast, we argue that perceiving ingroup flaws as more uniquely human than outgroup flaws would not provide the same kind of mitigation as humanizing them on HN. Ingroup identity threat was therefore not expected to moderate judgments of ingroup and outgroup traits on HU.

Method

Participants. In return for course credit, 54 undergraduates (83% female), ranging from 17 to 28 years of age ($M = 19.07$, $SD = 2.31$), participated. Participants were Australian citizens who had lived in Australia for at least 15 years, and most (89%) were Australian born.

Materials and procedure. This experiment closely replicated Study 3, with the important addition of an ingroup threat manipulation. Participants were randomly assigned to either the ingroup threat condition ($n = 28$) or the control condition ($n = 26$). Participants read a modified version of the mock scientific articles used in Study 3, in which a pair of traits was attributed to participants' ingroup (Australians) and outgroup (South Africans). In the ingroup threat condition the mock article also reported that compared to outgroup traits, the ingroup's traits were "associated with more frequent interpersonal conflicts, reduced feelings of social-connectedness and a smaller social network" and concluded by stating that, relative to the interpersonal style of South Africans, "the interpersonal style prevalent among Australians predicts lower success in social relationships and may therefore be characterized as 'socially maladaptive.'" In contrast, the mock article in the control condition informed participants that ingroup and outgroup traits "were not associated with differing frequency of interpersonal conflicts, feelings of social-connectedness or social network size" and concluded by stating that "these findings suggest that the interpersonal styles prevalent among Australians and South Africans predict equal success in social relationships and may therefore be characterized as similarly 'socially adaptive.'" Two additional manipulation check items designed to measure ingroup identity threat were included: "How do you feel about the way Australians were described in the article?" rated on a scale from 1 (*it does not bother me at all*) to 7 (*I feel that an important aspect of my identity has been threatened*) and "Did it worry you to learn that the Australian interpersonal style leads to [reduced/equal] success in social relationships compared to the South African interpersonal style and is therefore [similarly] 'socially [mal]adaptive'?" rated on a scale from 1 (*did not worry me at all*) to 7 (*worried me a great deal*). These two items were combined into a composite identity-threat measure (α = .66).

Results and Discussion

Responses to the comprehension items indicated that all participants had read and understood the mock article. Ratings of agreement with the conclusions of the mock article on a composite measure (α = .65) did not differ between the two versions of the article ($p = .99$).[3] However, participants in the ingroup threat condition reported significantly lower agreement with the mock article ($M = 3.42$, $SD = 1.16$) compared to control participants ($M = 4.08$, $SD = 1.16$), $t(52) = 2.09$,
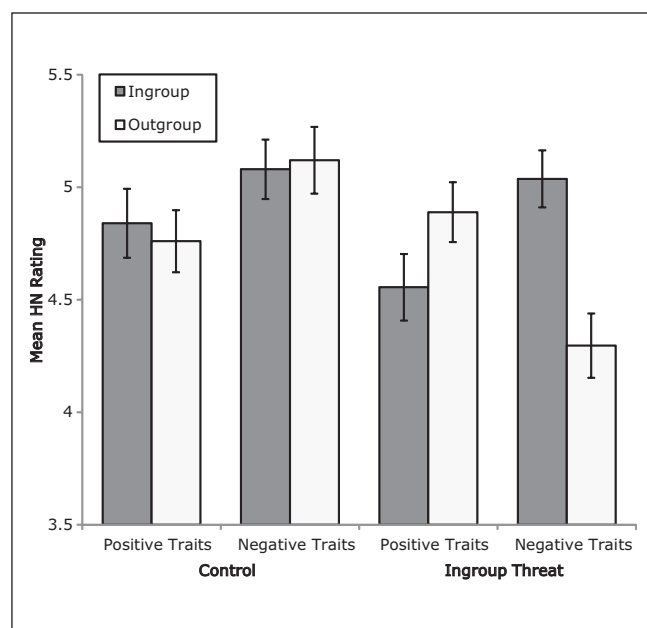
Figure 2. Mean human nature (HN) ratings of positive and negative traits attributed to the ingroup and outgroup in the control and ingroup threat conditions in Study 4
Error bars represent one standard error of the mean.

$p = .04$. Importantly, agreement ratings did not significantly moderate any of the effects reported below.

Scores on the composite identity-threat measure served as a manipulation check. Participants in the ingroup threat condition experienced significantly higher identity threat after reading the mock article ($M = 2.95$, $SD = 1.36$) compared to control participants ($M = 1.73$, $SD = 0.71$), $t(52) = 4.08$, $p < .001$, indicating that the manipulation succeeded in threatening the ingroup identity of participants in the ingroup threat condition, relative to participants in the control condition.

To test our main hypothesis that the HIF effect would be moderated by ingroup identity threat, we subjected participants' HN ratings of the four experimental traits to a 2 (target: ingroup vs. outgroup) × 2 (trait valence: positive vs. negative) × 2 (threat condition: ingroup threat vs. control) mixed ANOVA.[4] Target and trait valence were within-participants factors, and threat condition was a between-participants factor. The two-way interaction between target and trait valence, representing the HIF effect, was marginally significant, $F(1, 50) = 3.87$, $p = .06$, $\eta^2_p = .07$, replicating the findings of Studies 2 and 3. However, as predicted, this effect was qualified by a significant three-way interaction with threat condition, $F(1, 50) = 6.07$, $p = .02$, $\eta^2_p = .11$, indicating that ingroup identity threat significantly moderated the HIF effect (see Figure 2). No other effects in the model were significant (all $ps > .35$).

To decompose this three-way interaction, we conducted two 2 (target) × 2 (trait valence) repeated measures ANOVAs

separately for participants in the ingroup threat and control conditions. These analyses revealed that the interaction between target and trait valence interaction (representing the HIF effect) was significant only for participants in the ingroup threat condition, $F(1, 26) = 8.63$, $p = .01$, $\eta^2_p = .25$, and not for participants in the control condition, $F(1, 24) = 0.15$, $p = .70$, $\eta^2_p = .01$. No other effects were significant in either model. Simple effects analyses indicated that, as predicted, participants in the ingroup threat condition rated flaws significantly higher on HN when attributed to their ingroup ($M = 5.04$, $SD = 1.16$) versus the outgroup ($M = 4.30$, $SD = 1.51$), $F(1, 26) = 5.73$, $p = .02$, $\eta^2_p = .18$, whereas positive traits were rated as equally part of HN when associated with the ingroup ($M = 4.56$, $SD = 1.67$) and outgroup ($M = 4.89$, $SD = 1.40$), $F(1, 26) = 1.30$, $p = .265$, $\eta^2_p = .05$ (see Figure 2). These findings indicate that threatening ingroup identity triggered a motivation to humanize ingroup flaws relative to outgroup flaws among participants in the ingroup threat condition, whereas the same motivation was not triggered among participants in the control condition.

Although the failure to find a significant target by trait valence interaction in the control condition was unexpected, it may be the result of the fact that the mock article used in the control condition inadvertently served as an ingroup affirmation. Specifically, participants in the control condition were informed that their ingroup's traits lead to successful social functioning and are "socially adaptive" (see materials and procedure). Although this text was designed to mirror as closely as possible the wording of the article in the ingroup threat condition, it may have reassured participants that their ingroup's negative traits are not detrimental to their group's status, which appears to have counteracted their motivation to humanize their ingroup flaws.

To determine whether the HIF effect was exclusive to the HN dimension or was more generally related to ratings of humanness, we subjected participants' HU ratings of the experimental traits to the same 2 (target) × 2 (trait valence) × 2 (threat condition) mixed ANOVA. Only the main effect of trait valence was significant, $F(1, 52) = 16.53$, $p < .001$, $\eta^2_p = .25$, such that positive traits were rated significantly higher on HU ($M = 4.02$, $SD = 1.61$) than negative traits ($M = 2.86$, $SD = 1.61$). There were no other significant effects. These results indicate that, in contrast to HN ratings, participants did not humanize ingroup flaws on the HU dimension regardless of whether or not their ingroup identity was threatened. These findings support the notion that HN and HU may serve different functions in intergroup perception. Although judging ingroup flaws to be higher on HN than outgroup flaws appears to serve a group-protective function, the same is not true for HU judgments.

Finally, participants' desirability ratings of traits were analyzed using the same 2 (target) × 2 (trait valence) × 2 (threat condition) mixed ANOVA. Again, only the trivial main effect of trait valence emerged as significant, $F(1, 52) = 1029.38$, $p < .001$, $\eta^2_p = .95$, such that positive traits were rated as

significantly more desirable ($M = 6.54$, $SD = 0.61$) than negative traits ($M = 1.96$, $SD = 0.79$). No other effects in the model were significant. These findings indicate that participants did not judge traits attributed to their ingroup as more desirable than outgroup traits, regardless of whether their ingroup identity was threatened. Furthermore, consistent with the findings of Study 3, these results demonstrate that the HIF effect does not enhance the positivity of ingroup traits.

General Discussion

The current studies demonstrate that people tend to see their ingroup's flaws as higher on HN than the flaws of outgroups. In Study 1, participants preferentially ascribed high HN flaws to their ingroup relative to two outgroups. In contrast, participants did not attribute more desirable flaws to their ingroup than to outgroups. The independence of these effects adds weight to the argument that HN is an important dimension of social perception, independent of valence (e.g., Haslam et al., 2005).

Study 2 demonstrated that people humanize flaws that are attributed to their ingroup. However, this was not the case for positive traits, suggesting that our findings do not reflect a general tendency to rate all ingroup characteristics as high on HN (cf. Paladino & Vaes, 2009). Study 3 replicated this asymmetrical HIF effect using a different set of personality traits and found that the effect was not moderated by ingroup identification or based on a tendency to see ingroup flaws as more desirable than outgroup flaws. In contrast, HU judgments did not display the same asymmetrical pattern: Both positive and negative traits were rated higher on HU when ascribed to the ingroup. Finally, Study 4 demonstrated that ingroup identity threat amplifies the HIF effect. In contrast, threatening participants' ingroup identity had no effect on their HU or desirability judgments of ingroup or outgroup traits. Together, these findings support the hypothesis that people are motivated to see their ingroup's flaws as higher in HN to mitigate these flaws.

Does HIF Serve a Group-Protective Function?

Throughout this article, we have argued that judging ingroup flaws to be higher on HN than outgroup flaws serves a group-protective function by mitigating ingroup flaws. Although the current findings are consistent with this group-protective account, alternative explanations should be considered. For instance, the correlational nature of Study 1 makes it impossible to establish whether believing flaws to be aspects of HN caused people to preferentially ascribe such flaws to their ingroup. Instead, this ingroup bias may reflect a tendency to generalize the characteristics of a specific ingroup to a super-ordinate ingroup (humans), as proposed by the *ingroup projection model* (Wenzel, Mummendey, & Waldzus, 2007). Alternatively, this finding may reflect a more general overlap

between cognitive representations of specific ingroups and the more inclusive human group (cf. Otten, 2005; Smith, Coats, & Walling, 1999).

However, these alternative accounts would predict that both positive and negative characteristics should be considered more human when attributed to the ingroup and are therefore inconsistent with the asymmetrical HIF effect observed in Studies 2, 3, and 4. Rather, the fact that people repeatedly humanized negative, but not positive, ingroup traits on the HN dimension is suggestive of a motivational explanation. Although being reminded of an ingroup's virtues is not threatening, having negative traits attributed to one's ingroup may drive people to seek consolation by attributing ingroup flaws to HN, relative to the flaws of other groups.

Furthermore, Study 4 provides more direct evidence of the group-protective function of humanizing ingroup flaws by showing that the HIF effect is amplified by ingroup identity threat. Drawing on previous research in which defensive motivation has been shown to increase under threat (see Roese & Olson, 2007, for a review), we argue that the findings of Study 4 provide strong support for a group-protective explanation of the HIF effect. Although we predicted that ingroup identity threat would increase the motivation to humanize ingroup flaws, we note that the failure to replicate the HIF effect in the control condition of Study 4 appears to be at odds with the findings of Studies 2 and 3. There may be various reasons for these divergent results. As already mentioned, the mock article used in the control condition of Study 4 may have served to affirm participants' ingroup identities by casting the ingroup's traits as functional and adaptive, eliminating the need to humanize the ingroup's flaws. On the other hand, given the lack of further context, participants in Studies 2 and 3 may have experienced the attribution of flaws to their ingroup as identity threatening. If this were the case, it would imply that the HIF effect occurs only when the status of the ingroup is undermined, rather than when any negative trait is ascribed to the ingroup. More research is needed to further clarify under which conditions the HIF effect is magnified or attenuated.

In sum, although it may be too early to conclude that the HIF effect is driven exclusively by a group-protective motivation, especially given the variety of nonmotivational factors that are known to be involved in similar biases (Alicke & Sedikides, 2009), the current studies provide strong support for a group-protective account of humanizing ingroup flaws.

Ingroup Identification

Study 3 failed to find evidence that ingroup identification moderates the tendency to humanize ingroup flaws. Although, as with any null result, this should be interpreted with caution, we nevertheless offer two possible explanations for this finding. First, even low identifiers may appraise negative

evaluations of their ingroup as threatening (Branscombe et al., 1999) and may therefore, along with high identifiers, seek to mitigate their ingroup's flaws by judging them to be higher on HN compared to outgroup flaws. This is particularly likely to be the case in relation to socially significant, "natural" groups with relatively impermeable boundaries (Ellemers et al., 2002), such as the national groups that featured in the current research.

Second, although biases that *enhance* the positivity of the group appear to be more pronounced among high identifiers (e.g., Branscombe & Wann, 1994), group-*protective* strategies may be equally prevalent among high and low identifiers. For example, both high and low ingroup identifiers respond to negative feedback about their ingroup by undermining the reliability of the feedback or downplaying its importance (e.g., Cadinu & Cerchioni, 2001). Studies 3 and 4 demonstrated that the HIF effect is not a straightforward enhancement effect based on judging ingroup flaws to be more desirable than outgroup flaws. Rather, we argue that HIF may serve to protect ingroup identity by normalizing ingroup flaws as innate and universal to all humans and thus making them appear more acceptable. Rather than enhancing the positivity of the ingroup, HIF may protect the group by lowering the rest of humanity to the level of the flawed ingroup. As a consequence, both high and low ingroup identifiers may equally engage in this protective strategy.

Distinct Senses of Humanness With Distinct Functions

The current research provides further support for the distinctness of the two senses of humanness proposed by Haslam and colleagues (2005, 2008). In Study 1, the preferential ascription of high HN flaws to the ingroup (relative to outgroups) was independent of the ascription of high HU flaws. In Studies 3 and 4, the predicted asymmetrical HIF effect was found only for the HN dimension. In contrast, participants rated both negative and positive traits attributed to their ingroup higher on HU in Study 3, replicating Paladino and Vaes (2009). Furthermore, Study 4 demonstrated that ingroup identity threat significantly influenced HN ratings of ingroup flaws but had no impact on participants' judgments of ingroup traits on the HU dimension. Together, these findings indicate that people meaningfully distinguish between these two senses of humanness and suggest that each kind of humanness may serve a distinct function in intergroup perception.

We argue that the distinct effects observed for HN and HU judgments in the current studies stem from important differences between the two senses of humanness, which may make HN more amenable to excusing ingroup flaws. In particular, HN characteristics tend to be seen as more deep seated (i.e., essentialized) than HU characteristics (Haslam et al., 2005), so perceiving flaws as aspects of HN frames them as more fixed and uncontrollable, and thus arguably

more excusable. In addition, HN traits are seen as widely shared or universal (Haslam et al., 2004; Haslam et al., 2005), so perceiving flaws as part of HN may help mitigate them by making them seem common and normative. Finally, HN is the more ambiguous of the two senses of humanness because HU's content is more constrained by the explicit human–animal contrast. Consequently the HN concept is more open to idiosyncratic redefinition and reappraisal, and thus more able to accommodate the ingroup's flaws under its forgiving umbrella.

Although HU may not serve to mitigate an ingroup's flaws, it may nevertheless serve a different function. Vaes et al. (2010) suggest that the tendency to see ingroup characteristics as higher on HU serves a terror management function. Because mortality salience induces people to "deny their similarity to animals . . . [and] emphasize the uniquely human aspects of the self" (Vaes et al., 2010, p. 750) and to identify strongly with their ingroups, it follows that existential threats may also induce people to see their ingroups as possessing more HU characteristics. In support of this hypothesis, Vaes et al. (2010) found that participants who were exposed to a mortality-salience prime displayed an increased association between ratings of traits as ingroup descriptive and HU. In short, endowing the ingroup with desirable and undesirable HU qualities—symbols of human culture and socialization—ameliorates existential threats by helping to distance people from their animal nature, which entails death. Research has only just begun to explore the motivational functions of humanness. However, considered together, Vaes et al. and the present studies are beginning to clarify the specific functions of each sense of humanness. HN, more than HU, may serve to protect against the threat of acknowledging ingroup flaws by allowing people to excuse these flaws as only human. In contrast, HU may be better suited to protecting us against existential threats because it distinguishes and distances humans from animals (Haslam et al., 2008). Further exploration of the unique motivational functions of each sense of humanness would be a fruitful avenue for future research.

Implications of the Findings

The findings of the current studies have several implications for research on humanness and dehumanization. First, until recently most research and theory linked the HU dimension to the intergroup domain and the HN dimension to the interpersonal domain. The present studies are among the first to demonstrate that HN is also relevant in the intergroup context. Second, the current findings support the notion that lay conceptions of humanness are flexible and functional. Studies 2, 3, and 4 show that HN ratings of traits are malleable, depending on whether the traits are associated with the ingroup or outgroup. These findings complement Paladino and Vaes's (2009) finding that people's judgments of characteristics on the HU dimension also vary depending on

whether they are attributed to the ingroup or outgroup. Third, the present studies highlight the importance of investigating specific motives underlying attributions or denials of humanness. The few studies that have examined possible motivations underlying such effects have focused mainly on motives that might account for the denial of HU to outgroups (e.g., Castano & Giner-Sorolla, 2006; Zebel, Zimmerman, Viki, & Doosje, 2008). The current studies explored the previously unexamined question of why people may be motivated to humanize ingroup flaws on the HN dimension. Our findings suggest that a group-protective motivation may drive people to imbue their ingroup's flaws with greater HN.

Limitations and Future Directions

The current studies demonstrate that people tend to see their ingroup's flaws as higher in HN than outgroup flaws, either by preferentially ascribing negative traits rated high in HN to their ingroup or by increasing their HN ratings of flaws attributed to their ingroup. However, the HIF effect was operationalized as a relative effect in these studies (i.e., ingroup relative to outgroup). As such, it is unclear to what extent the effect represents ingroup humanizing, outgroup dehumanizing, or a combination of the two processes. Further studies are needed to tease apart these possibly distinct effects and consider the functions each may serve in intergroup perception.

Although the current studies provide initial support for the group-protective function of humanizing ingroup flaws, additional research exploring the consequences of humanizing ingroup flaws is needed. It is unclear whether the HIF bias confers psychological benefits such as reductions in negative affect. Furthermore, future research might also be able to shed more light on the processes underlying the tendency to humanize ingroup flaws by examining other possible moderators or mediators of the effect. A possible dispositional moderator of the effect may be sensitivity to ingroup identity threats. For instance, Sassenberg and Hansen (2007) found that prevention-focused individuals experienced stronger negative affective responses after their ingroup was evaluated unfavorably by an outgroup, and this might motivate stronger tendencies to humanize the ingroup's supposed failings. A possible mediator of the HIF effect may be the tendency to judge ingroup flaws as less deviant when they are associated with HN, in line with previous research showing that HN characteristics are perceived as highly prevalent in the population (Haslam et al., 2004).

Conclusion

The current studies investigated whether people tend to see ingroup flaws as more a part of HN than outgroup flaws. Our findings suggest that people may ameliorate the threat posed by ingroup flaw acknowledgment either by preferentially attributing HN flaws to their ingroup or by humanizing their ingroup's flaws. Believing an ingroup's flaws to be aspects of an essential HN may help to protect the group by making those flaws appear less blameworthy. Centuries of philosophical scholarship has shown that HN is an ambiguous but reassuring concept (Pojman, 2006), and it may be precisely these features of the concept that make HN particularly well suited to mitigating our flaws.

Notes

1. At the trait level (aggregating across all participants) outgroup-descriptiveness ratings of the flaws for the two outgroups correlated at $r = .86$. Furthermore, when the below-reported analysis was repeated using intergroup-comparison scores calculated separately for Nigerian and Estonian outgroup-descriptiveness ratings, the results did not differ from those obtained using average intergroup-comparison scores.
2. Separate analyses testing the five subscales of Leach et al.'s (2008) ingroup identification measure as moderators in each of the reported models yielded no significant interactions.
3. Initial analyses revealed that none of the predicted effects were significantly qualified by article version. We therefore excluded article version from all analyses reported in Study 4.
4. Analyses of HN ratings were based on $n = 52$ because of missing data for 2 participants.

References

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology*, *20*, 1-48.

Bain, P., Park, J., Kwok, C., & Haslam, N. (2009). Attributing human uniqueness and human nature to cultural groups: Distinct forms of subtle dehumanization. *Group Processes and Intergroup Relations*, *12*, 789-805.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K., D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323-370.

Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The context and content of social identity threat. In N. Ellemers, R. Spears, & B. Doosje (Eds.), *Social identity: Context*, *commitment*, *content* (pp. 35-58). Oxford, UK: Blackwell.

Branscombe, N. R., & Wann, D. L. (1994). Collective self-esteem consequences of outgroup derogation when a valued social identity is on trial. *European Journal of Social Psychology*, *24*, 641-657.

Brewer, M. B., & Gardner, W. (1996). Who is this "we"? Levels of identity and self representations. *Journal of Personality and Social Psychology*, *71*, 83-93.

Cadinu, M. R., & Cerchioni, M. (2001). Compensatory biases after ingroup threat: "Yeah, but we have a good personality." *European Journal of Social Psychology*, *31*, 353-367.

Campbell, W. K., & Sedikides, C. (1999). Self-threat magnifies the self-serving bias. *Review of General Psychology*, *3*, 23-43.

Castano, E., & Giner-Sorolla, R. (2006). Not quite human: Infra-humanization in response to collective responsibility for intergroup killing. *Journal of Personality and Social Psychology*, *90*, 804-818.

Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual Review of Psychology*, *53*, 161-186.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*, 121-138.

Green, J. D., Sedikides, C., & Gregg, A. P. (2008). Forgotten but not gone: The recall and recognition of self-threatening memories. *Journal of Experimental Social Psychology*, *44*, 547-561.

Greenaway, K. H., & Louis, W. R. (2010). Only human: Hostile human norms can reduce legitimisation of intergroup discrimination by perpetrators of historical atrocities. *British Journal of Social Psychology*, *49*, 765-783.

Haslam, N., & Bain, P. (2007). Humanizing the self: Moderators of the attribution of lesser humanness to others. *Personality and Social Psychology Bulletin*, *37*, 57-68.

Haslam, N., Bain, P., Douge, L., Lee, M., & Bastian, B. (2005). More human than you: Attributing humanness to self and others. *Journal of Personality and Social Psychology*, *89*, 937-950.

Haslam, N., Bastian, B., & Bissett, M. (2004). Essentialist beliefs about personality and their implications. *Personality and Social Psychology Bulletin*, *30*, 1661-1673.

Haslam, N., Loughnan, S., Kashima, Y., & Bain, P. (2008). Attributing and denying humanness to others. *European Review of Social Psychology*, *19*, 55-85.

Hornsey, M. J. (2005). Why being right is not enough: Predicting defensiveness in the face of group criticism. *European Review of Social Psychology*, *16*, 301-334.

Leach, C. W., van Zomeren, M., Zebel, S., Vliek, M. L., Pennekamp, S. F., Doosje, B., . . . Spears, R. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology*, *95*, 144-165.

Leyens, J. P., Cortes, B., Demoulin, S., Dovidio, J., Fiske, S., Gaunt, R., . . . Vaes, J. (2003). Emotional prejudice, essentialism and nationalism: The 2002 Tajfel lecture. *European Journal of Social Psychology*, *33*, 703-717.

Leyens, J. P., Demoulin, S., Vaes, J., Gaunt, R., & Paladino, M. P. (2007). Infra-humanization: The wall of group differences. *Social Issues and Policy Review*, *1*, 139-172.

Leyens, J. P., Paladino, M.-P., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The emotional side of prejudice: The role of secondary emotions. *Personality and Social Psychology Review*, *4*, 186-197.

Leyens, J. P., Rodriguez-Perez, A., Rodriguez-Torres, R., Gaunt, R., Paladino, M.-P., Vaes, J., & Demoulin, S. (2001). Psychological essentialism and the attribution of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology*, *31*, 395-411.

Loughnan, S., & Haslam, N. (2007). Animals and androids: Implicit associations between social categories and nonhumans. *Psychological Science*, *18*, 116-121.

Loughnan, S., Haslam, N., & Kashima, Y. (2009). Understanding the relationship between attribute-based and metaphor-based dehumanization. *Group Processes and Intergroup Relations*, *12*, 747-762.

Loughnan, S., Leidner, B., Haslam, N., Kashima, Y., Doron, G., Tong, J., & Yeung, V. (2010). Universal biases in self-perception: Better and more human than average. *British Journal of Social Psychology*, *49*, 627-636.

McCrae, R. R., & Costa, P. T. (1985). Updating Norman's "adequate taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology*, *49*, 710-721.

Morton, T. A., & Postmes, T. (2011). Moral duty or moral defense? The effects of perceiving shared humanity with the victims of ingroup perpetrated harm. *European Journal of Social Psychology*, *41*, 127-134.

Mummendey, A., & Otten, S. (1998). Positive-negative asymmetry in social discrimination. *European Review of Social Psychology*, *9*, 107-143.

Mummendey, A., Otten, S., Berger, U., & Kessler, T. (2000). Positive-negative asymmetry in social discrimination: Valence of evaluation and salience of categorization. *Personality and Social Psychology Bulletin*, *26*, 1258-1270.

Nezlek, J. B. (2001). Multilevel random coefficient analyses of event- and interval-contingent data in social and personality psychology research. *Personality and Social Psychology Bulletin*, *27*, 771-785.

Otten, S. (2005). The ingroup as part of the self: Reconsidering the link between social categorization, ingroup favoritism, and the self-concept. In M. D. Alicke, D. A. Dunning, & J. I. Krueger (Eds.), *The self in social judgment* (pp. 241-265). New York, NY: Psychology Press.

Paladino, M. P., & Vaes, J. (2009). Ours is human: On the pervasiveness of infra-humanization in intergroup relations. *British Journal of Social Psychology*, *48*, 237-251.

Pojman, L. P. (2006). *Who are we? Theories of human nature*. New York, NY: Oxford University Press.

Reynolds, K. J., Turner, J. C., & Haslam, S. A. (2000). When are we better than them and they worse than us? A closer look at social discrimination in positive and negative domains. *Journal of Personality and Social Psychology*, *78*, 64-80.

Roese, N. J., & Olson, J. M. (2007). Better, stronger, faster: Self-serving judgment, affect regulation, and the optimal vigilance hypothesis. *Perspectives on Psychological Science*, *2*, 124-141.

Sassenberg, K., & Hansen, N. (2007). The impact of regulatory focus on affective responses to social discrimination. *European Journal of Social Psychology*, *37*, 421-444.

Sedikides, C., & Brewer, M. B. (Eds.). (2001). *Individual self, relational self, collective self*. Philadelphia, PA: Psychology Press.

Smith, E. R., Coats, S., & Walling, D. (1999). Overlapping mental representations of self, in-group, and partner: Further response time evidence and a connectionist model. *Personality and Social Psychology Bulletin*, *25*, 873-882.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 33-48). Chicago, IL: Nelson Hall.

Vaes, J., Heflick, N. A., & Goldenberg, J. L. (2010). "We are people": In-group humanization as an existential defense. *Journal of Personality and Social Psychology*, *98*, 750-760.

Vaes, J., & Paladino, M. P. (2010). The uniquely human content of stereotypes. *Group Processes and Intergroup Relations*, *13*, 23-39.

Vaes, J., Paladino, M. P., Castelli, L., Leyens, J. P., & Giovanazzi, A. (2003). On the behavioral consequences of infrahumanization: The implicit role of uniquely human emotions in intergroup relations. *Journal of Personality and Social Psychology*, *85*, 1016-1034.

Wenzel, M., Mummendey, A., & Waldzus, S. (2007). Superordinate identities and intergroup conflict: The ingroup projection model. *European Review of Social Psychology*, *18*, 331-372.

Zebel, S., Zimmerman, A., Viki, G. T., & Doosje, B. (2008). Dehumanization and guilt as distinct but related predictors of support for reparation policies. *Political Psychology*, *29*, 193-219.