Normative Dehumanization and the Ordinary Concept of a True Human

Ben Phillips

Arizona State University

bsphilli@asu.edu

Forthcoming in *Current Research in Ecological and Social Psychology*

(Special issue: 'Painting an Integrated Portrait of Dehumanization,' edited by Alexander Landry & Katrina Fincher)

Abstract. Recently, I presented evidence that there are two broad kinds of dehumanization: *descriptive dehumanization* and *normative dehumanization*. An individual is *descriptively dehumanized* when they are perceived as less than fully human in the biological-species sense; whereas an individual is *normatively dehumanized* when they are perceived as lacking a deep-seated commitment to good moral values. Here, I develop the concept of normative dehumanization by addressing skepticism about two hypotheses that are widely held by dehumanization researchers. The first hypothesis is that dehumanization is distinct from mere dislike and other non-dehumanizing attitudes. The second hypothesis is that dehumanization is an important predictor of intergroup hostility. Across four studies, I found evidence that normative dehumanization is distinct from mere dislike, and denials of ideal humanness. I also found that it is a unique predictor of intergroup hostility. These findings suggest that research into dehumanization and intergroup hostility will benefit from recognizing the distinction between descriptive and normative dehumanization.

Keywords: dehumanization; blatant dehumanization; intergroup hostility; passive harm; active harm

1. Introduction

People sometimes explicitly deny that someone is a "true" or "real" human when they perceive them as highly immoral. For example, various websites sell t-shirts for vegans and vegetarians with the following slogan printed on it: "Real humans don't kill non-humans." In explaining these sorts of cases, I have argued that the ordinary concept HUMAN has independent *descriptive* and *normative* senses (Phillips, 2022); and that it therefore belongs to a class of social concepts known as "dual character concepts" (Knobe, Prasada, and Newman, 2013). Being human in the descriptive sense is a matter of belonging to the biological species, *Homo sapiens*; whereas being human in the normative sense is primarily a matter of having a deep commitment to good moral values.

For example, in one experiment, I found that when presented with a character, Jim, who was described as an evil *Homo sapiens*, participants tended to agree that there is a biological sense in which Jim is human, but that there is a "deeper sense" in which Jim is "not a true human after all." And when presented with a character, Xanthon, who was described as a kind and generous alien, participants tended to agree that there is a biological sense in which Xanthon is clearly not human, but that there is a "deeper sense" in which Xanthon is "a true human after all" (2022, pp. 5–9). In two follow-up experiments, I also found evidence that people regard the ability to experience emotions as somewhat central to being a true human, but as less central than good moral character (2022, pp. 11–16).

These findings suggests that we should recognize a distinction between two broad kinds of dehumanization: *descriptive dehumanization* and *normative dehumanization* (Phillips, 2022, pp. 20–21). *Descriptive dehumanization* occurs when someone is categorized as less-than-fully human (or subhuman) in the descriptive (biological) sense, while *normative dehumanization*

occurs when someone is categorized as less-than-fully human (or subhuman) in the normative sense.

Nonetheless, some basic questions about normative dehumanization still need to be addressed. One issue is whether "normative dehumanization" really is a genuine form of dehumanization. For instance, it is possible that when people agree that a highly immoral individual is not a "true human," they are merely using this phrase to express dislike or a prejudicial attitude. A related issue is whether normative dehumanization predicts intergroup hostility, over and above various sorts of non-dehumanizing attitudes. Below, I elaborate on these two issues, and outline the hypotheses that I go on to test.

1.1. Skepticism about the explanatory power of dehumanization models

Various theorists have expressed skepticism about the general hypothesis that outgroup members are often perceived as less than fully human, and that they are more vulnerable to harm as a result (Bloom, 2017, 2022; Enock et al., 2021; Enock and Over, 2022; Enock, Tipper, and Over, 2021; Lang, 2010; Manne, 2016, 2018, chapter 5; Over, 2021).

One source of skepticism concerns the observation that in paradigmatic cases of dehumanization, the perpetrators often attribute uniquely human traits to their victims, such as evilness, corruption, and criminality. This is sometimes referred to as the "paradox of dehumanization" (Smith, 2016; see also Smith, 2011, 2014, 2020). According to Over (2021, p. 6), it follows that the perpetrators in these paradigmatic cases are perceiving their victims as human after all (see also Bloom, 2017, 2022; Manne, 2016). Relatedly, some skeptics have questioned whether dehumanization is a unique predictor of intergroup hostility (Bloom, 2017, Enock and Over, 2022; Lang, 2010; Manne, 2016, 2018, chapter 5; Over, 2021). For instance, Over (2021, pp. 6–7) argues that the victims in paradigmatic cases of dehumanization are often

persecuted precisely *because* they are perceived as having the sorts of uniquely human, albeit negative, traits mentioned above.

One way to address these forms of skepticism is to invoke the distinction between *descriptive* and *normative* dehumanization. For example, the Nazis may well have perceived Jewish people as human in the *descriptive (biological)* sense; however, they may have perceived them as subhuman in the *normative* sense because they regarded them as evil, corrupt, and criminal (Phillips, 2022, pp. 17–18; see also Smith, 2023). By the same token, the Nazis may have subjected Jewish people to various forms of hostility, such as violent punishment, humiliation, and shaming, in part, because they perceived them as subhuman in the normative sense.

Another source of skepticism concerns the possibility that extant measures of dehumanization fail to distinguish between attributing negative traits to someone versus perceiving them as less than fully human (Bloom, 2022; Enock et al., 2021; Enock and Over, 2022; Enock, Tipper, and Over, 2021; Over, 2021). For instance, in responding to my finding that people tend to deny that someone is a "true human" when they perceive them as evil (Phillips, 2022), Paul Bloom raises the following concern:

> To think of someone as evil, then, is to dehumanize them, because the ideal person would not be evil. At this point, the attribution of any negative trait to others, so long as it is substantial enough, counts as dehumanization, and everything from prejudicial attitudes to moral condemnation now falls into the category. (Bloom, 2022, p. 539)

The suggestion here is that dehumanization is being conflated with the attribution of negative traits, such as those immoral traits that the *ideal* human is seen as lacking. To illustrate, consider the following analogy. Most dog owners probably have a concept of *the ideal dog*. Perhaps the ideal dog has the following sorts of traits: perfectly obedient; super friendly; extremely

intelligent; and doesn't smell bad. It does not follow that when people think of Max as a more ideal dog than Fido, they are perceiving Max as more "doglike" than the Fido. What this suggests is that the modifier "ideal" in the phrase "ideal dog" might just be functioning as a term of appraisal, much like "good" or "likeable." In the same way, if "true human" is synonymous with "ideal human," the modifier "true" in the former phrase may just be functioning as a term of appraisal. If so, this would mean that when people assert that someone is not a "true human," they are merely expressing the belief that this individual is *less than ideal*, but not the belief that this individual is *less than fully human*.

In addressing the varieties of skepticism outlined above, I aimed to test the hypothesis that judgments of true humanness come apart from judgments concerning ideal humanness (Studies 1a, 1b, and 2), as well as dislike (Study 2). I also aimed to test the hypothesis that denials of true humanness predict intergroup hostility, even when controlling for these non-dehumanizing attitudes (Study 3).

1.2. True versus ideal humanness

To see how judgments concerning true humanness might come apart from judgments concerning ideal humanness, as well as dislike, consider two individuals, Michael and William. Michael is morally perfect, whereas, William tries, but sometimes fails, to do the right thing. People may think that Michael is more of an ideal human than William: they may also like Michael more than William. However, they may think that William's moral fallibility makes him more of a true human than Michael.

Why might people regard a morally fallible character as more of a true human than a morally infallible one? One possibility is that people think that a morally fallible character is constantly fighting off the corrupting influence of irrational impulses, which requires them to maintain a

robust, emotionally laden, commitment to good moral values. In contrast, people may think that a morally infallible character is not as deeply committed to good moral values because they embody them in a relatively emotionless and "robotic" manner (for some relevant studies, see Lapka et al., 2022). This is consistent with my previous finding that people regard the ability to experience emotions as somewhat central to true humanness, but as less central than good moral character (2022, pp. 11–16).

1.3. Overview of studies

Four studies provide evidence that denials of true humanness are (i) distinct from mere dislike and denials of ideal humanness; and (ii) a unique predictor of intergroup hostility. More specifically, Studies 1a and 1b examined the hypothesis that judgments of true humanness are distinct from judgments concerning ideal humanness. Study 2 examined whether judgments of true humanness are distinct from both dislike and judgments concerning ideal humanness. This study also examined whether, in addition to good moral character, people associate emotionality with being a true human. Study 3 examined whether hostility towards outgroup members is driven by judgments concerning true humanness, over and above judgments concerning ideal humanness and dislike. The materials and data for each study are available at https://osf.io/bpr9h/. Studies 1a and 2 were not pre-registered. Study 1b (https://osf.io/3szb6) and Study 3 (https://osf.io/hztpf) were both pre-registered through OSF.

2. Study 1a

The aim of Study 1a was to examine whether denials of true humanness are distinct from denials of ideal humanness. To investigate this issue, I presented participants with a vignette describing two characters side-by-side: a morally infallible character, and a morally fallible one. Each participant was asked to rate both characters in terms of "true humanness" and "ideal

humanness." I predicted that they would rate the fallible character as more of a true human than the infallible character; and that they would rate the infallible character as more of an ideal human than the fallible character. If so, this will constitute evidence that people do not equate true humanness with ideal humanness.

Participants were also randomly assigned to one of two conditions. Those in the *infallible-more-humanlike* condition were asked to rate a statement asserting that there is a sense in which the infallible character is more "humanlike" than the fallible character; while participants in the *fallible-more-humanlike* condition were asked to rate a statement asserting that there is a sense in which the fallible character is more "humanlike" than the infallible character. The purpose of including these items was to further test the hypothesis that normative dehumanization is distinct from denials of ideal humanness. I predicted that participants would agree that there is a sense in which the fallible character is more "humanlike" than the infallible character, because they would perceive the fallible character as more of a true human. On the other hand, I predicted that participants would disagree that there is a sense in which the infallible character is more "humanlike" than the fallible character, because they do not regard idealness as central to humanness. If these predictions are correct, it will provide additional evidence that when people perceive someone as a *less-than-true* human, they are not merely perceiving them as a *less-than-ideal* human.

2.1. Materials and methods

Two hundred and twenty participants (110 female, 110 male; M_{age} = 27.8 years) were recruited from Prolific in exchange for $0.45. Sixteen participants were excluded from the final analysis because they failed the comprehension check. This brought the final sample size down to 204. An *a priori* power analysis was not performed. However, the R package *pwr* (Champley,

2020) was used to conduct a post hoc sensitivity analysis. For the main analyses reported below, the sensitivity analysis revealed that 204 participants would be sufficient to detect a small effect (Cohen's $d = 0.2$) with 80% power.

Each participant was presented with the same vignette, along with items measuring perceptions of true and ideal humanness. Condition (*infallible-more-humanlike* vs *fallible-more-humanlike*) was a between-subjects factor. The vignette read as follows (the order of information about Michael and William was counterbalanced across participants):

Michael and William differ in the following way:

- Michael is morally perfect. For example, there has never been a single situation in which he has done something immoral, such as lying, cheating, or manipulating someone for his own gain. In general, Michael aims to be kind and to do the right thing by others, and he never fails to achieve perfection in this regard.
- William is not morally perfect. For example, there have been situations in which he has done immoral things, such as lying, cheating, or manipulating someone for his own gain. However, in general, William aims to be kind and to do the right thing by others, even if he occasionally falls short of achieving perfection in this regard.

Each participant was asked to rate the extent to which Michael and William are ideal humans (0 = Very much not an ideal human, 100 = Very much an ideal human). They were also asked to rate the extent to which Michael and William are true humans (0 = Not a true human, 100 = True human). The instructions for the true humanness item read as follows:

There are different ways to think about what it means to be "human." For instance, one might think that being human is just a matter of being a member of the biological species that you

and I belong to. Biologists use the term "*Homo sapiens*" when referring to this species. One might think that there is also a deeper way of thinking about what it means to be human. According to this way of thinking, ultimately, certain individuals do not count as "true humans," even if they are members of our biological species (i.e. *Homo sapiens*).

Use the sliding scale below to indicate the extent to which [Michael/William] is or isn't a true human in this deeper sense.

Those participants who were randomly assigned to the *infallible-more-humanlike* condition were asked to rate the extent to which they agree/disagree with the following statement (1 = Strongly disagree, 7 = Strongly agree):

"There is a sense in which Michael is more humanlike than William."

Those participants who were randomly assigned to the *fallible-more-humanlike* condition were asked to rate the following statement:

"There is a sense in which William is more humanlike than Michael."

Each participant also completed a comprehension check.

2.2. Results and discussion

The mean true and ideal humanness ratings are shown in Fig. 1. As expected, participants tended to rate the infallible agent, Michael, as a more ideal human than the fallible agent, William (M = 75.7, SD = 25.5, and M = 56.3, SD = 24.8, respectively), $t(201) = 6.93$, $p < .001$, Cohen's $d = 0.77$. And, as expected, participants tended to rate the fallible agent, William, as more of a true human than the infallible agent, Michael (M = 83.8, SD = 20.0, and M = 64.8, SD = 33.6, respectively), $t(202) = -6.92$, $p < .001$, Cohen's $d = 0.69$.

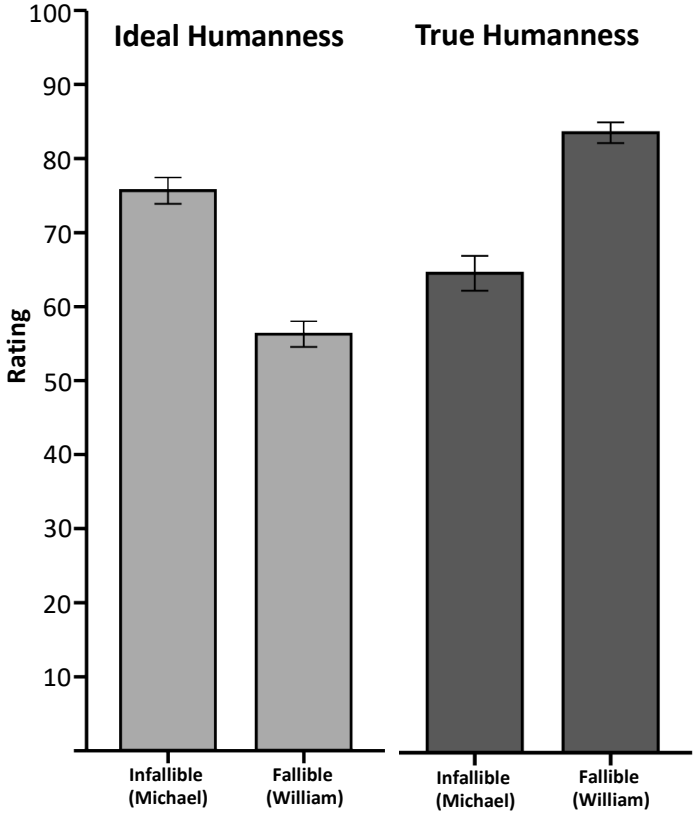

Fig. 1. Mean true humanness and ideal humanness ratings in Study 1a (error bars show SE mean).

The mean ratings for each condition (*infallible-more-humanlike* vs *fallible-more-humanlike*) are shown in Fig. 2. The mean rating for the statement asserting that there is a sense in which the fallible agent, William, is more humanlike than the infallible agent, Michael (M = 5.37, SD = 1.67) was significantly higher than the mean rating for the statement asserting that there is a sense in which the infallible agent, Michael, is more humanlike than the fallible agent, William (M = 2.52, SD = 1.43), t(196.44) = -13.05, p <.001, Cohen's *d* = 1.83.

One-sample t tests were also conducted to compare the mean ratings for each statement against the scale's midpoint. Participants tended to agree that there is a sense in which the fallible agent, William, is more humanlike than the infallible agent, Michael, t(100) = 8.24, p < .001, Cohen's *d* = 0.82; whereas they tended to disagree that there is a sense in which the

infallible agent, Michael, is more humanlike than the fallible agent, William, $t(102) = -10.45$, $p < .001$, Cohen's $d = 1.03$.

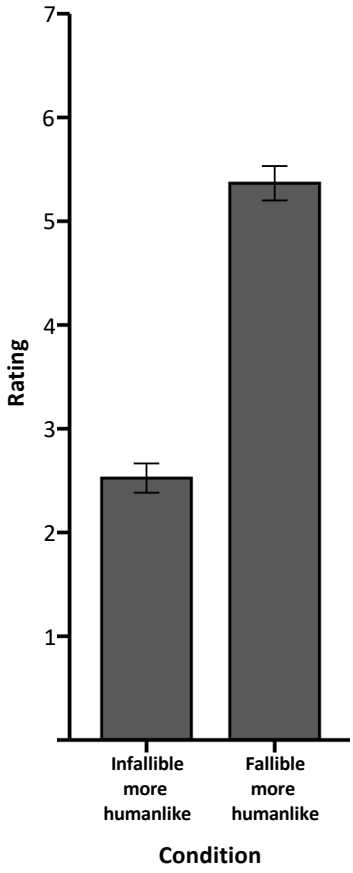

Fig. 2. Mean ratings by condition in Study 1a (error bars show SE mean).

These findings provide evidence that denials of true humanness are distinct from denials of ideal humanness. Participants tended to rate the fallible character as more of a true human than the infallible character, and they tended to rate the infallible character as more of an ideal human than the fallible character. If people tend to equate true humanness with ideal humanness, this is not the pattern of judgments that we would expect to find.

I also found evidence that when people regard one individual as more of a true human than another individual, they tend to perceive the former as more "humanlike" than the latter. In contrast, though, I found evidence that when people regard one individual as more of an ideal human than another individual, they do not necessarily perceive the former as more "humanlike" than the latter. This asymmetry in judgments concerning true and ideal humanness has two key implications. First, it provides additional evidence that people do not equate true humanness with ideal humanness, and that normative dehumanization is therefore distinct from denials of ideal humanness. Second, it suggests that when someone is perceived as a *less-than-true* human, they are thereby dehumanized (i.e. they are seen as less-than-fully humanlike); but when someone is perceived as a *less-than-ideal* human, they are not necessarily dehumanized (I elaborate on this second point in section 6.3.).

3. Study 1b

The purpose of Study 1b was to further examine the hypothesis that people do not equate true humanness with ideal humanness. Once again, participants were presented with a vignette comparing two characters, Michael and William. However, instead of asking participants to rate the true and ideal humanness of these characters, they were asked to imagine that Michael is an ideal human, and to imagine that William is a true human. Then, as in Study 1a, participants in one condition were asked whether they agree that there is a sense in which the ideal human, Michael, is more humanlike than the true human, William; while those in the other condition were asked whether they agree that there is a sense in which the true human, William, is more humanlike than the ideal human, Michael. If participants simply equate true humanness with ideal humanness, they should not agree with either statement. However, as in Study 1a, if participants agree that there is a sense in which the true human is more "humanlike" than the

ideal human, but they disagree that there is a sense in which the ideal human is more "humanlike" than the true human, this will provide more evidence that normative dehumanization is distinct from denials of ideal humanness.

3.1. Materials and methods

In Study 1a, condition (*fallible-more-humanlike* vs *infallible-more-humanlike*) had a large effect on ratings (Cohen's $d = 1.83$). For Study 1b, an *a priori* power analysis using G*Power (Faul et al., 2007) determined that 100 participants would be sufficient to detect an effect of this magnitude with 95% power. One hundred and fifteen participants (58 female, 57 male; M_{age} = 27.0 years) were therefore recruited from Prolific in exchange for $0.45. Seven participants were excluded from the final analysis because they failed the attention check. This brought the final sample size down to 108.

All participants were presented with the following vignette:

> Think about the traits of the *ideal* human. For example, you might think of the ideal human as someone who is morally perfect (i.e. someone who never does the wrong thing). You might also think of the ideal human as having various other traits. Suppose that Michael has all these traits—that is, suppose that **Michael is the *ideal* human**.
>
> Now, think about the traits of a *true* human. For example, you might think of a true human as someone who *tries* to do the right thing, even if they sometimes fail. You might also think of a true human as having various other traits. Suppose that William has all these traits—that is, suppose that **William is a *true* human.**

The order of information about Michael and William was counterbalanced across participants. Participants in the *ideal-more-humanlike* condition rated the statement, "There is a sense in

which Michael is more human-like than William (1 = Strongly disagree, 7 = Strongly agree); while those in the *true-more-humanlike* condition rated the statement, "There is a sense in which William is more human-like than Michael." Participants also completed an attention check.

3.2. Results and discussion

The mean ratings and distributions are shown in Fig. 3. Participants' mean rating for the statement asserting that there is a sense in which the true human, William, is more humanlike than the ideal human, Michael (M = 5.89, SD = 1.11) was significantly higher than their mean rating for the statement asserting that there is a sense in which the ideal human, Michael, is more humanlike than the true human, William (M = 2.60, SD = 1.51), $t(93.03) = -12.85$, $p < .001$, Cohen's $d = 2.50$.

One-sample t tests were also conducted to compare the mean ratings for each statement against the scale's midpoint. Participants tended to agree that there is a sense in which the true human, William, is more humanlike than the ideal human, Michael, $t(55) = 12.8$, $p < .001$, Cohen's $d = 1.71$; but they tended to disagree that there is a sense in which the ideal human, Michael, is more humanlike than the true human, William, $t(51) = -6.70$, $p < .001$, Cohen's $d = 0.93$.

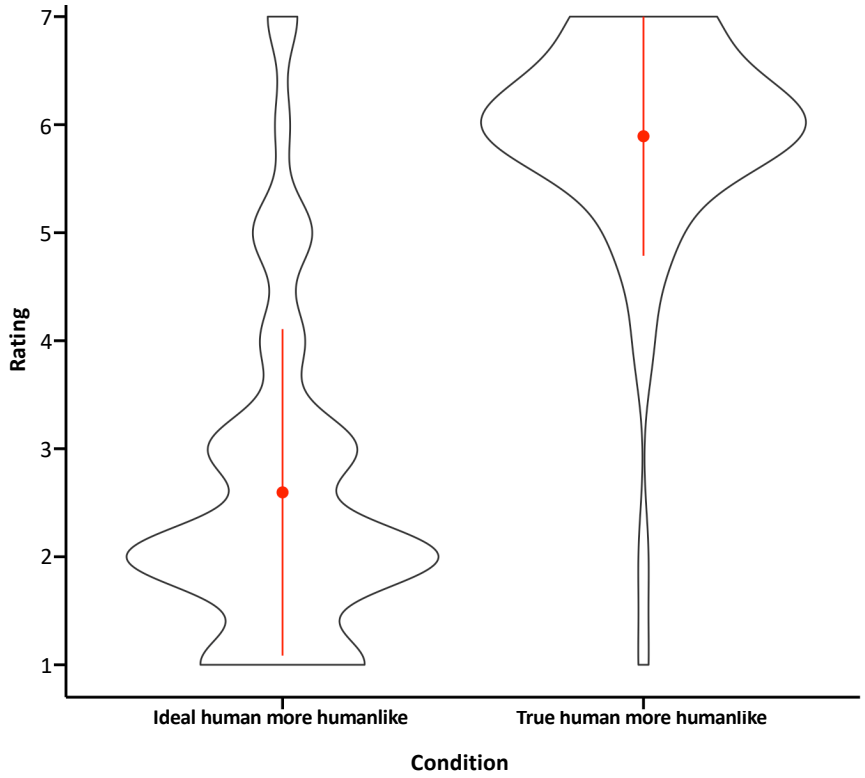

Fig. 3. Violin plots showing the distribution of ratings for each statement in Study 1b. Each plot also displays the mean ratings (red dots), and the standard deviations (red vertical lines).

These findings converge with those of Study 1a, thereby providing additional evidence for two related hypotheses. The first hypothesis is that people tend not to equate true humanness with ideal humanness. The second hypothesis is that when people perceive someone as a *less-than-true* human, they perceive them as less-than-fully humanlike; but when people perceive someone as a *less-than-ideal* human, they do not necessarily perceive them as less-than-fully-humanlike.

4. Study 2

Studies 1a and 1b provide evidence that denials of true humanness are distinct from denials of ideal humanness. The main aim of Study 2 was to build on these findings by examining whether

denials of true humanness are also distinct from dislike. Participants were presented with a vignette that describes a character as either morally infallible, morally fallible, or evil. I predicted that participants' feelings towards the morally infallible character would be as warm, if not warmer, than their feelings towards the morally fallible character. As in Studies 1a and 1b, I also predicted that they would regard the morally infallible character as more of an ideal human than the morally fallible character. Importantly, though, I predicted that participants would rate the morally fallible character as more of a true human than both the morally infallible character and the evil character. If these predictions are correct, this would suggest that when people deny that someone is a "true human," they are not just expressing dislike or the belief that this individual is less than ideal.

Another aim of Study 2 was to explore the possibility that in Studies 1a and 1b, people perceived the morally fallible character as more of a true human than the morally infallible one, in part, because they perceived the latter as less capable of experiencing emotions. If so, this would be consistent with my initial studies of normative dehumanization, which suggested that in addition to good moral character, people also regard emotionality as somewhat central to true humanness (2022, pp. 11–16).

4.1. Materials and methods

Three hundred and twenty participants (160 female, 160 male; M_{age} = 30.7 years) were recruited from Prolific in exchange for $0.60. Fifteen participants were excluded from the final analysis because they failed the comprehension check. This brought the final sample size down to 305 (the aim was 100 participants per condition). An *a priori* power analysis was not performed. However, the R package *pwr* (Champley, 2020) was used to perform a post hoc sensitivity analysis. For the one-way ANOVAs reported below, the sensitivity analysis revealed

that 305 participants would be sufficient to detect a small-to-medium effect (η_p^2= .031) with 80% power.

Participants were randomly assigned to one of three vignettes. In each case, the vignette described a character, William, as a *Homo sapiens* who is either morally perfect, morally fallible, or evil. The vignettes read as follows:

> William is a member of the human species. In other words, he is what scientists refer to as a "*Homo sapiens*" (e.g. he possesses *Homo sapiens* DNA). He is morally perfect. For example, there has never been a single situation in which he has done something immoral, such as lying, cheating, or manipulating someone for his own gain. In general, William always aims to do the right thing by others, and he never fails to achieve perfection in this regard.

> William is a member of the human species. In other words, he is what scientists refer to as a "*Homo sapiens*" (e.g. he possesses *Homo sapiens* DNA). He is not morally perfect. For example, there have been situations in which he has done immoral things, such as lying, cheating, or manipulating someone for his own gain. However, in general, William aims to do the right thing by others, even if he occasionally falls short of achieving perfection in this regard.

> William is a member of the human species. In other words, he is what scientists refer to as a "*Homo sapiens*" (e.g. he possesses *Homo sapiens* DNA). He is also evil. For example, regardless of the situation, he always lies; he always cheats; and he always manipulates others for his own gain. In general, William never aims to do the right thing by others: instead, he aims to pursue evil in every way that he can.

Participants were then asked four questions, presented in random order. One question asked them to rate the extent to which William is a "true human" (0 = Not a true human, 100 = True human). The instructions for this item were the same as those utilized in Study 1a. Participants were also asked to rate the statement, "William is an ideal human" (1 = Strongly disagree, 7 = Strongly agree); whether William is capable of experiencing emotions (1 = Very incapable, 7 = Very capable); and whether they feel cold/unfavorable or warm/favorable towards him (0 = Very cold, 100 = Very warm).

In addition to these items, participants also completed a comprehension check.[1]

4.2. Results and discussion

Table 1 displays the zero-order correlations among the variables. The normalized mean ratings for *true humanness; ideal humanness*; *feeling thermometer;* and *emotionality* are displayed in Fig. 4.

Table 1. Zero-order correlations from Study 2

	1	2	3	4
1. True Humanness		.29***	.39***	.50***
2. Ideal humanness			.73***	.30***
3. Feeling thermometer				.46***
4. Emotionality				

[1] Participants were also asked to rate the following three statements: "William embodies the true purpose of being human;" "William embodies the values that are an essential part of being human"; and "William is a typical human." The purpose of including the first two statements was to examine whether the concept of true humanness is associated with either value-based essentialism or teleological essentialism (for an overview of the different varieties of essentialism, see Neufeld, 2022). The purpose of eliciting typicality judgments was to examine whether people regard true humans as typical humans. However, given that these issues are not central to the focus of this paper, participants' responses to these items are only reported in the Supplementary Material (S1).

***p <.001

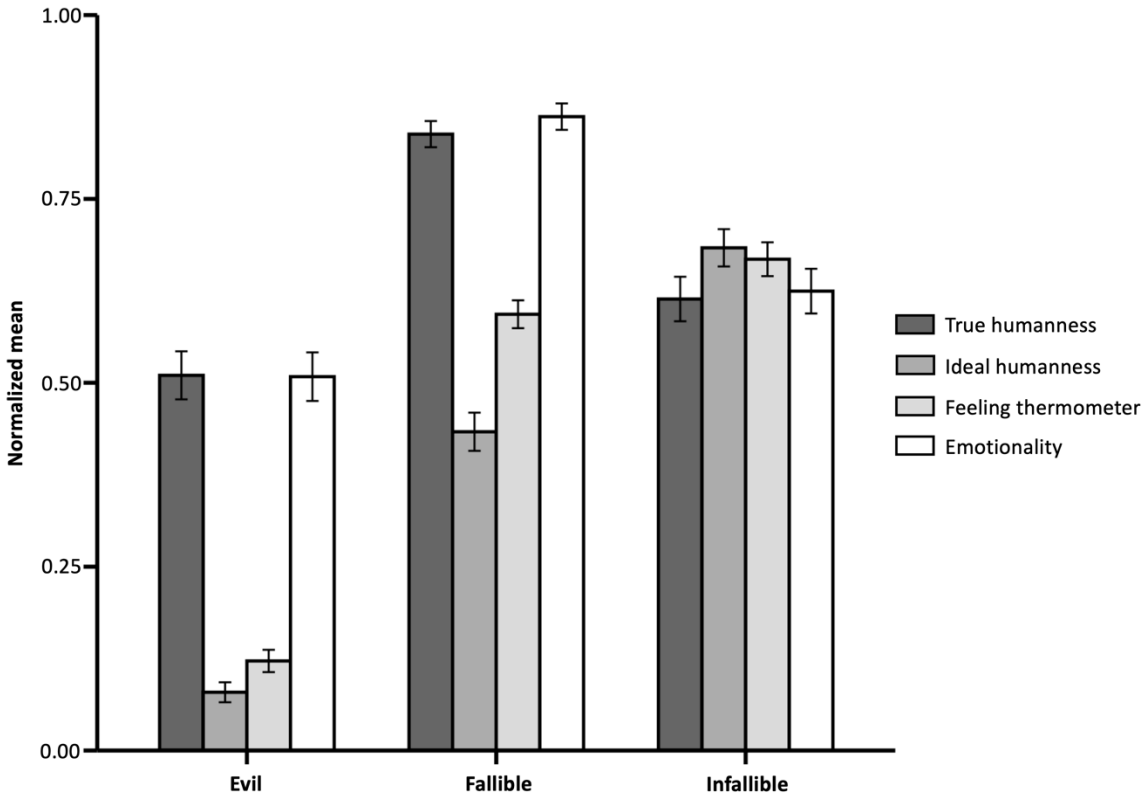

Fig. 4. Normalized mean ratings by condition in Study 2 (error bars show SE of the normalized mean).

To assess the effect of condition on perceptions of *true humanness*, *ideal humanness*, *feeling thermometer*, and *emotionality*, a series of one-way ANOVAs was carried out. In each case, post hoc comparisons were conducted using the Tukey HSD test.

The effect of condition on *true humanness* was significant, $F(2, 229) = 37.76$, $p < .001$, $\eta_p^2 = .202$. As predicted, participants tended to rate the *fallible* character as more of a true human (M = 83.8, SD = 18.2) than the *infallible* character (M = 61.3, SD = 30.4) ($p < .001$, Cohen's $d = 0.81$). They also tended to rate the *evil* character as less of a true human (M = 50.9, SD = 33.0) than both the *fallible* character ($p < .001$, Cohen's $d = 1.19$), and the *infallible* character ($p = 0.02$, Cohen's $d = 0.38$).

Condition also had a significant effect on *ideal humanness*, $F(2, 302) = 180.1$, $p < .001$, $\eta_p^2 = .544$. Participants tended to rate the *infallible* character as more of an ideal human ($M = 5.10$, $SD = 1.52$) than both the *fallible* character ($M = 3.60$, $SD = 1.60$) ($p < .001$, Cohen's $d = 1.10$) and the *evil* character ($M = 1.48$, $SD = 0.82$) ($p < .001$, Cohen's $d = 2.67$). Participants also tended to rate the *fallible* character as more of an ideal human than the *evil* character ($p < .001$, Cohen's $d = 1.56$).

The effect of condition on *feeling thermometer* was significant, $F(2, 302) = 234.8$, $p < .001$, $\eta_p^2 = .609$. Importantly, participants tended to report warmer feelings towards the *infallible* character ($M = 66.8$, $SD = 22.9$) than they did towards the *fallible* character ($M = 59.3$, $SD = 19.4$) ($p = 0.02$, Cohen's $d = 0.39$). As expected, participants tended to report colder feelings towards the *evil* character ($M = 12.2$, $SD = 15.2$) than both the *infallible* character ($p < .001$, Cohen's $d = 2.81$) and the *fallible* character ($p < .001$, Cohen's $d = 2.43$).

Finally, condition also had a significiant effect on *emotionalit*y ratings, $F(2, 302) = 43.11$, $p < .001$, $\eta_p^2 = .222$. Participants tended to rate the *fallible* character as more capable of experiencing emotions ($M = 6.17$, $SD = 1.10$) than both the *infallible* character ($M = 4.75$, $SD = 1.81$) ($p < .001$, Cohen's $d = 0.85$) and the *evil* character ($M = 4.05$, $SD = 1.99$) ($p < .001$, Cohen's $d = 1.27$). Participants also tended to rate the *infallible* character as more capable of experiencing emotions than the *evil* character ($p = 0.01$, Cohen's $d = 0.42$).

The key finding from this study is that participants tended to rate the fallible character as less likeable and as less of an ideal human than the infallible character, however, they tended to rate the fallible character as more of a true human than infallible character. This suggests that when people deny that someone is a "true human," they are not merely expressing dislike or the belief that the target is less than ideal. Given that participants also tended to rate the fallible character

as more capable of experiencing emotions than both the infallible character and the evil character, these findings are more consistent with the hypothesis that people regard both a commitment to good moral values, and the ability to experience emotions, as central to true humanness.

It is not clear why people regard both good moral character and emotionality as central to true humanness. One possibility, described above, is that people think of a true human as having an *emotionally laden* commitment to good moral values. For instance, participants may have viewed the morally infallible character as less of a true human than the morally fallible character, in part, because they perceived the infallible character as implementing good moral values in a relatively emotionless and robotic manner (I elaborate on this hypothesis below in section 6.2.).

5. Study 3

Studies 1a, 1b, and 2 suggest that denials of true humanness are distinct from mere dislike, as well as denials of ideal humanness. The aim of Study 3 was to examine whether denials of true humanness are also a unique predictor of intergroup hostility, over and above dislike and denials of ideal humanness. Participants were asked to rate the average member of twelve social groups in terms of *true humanness*; *ideal humanness*; and *feeling thermometer*. To measure hostility, I adapted the measures of active and passive harm introduced by Cuddy, Fiske, and Glick (2007). Passive harm involves diminishing a group's social worth through exclusion and neglect; while active harm involves explicitly intending to hurt a group through violence and verbal harassment (2007, p. 633). The purpose of including measures of both active and passive harm, rather than a single measure of hostility, was to assess the generalizability of any effect of normative dehumanization on intergroup hostility. For instance, it is possible that when controlling for

dislike and denials of ideal humanness, normative dehumanization is only a unique predictor of certain forms of hostility.

5.1. Materials and methods

The R package *SimR* (Green and MacLeod, 2016) was used to conduct an *a priori* sensitivity analysis for the mixed effects models reported below. The analysis indicated that 100 participants would provide these models with more than 95% power to detect an effect size of 0.15 (this was the smallest standardized coefficient found in a pilot study). One hundred and twenty participants (60 female, 60 male; M_{age} = 32.2 years) were therefore recruited from Prolific in exchange for $2. Eight participants were excluded from the final analysis because they failed the attention check. This brought the final sample size down to 112.

Each participant rated the same twelve social groups (presented in random order). In a pilot study, some of these groups were rated as highly moral (veterinarians, aid workers, grade school teachers, and firefighters); some as moderately immoral (billionaires, journalists, telemarketers, and celebrities); while others were rated as extremely immoral (pedophiles, serial killers, White supremacists, and psychopaths).

Each participant received the same nine questions, which were presented in random order. They were asked to rate the average member of each group in terms of *true humanness* (0 = Not a true human, 100 = True human). The instructions for this item were the same as those utilized in Study 1a and Study 2. Each participant was also asked to rate the average member of each group in terms of *ideal humanness* (0 = Very much not an ideal human, 100 = Very much an ideal human), and *feeling thermometer* (0 = Very cold, 100 = Very warm). The items measuring *active harm* were as follows (0 = Not at all; 50 = A moderate amount; 100 = An extreme amount):

How much do you want to *attack* the average member of each group?

How much do you want to *lash out* at the average member of each group?

How much do you want to *hit* the average member of each group?

The items measuring *passive harm* were as follows (0 = Not at all; 50 = A moderate amount; 100 = An extreme amount):

How much do you want to *socially exclude* the average member of each group?

How much do you want to *avoid* the average member of each group?

How much do you want to *limit your contact* with the average member of each group?

Participants also completed an attention check.

5.2. Results and discussion

Participants' responses to the three items measuring *active harm* were combined into a single scale (α = .96), as were their responses to the three items measuring *passive harm* (α = .95). Table 2 displays the zero-order correlations. The mean ratings for each item (by group) are displayed in Fig. 5.

Table 2. Zero-order correlations from Study 3

Correlations	1	2	3	4	5
1. True Humanness		.66***	.68***	-.69***	-.70***
2. Ideal Humanness			.88***	-.69***	-.82***
3. Feeling thermometer				-.73***	-.88***

4. Active harm	.80***
5. Passive harm	

***p <.001

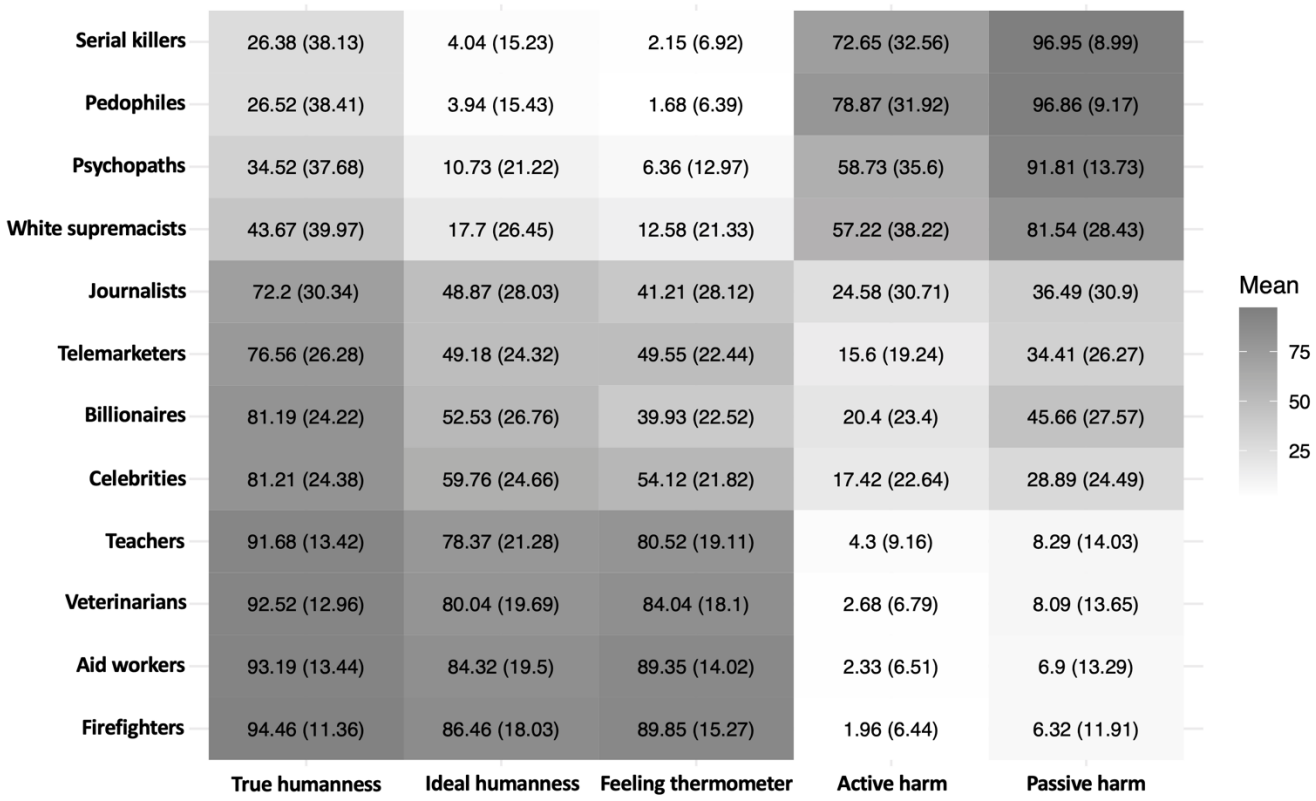

Fig. 5. Heat map displaying means for each item (by group) in Study 3 (standard deviations are in parentheses). Darker shades indicate higher mean ratings.

To examine which sorts of judgments predict active and passive harm, two linear mixed effects models were conducted in R, with the package *lme4* (Bates et al., 2015). In both models, *true humanness*; *ideal humanness*; and *feeling thermometer* were specified as fixed effects; while *participant* and *group* were specified as random effects. The two mixed models were expressed in *lme4* syntax as follows (all variables were centered and scaled):

lmer(active/passive harm) ~ true_humanness + ideal_humanness + feeling_thermometer + (1 + true_humanness + ideal_humanness + feeling_thermometer|group) + (1 true_humanness + ideal_humanness + feeling_thermometer |participant))

The results of the two models are displayed in Table 3. *True humanness*, *ideal humanness*, and *feeling thermometer* were all significant predictors of both active and passive harm. *True humanness* and *feeling thermometer* were the strongest predictors of active harm, but *true humanness* was the weakest predictor of passive harm.

Table 3. Linear mixed effects models predicting *active harm* and *passive harm* from *true humanness*, *ideal humanness*, and *feeling thermometer* in Study 3

	Active harm[a]			**Passive harm[b]**		
Predictor	*ß*	95% CI	VIF	*ß*	95% CI	VIF
(Intercept)	-0.20**	[-0.32, -0.08]		-0.16*	[-0.30, -0.02]	
True humanness	-0.25***	[-0.35, -0.16]	1.03	-0.17***	[-0.24, -0.11]	1.84
Ideal humanness	-0.20***	[-0.30, -0.10]	1.08	-0.22***	[-0.31, -0.13]	1.39
Feeling thermometer	-0.24***	[-0.34, -0.14]	1.05	-0.38***	[-0.50, -0.27]	1.51

Note. *ß* indicates standardized regression weights. VIF = variance inflation factor.

[a] Marginal $R^2 = 0.46$ (Conditional $R^2 = 0.88$)

[b] Marginal $R^2 = 0.66$ (Conditional $R^2 = 0.91$)

* indicates <.05

** indicates <.01

*** indicates p <.001

These findings indicate that normative dehumanization is an important predictor of intergroup hostility, over and above dislike and denials of ideal humanness. Together with the results of

Study 2, this suggests that when outgroup members are perceived as sufficiently immoral, they are perceived as less than fully "human," in the normative sense of the term; and are more vulnerable to both active and passive harm as a result.[2]

6. General Discussion

The findings reported here support two main hypotheses. The first hypothesis is that denials of true humanness are distinct from denials of ideal humanness (Studies 1a, 1b, and 2) and dislike (Study 2). The second hypothesis is that denials of true humanness predict intergroup hostility, even when controlling for denials of ideal humanness and dislike (Study 3).

6.1. Addressing skepticism about the explanatory power of normative dehumanization

These findings thereby address some recent doubts about dehumanization's construct validity (Bloom, 2017, 2022; Enock et al., 2021; Enock and Over, 2022; Enock, Tipper, and Over, 2021; Lang, 2010; Manne, 2016, 2018, chapter 5; Over, 2021). They suggest that when people deny that someone is a true human, they are not just expressing the belief that this individual is less-than-ideal; nor are they just expressing cold feelings towards the given individual.

It is also worth emphasizing that, *pace* Bloom (2022), the model of normative dehumanization that I have been developing here does not entail that *all* instances of moral condemnation are instances of normative dehumanization. If that were the case then participants should have perceived the morally fallible character as less of a true human than the morally infallible character: instead, it was the other way around. Moreover, in previous research, I found that people will not deny that someone is a true human just because they perceive them as engaging

[2] To address a concern about floor effects, I ran a version of each model in which the following were excluded from the random factor, *group*: White supremacists, psychopaths, pedophiles, and serial killers. These groups all received very low mean ratings for ideal humanness and feeling thermometer (see Fig. 5). However, despite removing these groups, the overall pattern of findings remained the same. See Supplementary Material (S2) for details.

in immoral behavior. Rather, people only tend to deny that someone is a true human when they perceive them as being morally bad "deep down" in their true self (Phillips, 2022, pp. 9–11; see also Strohminger and Nichols, 2014). This suggests that people do not have a concept of true humanness according to which it is about successfully *implementing* certain values: instead, it suggests that people have a concept of true humanness according to which it is about harboring a deep-seated commitment to these values, even if one fails to implement them from time to time. Thus, overall, the evidence suggests that not all instances of moral condemnation count as normative dehumanization. Instead, normative dehumanization seems to be limited to those forms of condemnation in which the target is perceived as lacking a *deep-seated* commitment to moral values.

One possibility that the studies reported here do not address directly is that when participants agree that someone is not a "true human," they are interpreting this phrase figuratively (for a relevant discussion, see Over, 2021). As an analogy, consider the scene in the film *Crocodile Dundee*, during which a mugger pulls a knife on Mick, the main character. Mick responds by saying "That's not a knife. *That's* a knife!" As he says this, he pulls out a very large knife from his jacket. In this context, the phrase "That's not a knife" is being used figuratively to convey the idea that the mugger's knife is inadequate: clearly, Mick does not think that the mugger is not holding a knife. In the same way, when people deny that someone is a "true human," they might just be using this phrase figuratively to convey the idea that this person is an inadequate human, as opposed to the idea that they are not human.

Studies 1a and 1b provide indirect evidence against this sort of view. In both cases, participants tended to agree that agent who was portrayed as less of a "true human" than the other agent seemed less humanlike; however, they tended to disagree that the agent who was

portrayed as less of an "ideal human" than the other agent seemed less humanlike. This provides support for the hypothesis that when people deny that someone is a "true human," they are not just engaging in figurative speech: instead, they appear to be reporting that the target seems relatively un-humanlike to them. In examining this issue more directly, future studies could include a measure that asks participants to specify whether they are speaking literally or figuratively when they use the phrase "not a true human."

6.2. Which traits do people encode as central to true humanness?

Another issue that warrants more attention is which traits people regard as central to true humanness. Previously, I found evidence that people regard a commitment to morally good values as more central to true humanness than various other non-moral traits and capacities (Phillips, 2022, pp. 11–15). This squares away with other studies suggesting that perceptions of moral character are a key predictor of dehumanization (e.g., see Fincher & Tetlock, 2016; Fincher, Tetlock, and Morris, 2017; Kteily and Landry, 2022; Puryear et al., 2022; Schwartz and Struch, 1989). Nonetheless, my previous research also suggests that people regard emotionality as somewhat central to true humanness, even if they regard it as *less* central than good moral character (2022, pp. 11–16). The findings from Study 2 are consistent with this hypothesis. Participants tended to perceive the fallible character as more of a true human than both the infallible character and the evil character: they also tended to perceive the fallible character as the most capable of experiencing emotions.

In further examining this issue, one possibility to consider is that people regard emotionality as necessary for maintaining the sort of commitment that is central to being a true human. To illustrate, participants may have perceived the fallible character as more of a true human than the infallible character because they perceived the former as having a more *emotionally laden*

commitment to good moral values. In other words, they may have thought that the infallible character is less of a true human because he implements good moral values in a relatively "robotic" and emotionless way (for some relevant findings, see Lapka et al., 2022). To examine this hypothesis more directly, future studies could include an item that asks participants whether the target "implements" or "displays" good moral values; as well as an item that asks them whether the target "feels deeply committed" to good moral values.[3]

6.3. True humanness, ideal humanness, and extant measures of dehumanization

Studies 1a and 1b suggest that when people regard one person as less of a true human than another person, they perceive the former as less humanlike. However, these studies also suggest that when people regard one person as less of an *ideal* human than another person, they do not necessarily perceive the former as less humanlike. This may have important implications for other measures of blatant dehumanization. For example, Kteily and colleagues' (2015) measure presents participants with the Ascent of Man image, along with a prompt asking them to rate various targets in terms of how "humanlike" and "evolved" they seem (on a scale from 0 to 100, with 100 representing someone who is maximally "evolved" and "humanlike"). It is plausible that in various contexts, participants interpret 100 as representing the *ideal* human (see Kteily and Landry, 2022). If so, the findings reported here suggest that when participants give someone a rating below 100, they may not be expressing the belief that the target is less-than-fully humanlike: instead, they may just be expressing the belief that the target is not an *ideal* human. However, it does not follow that the Ascent scale fails to measure *any* dehumanizing attitudes,

[3] The findings reported in the Supplementary Material (S1) indicate that participants did not think that the fallible character "embodies the values that are an essential part of being human" to a greater degree than the infallible character. However, this is compatible with the hypothesis that participants perceived the fallible character as more of a true human than the infallible one, in part, because they perceived the former as maintaining a more emotional commitment to good moral values.

for when the target is perceived as highly immoral, participants may utilize the scale to express the belief that the target is not a true human. Moreover, when the target is perceived as having some sort of biological impairment, participants may utilize the scale to express the belief that the target is less than fully human in the descriptive (biological) sense.

Similarly, consider Haslam and colleagues' (2005) measure of "animalistic dehumanization." Typically, participants are asked to rate the extent to which the target possesses the following sorts of traits, which people tend to regard as uniquely human: intelligence, rationality, open mindedness, culturedness, moral sensibility, etc. Plausibly, people regard these traits as reflecting the *ideal* human. If so, when people deny that someone possesses these sorts of traits to the same extent as themselves—thereby engaging in "animalistic dehumanization"—they may just be expressing the belief that this person is a less-than-ideal human, as opposed to the belief that they are less-than-fully humanlike. This is consistent with some recent studies suggesting that when *undesirable* uniquely human traits, such as corruption, are also included in measures of "animalistic dehumanization," it becomes indistinguishable from ingroup bias and stereotyping (Enock et al., 2021; however, see Vaes, 2023, for a reply).

6.4. Normative dehumanization and intergroup hostility

Study 3 provides evidence that normative dehumanization predicts both active and passive harm, over and above dislike and judgments concerning ideal humanness. Nonetheless, Study 3 has various limitations that could be addressed in future research.

First, it is worth noting that in Study 2, the morally infallible agent was perceived as less of a true human than the morally fallible agent; however, participants had slightly more favorable feelings towards the infallible agent. This suggests that in certain cases, normative dehumanization is *not* linked to harm. In fact, various other studies suggest that there are

interesting exceptions to the generalization that dehumanized targets are vulnerable to harm (e.g., see Bastian et al., 2013; Vaes & Muratore, 2013; Vaes et al., 2021; see also Nussbaum, 1995).

Another limitation of Study 3 is that the measure of normative dehumanization that was utilized does not elicit categorical judgments about true humanness: instead, it asks participants to indicate the *degree* to which the target counts as a true human. In further examining the relation between normative dehumanization and intergroup hostility, it will be important to examine the role played by categorical denials of true humanness. For instance, it is possible that categorical denials of true humanness predict certain varieties of active harm, such as violent retribution, because the individual in question is seen as categorically subhuman and thus irredeemable (for some relevant findings, see Maffly-Kipp et al., 2022; Martin and Heiphetz, 2021).

Directly examining whether certain forms of hostility are uniquely driven by *categorical* denials of true humanness will also address a recent concern raised by Smith (2023), who points out that influential models of dehumanization tend to construe it as coming in degrees. Smith regards this as problematic because in paradigmatic cases of dehumanization, such as Nazi antisemitism, the perpetrators did not just think of their victims as less than fully human: they thought of them as categorically subhuman.

While the studies reported here suggest that people regard true humanness as coming in degrees, this is entirely compatible with the thesis that in certain cases, people will categorically deny that someone is a true human. In fact, my initial research into normative dehumanization utilized a categorical measure (Phillips, 2022). This measure did not ask participants to indicate the degree to which the target is a true human: instead, it asked participants whether they agree that the target is "not a true human." Thus, to examine the forms of hostility that tend to occur in

paradigmatic cases of dehumanization, such as Nazi antisemitism, future studies could utilize this measure of categorical dehumanization (a forced-choice question that simply asks participants whether the target is, or is not, a true human could also be utilized).

A third limitation of Study 3 is that it did not include a measure of descriptive dehumanization (i.e. denials of biological humanness). It is likely that descriptive and normative dehumanization give rise to distinct patterns of hostility, including patterns that are not necessarily captured by the distinction between active and passive harm. For example, they might fuel different forms of punishment (retributivist versus restorative); different forms of violence (instrumental versus moral); and distinct emotional responses on the part of the dehumanizer (for an overview of relevant studies, see Kteily and Landry, 2022, p. 229). To illustrate, consider Rai and colleagues' (2017) finding that when dehumanization is construed as the denial of mind, it predicts instrumental, but not moral, violence. It is possible that mind denial is primarily a form of descriptive dehumanization, and that this explains why it does not predict moral violence (for a highly relevant discussion, see Fincher, Kteily, and Bruneau, 2018).

One way to systematically examine all these issues will be to study various outgroups that, from the perspective of the dehumanizer, occupy different locations on the graph depicted in Fig. 6, with the aim of assessing whether these different groups are subjected to distinctive patterns of hostility:

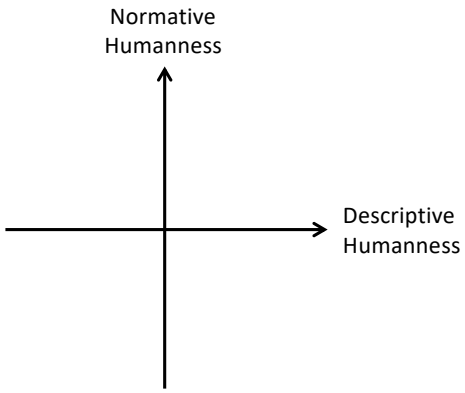

Fig. 6. The two dimensions of humanness.

For instance, when people see a political or ideological opponent as belonging to the same racial group as them, they might place them relatively high up on the descriptive dimension, but relatively low down on the normative dimension (for some relevant research, see Puryear et al., 2022). Moreover, if people tend to view traits such as intelligence and rationality as central to being human (in descriptive sense), they may view their political and ideological opponents as especially threatening because they perceive them as "entitative" groups who are highly capable of implementing their morally deviant values (see Phillips 2022b, 2022c). This is arguably how the Nazis tended to think of Jewish people: namely, as humans in the descriptive sense, with traits such as intelligence and rationality, but as subhumans in the normative sense, with traits such as evilness and criminality (Phillips, 2022, p. 18; see also Steizinger, 2018).

On the other hand, people might place individuals with certain physical or cognitive disabilities relatively high up on the normative dimension, but relatively low down on the descriptive dimension. For example, some individuals may see little people as less than fully human in the descriptive sense, but they may see them as fully human in the normative sense (for some relevant research, see Kunst, Kteily, and Thomsen, 2019).

Finally, people might place the members of those racial outgroups whom they regard as morally deviant at the lower end of both dimensions. For example, during the 1600s, Morgan Godwyn, a minister in the Church of England, travelled through Virginia and Barbados to spread the Gospel to enslaved Africans. When he return to England, Godwyn wrote a book in which he reported that fellow Englishmen told him "That the Negro's, though in their Figure they carry some resemblance of Manhood, yet are indeed *no Men*" (1708, p. 3). He also stated that White Colonialists saw Africans as "Creatures destitute of Souls, to be ranked among Brute Beasts, and treated accordingly" (1708, p. 3). This characterization of African slaves as beasts with no souls suggests that the Colonialists may have thought of them as subhuman in both the descriptive and the normative sense (for further discussion of this case study, see Smith, 2016, pp. 420–421; and 2020, chapter 9).

7. Conclusion

I found evidence that normative dehumanization is distinct from denials of ideal humanness and mere dislike. I also found evidence that normative dehumanization is a unique predictor of intergroup hostility. Future work on dehumanization could benefit from focusing on the distinction between descriptive and normative dehumanization. As was argued above, doing so might explain some of the paradoxical features of dehumanization. Focusing on this distinction may also reveal a more nuanced picture of the relations between dehumanization and intergroup hostility.

Funding

This research was supported by a *Publication Development Grant* from *The Institute for Humanities Research (IHR)* at Arizona State University.

References

Bastian, B., Jetten, J., Chen, H., Radke, H., Harding, J. F., & Fasoli, F. (2013). Losing our humanity: The self-dehumanizing consequences of social ostracism. *Personality and Social Psychology Bulletin*, 39, 156–169. https://doi.org/10.1177/0146167212471205

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bloom, P. (2017, November 20). The root of all cruelty? *The New Yorker*. Retrieved from https://www.newyorker.com/magazine/2017/11/27/the-root-of-all-cruelty

Bloom, P. (2022). If everything is dehumanization, then nothing is. *Trends in Cognitive Sciences*, 26, 539.

Champley, S. (2020) Package 'pwr': basic functions for power analysis. Version 1.3–0. https://cran.r-project.org/web/packages/pwr/pwr.pdf

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology, 92*(4), 631–648. https://doi.org/10.1037/0022-3514.92.4.631

Enock, F. E., Flavell, J. C., Tipper, S. P., & Over, H. (2021). No convincing evidence outgroups are denied uniquely human characteristics: Distinguishing intergroup preference from trait-based dehumanization. *Cognition*, *212*. https://doi.org/10.1016/j.cognition.2021.104682

Enock, F. E., & Over, H. (2022). Reduced helping intentions are better explained by the attribution of antisocial emotions than by 'infrahumanization'. *Scientific Reports*, 12, 7824. https://doi.org/10.1038/s41598-022-10460-0

Enock, F. E., Tipper, S. P., & Over, H. (2021). Intergroup preference, not dehumanization, explains social biases in emotion attribution. *Cognition*, 216, 104865: https://doi.org/10.1016/j.cognition.2021.104865.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fincher, K. M., Kteily, N. S., & Bruneau, E. G. (2018). Our humanity contains multitudes: Dehumanization is more than overlooking mental capacities. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(15), E3329-E3330. https://doi.org/10.1073/pnas.1800359115

Fincher, K. M., & Tetlock, P. E. (2016). Perceptual dehumanization of faces is activated by norm violations and facilitates norm enforcement. *Journal of Experimental Psychology: General, 145*(2), 131–146. https://doi.org/10.1037/xge0000132

Fincher, K. M., Tetlock, P. E., & Morris, M. W. (2017). Interfacing with faces: Perceptual humanization and dehumanization. *Current Directions in Psychological Science*, *26*(3), 288–293. https://doi.org/10.1177/0963721417705390

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. https://doi.org/10.1111/2041-210X.12504

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.

Knobe, J., Prasada, S., & Newman, G. E. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition*, *127*(2), 242–257.

Kteily, N., Bruneau, E., Waytz, A., & Cotterill, S. (2015). The ascent of man: Theoretical and empirical evidence for blatant dehumanization. *Journal of Personality and Social Psychology*, *109*(5), 901–931.

Kteily, N. S., & Landry, A. P. (2022). Dehumanization: Trends, insights, and challenges. *Trends in Cognitive Sciences, 26*(3), 222–240.

Kunst, J. R., Kteily, N., & Thomsen, L. (2019). "You little creep": Evidence of blatant dehumanization of short groups. *Social Psychological and Personality Science*, *10*(2), 160–171.

Lang, J. (2010). Questioning dehumanization: Intersubjective dimensions of violence in the Nazi concentration and death camps. *Holocaust and Genocide Studies*, *24*(2), 225–246.

Lapka, S. P., Kung, F. Y. H., Brienza, J. P., & Scholer, A. A. (2022). Determined yet dehumanized: People higher in self-control are seen as more robotic. *Social Psychological and Personality Science*. https://doi.org/10.1177/19485506221093109

Maffly-Kipp J., Flanagan, P. N., Schlegel, R. J., & Vess, M. (2022). True self-attributions shape judgments of blame in the context of addiction-relevant crime. *Addictive Behaviors*, 130, 107287. doi: 10.1016/j.addbeh.2022.107287.

Manne, K. (2016). Humanism: A critique. *Social Theory and Practice*, *42*, 389–415.

Manne, K. (2018). *Down girl*. Oxford, England: Oxford University Press.

Martin, J. W., & Heiphetz, L. (2021). "Internally wicked": Investigating how and why essentialism influences punitiveness and moral condemnation. *Cognitive Science*, 45(6), e12991. doi: 10.1111/cogs.12991. PMID: 34170019.

Neufeld, E. (2022). Psychological essentialism and the structure of concepts. *Philosophy Compass*, 17(5), e12823. https://doi.org/10.1111/phc3.12823

Nussbaum, M. (1995). Objectification. *Philosophy and Public Affairs*, 24(4), 249–291.

Over, H. (2021). Seven challenges for the dehumanization hypothesis. *Perspectives on Psychological Science*, *16*, 3–13. https://doi.org/10.1177/1745691620902133

Phillips, B. (2022). 'They're not true humans': Beliefs about moral character drive denials of humanity. *Cognitive Science*, 46(2), e13089. https://doi.org/10.1111/cogs.13089.

Phillips, B. (2022b). The roots of racial categorization. *Review of Philosophy and Psychology*, 13, 151–175. https://doi.org/10.1007/s13164-021-00525-w.

Phillips, B. (2022c). Entitativity and implicit measures of social cognition. *Mind & Language*, 37(5), 1030–1047. https://doi.org/10.1111/mila.12350

Puryear, C., Kubin, E., Schein, C., Bigman, Y., & Gray, K. (2022). Bridging political divides by correcting the basic morality bias. https://doi.org/10.31234/osf.io/fk8g6

Rai, T. S., Valdesolo, P., & Graham, J. (2017). Dehumanization increases instrumental violence, but not moral violence. *Proceedings of the National Academy of Sciences*, *114*(32), 8511–8516.

Schwartz, S. H., & Struch, N. (1989). Values, stereotypes, and intergroup antagonism. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Stroebe (Eds.), *Stereotyping and prejudice: Springer series in social psychology*. New York: Springer.

Smith, D. L. (2011). *Less than human: Why we demean, enslave, and exterminate others*. New York: St. Martin's Press.

Smith, D. L. (2014). Dehumanization, essentialism, and moral psychology. *Philosophy Compass*, *9*(11), 814–824.

Smith, D. L. (2016). Paradoxes of dehumanization. *Social Theory and Practice*, *42*(2), 416–443.

Smith, D. L. (2020). *On inhumanity: Dehumanization and how to resist it*. Oxford: Oxford University Press.

Smith, D. L. (2023). Some conceptual deficits of psychological models of dehumanization, *Current Research in Ecological and Social Psychology*, 4, 100117. https://doi.org/10.1016/j.cresp.2023.100117

Steizinger, J. (2018). The significance of dehumanization: Nazi ideology and its psychological consequences. *Politics, Religion and Ideology*, *19*(2), 139–157.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, 131, 159–171.

Vaes, J., & Muratore, M. (2013). Defensive dehumanization in the medical practice: A cross-sectional study from a health care worker's perspective. *British Journal of Social Psychology*, 52, 180–190. https://doi.org/10.1111/bjso.12008

Vaes, J., Paladino, M. P., & Haslam, N. (2021). Seven clarifications on the psychology of dehumanization. *Perspectives on Psychological Science*, *16*(1), 28–32. https://doi.org/10.1177/1745691620953767

Vaes, J. (2023). Dehumanization after all: Distinguishing intergroup evaluation from trait-based dehumanization. *Cognition*, 231, 105329. DOI: 10.1016/j.cognition.2022.105329