Normative Dehumanization and the Ordinary Concept of a True Human

Ben Phillips

Arizona State University

bsphilli@asu.edu

Abstract. Phillips (2022) argues that we should distinguish between 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 sense; whereas an individual is normatively dehumanized when they are perceived as lacking a deep-seated commitment to good moral values. In what follows, I develop this model by examining whether normative dehumanization is distinct from dislike and other sorts of non-dehumanizing judgments. In doing so, I also examine which form of essentialism predicts normative dehumanization. Across four experiments, I found evidence that normative dehumanization is predicted by value-based essentialism, even when controlling for dislike, as well as judgments concerning ideal and typical humanness. I also found that normative dehumanization is a unique predictor of intergroup hostility. Together, these findings clarify what it means to normatively dehumanize someone. They also suggest that research into dehumanization will benefit from recognizing the distinction between descriptive and normative dehumanization.

Keywords: dehumanization; dual character concepts; essentialism; intergroup hostility

1. Introduction

People sometimes explicitly deny that someone is a "true" or "real" human when they perceive them as highly immoral. For example, in 2016, *Amnesty Ireland* posted an image of some women protesting Ireland's abortion laws. One person responded with the remark, "Real

humans don't kill humans" (Anonymous, 2016). In 2015, the organization, *Animal Equality*, posted a quote from Jeremy Bentham regarding the moral status of animals on their Facebook page. One person replied with the following comment: "There are too many subhumans that don't care about other people let alone an animal. A true human being cares about all living beings" (Anonymous, 2015). In the same vein, various websites sell t-shirts for vegans and vegetarians with the following slogan printed on it: "Real humans don't kill non-humans."

A striking feature of the last two examples is that the speaker appears to be invoking two, independent, conceptions of humanness. On the one hand, they are drawing a *biological* distinction between humans and nonhuman animals. On the other hand, they appear to be invoking a *normative* distinction between "real" or "true" humans and those people who mistreat biologically nonhuman animals.

In addressing these sorts of cases, Phillips (2022) has argued that the ordinary concept HUMAN has independent *descriptive* and *normative* senses. 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 a matter of harboring a deep commitment to good moral values. For example, in one experiment, Phillips (2022, pp. 5–9) 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 not human, but that there is a "deeper sense" in which Xanthon is "a true human after all."

Phillips argues that HUMAN therefore belongs to a class of social concepts known as "dual character concepts" (Knobe, Prasada, and Newman, 2013). Research suggests that dual character concepts have two dimensions for categorization: one descriptive and one normative. These dimensions appear to be orthogonal, in the sense that someone can count as a member of the given category with respect to one dimension, but as a non-member with respect to the other dimension (for an overview, see Reuter, 2019). To illustrate, consider the ordinary concept SCIENTIST. Knobe, Prasada, and Newman (2013) presented participants with someone who is employed as a scientist, but who is not committed to scientific values. They found that participants tended to agree that there is a sense in which this person is clearly a scientist, but that there is also a "deeper sense" in which they are not a "true scientist." In contrast, when presented with someone who is not employed as a scientist, but who embodies scientific values in their day-to-day life, participants tended to agree that there is a sense in which this person is clearly not a scientist, but that there is a "deeper sense" in which they are a "true scientist."

1.1. Descriptive versus normative dehumanization

If HUMAN is a dual character concept, this has important implications for research into the nature and consequences of dehumanization. First and foremost, it 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 in the descriptive (biological) sense. Those who are perceived as having atypical biological features, such as individuals with physical disabilities, may be especially vulnerable to this form of dehumanization. *Normative dehumanization* occurs when someone is categorized as less-than-fully human in the normative sense. As was explained above, Phillips' (2022) findings suggest that those individuals who are

perceived as highly immoral are especially vulnerable to this form of dehumanization (for some relevant findings, see Fincher & Tetlock, 2016; Fincher, Tetlock, and Morris, 2017; Kteily and Landry, 2022; Puryear et al., 2022; Schwartz and Struch, 1989).

Examining whether the members of various social groups are vulnerable to descriptive dehumanization, normative dehumanization, or both, may deepen our understanding of the sorts of hostility to which they are subjected. Relatedly, it might explain why certain measures of dehumanization predict certain forms of violence, but not others. For instance, it might explain 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). More generally, it is plausible that those measures that exclusively detect *descriptive* dehumanization only tend to predict instrumental violence; while those measures that detect *normative* dehumanization tend to predict moral violence as well (I elaborate on this issue below in section 6.4).

Of course, it is likely that certain measures predict *both* sorts of violence because they detect both kinds of dehumanization. For example, consider Kteily and colleagues' (2015) Ascent scale. Participants are asked to rate how "humanlike" and "evolved" various targets seem after having viewed the Ascent of Man image, which depicts a sequence of ape-like creatures, through to a modern human. Participants might utilize the scale to express the belief that the target is "unevolved" in the descriptive (biological) sense; "unevolved" in the normative (moral) sense; or both (see Phillips, 2022, pp. 20–21).

Finally, distinguishing between descriptive and normative dehumanization provides a solution to the "paradox of dehumanization," which concerns the observation that in paradigmatic cases

of dehumanization, the perpetrators often attribute uniquely human traits to their victims, such as criminality and evilness (Smith, 2016). One way to resolve the paradox is to maintain that the perpetrators think of their victims as subhuman in the normative sense, but as human in the descriptive (biological) sense (see Phillips, 2022, pp. 17–18).

In sum, then, recognizing a distinction between descriptive and normative dehumanization may lead to progress on various fronts of dehumanization research. Nonetheless, some basic questions about normative dehumanization still need to be addressed. In this paper, I examine two related issues. The first issue concerns whether we should, in fact, construe "normative dehumanization" as 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. The second issue concerns which sort of essentialism, if any, drives normative dehumanization. Below, I explain how these two issues are related, and outline the hypotheses that I go on to test.

1.2. 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 stems from the observation that the perpetrators of intergroup violence often characterize their victims as evil, corrupt, and criminal. As Over (2021, p. 6) points out, it seems unlikely that the perpetrators would apply these sorts of terms to nonhuman animals (see also, Manne, 2016; Bloom, 2017, 2022). One way to address this sort of concern is to invoke the distinction between *descriptive* and *normative* dehumanization. For example, the

Nazis may have perceived Jewish people as subhuman in the *normative* sense precisely because they perceived them as evil, corrupt, and criminal (for further discussion, see Phillips, 2022, pp. 17–18).

Another source of skepticism concerns the worry that extant measures of dehumanization fail to adequately 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 Phillips' (2022) finding that people tend to deny that someone is a "true human" when they perceive them as evil, 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. (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. This suggests that when people assert that someone is not a "true human," they might

be expressing the belief that this individual is *less than ideal*, but not the belief that this individual is *less than fully human*.

In addressing these sorts of concerns, I aimed to test the hypothesis that judgments concerning true humanness are predicted by perceptions of moral character, even when controlling for dislike, and perceptions of ideal humanness. I also aimed to test the hypothesis that denials of true humanness predict hostility, even when controlling for these sorts of judgments.

1.3. What is a true human?

To see how judgments concerning true humanness and judgments concerning ideal humanness might come apart, 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 someone like Michael is more of an ideal human than someone like William. However, they may think that William's moral fallibility makes him more of a true human than Michael. Why might people regard moral fallibility as central to true humanness? Below, I outline some competing hypotheses.

1.3.1. Value-based essentialism and true humanness

Consider what Bailey, Knobe, and Newman (2021) call "value-based essentialism": the belief that embodying certain values is essential for being a member of the given category (see also Newman and Knobe, 2019). Bailey, Knobe, and Newman found that people tend to regard the embodiment of certain values as essential for being a member of various social groups, such as *Christians, Hippies*, and *Southerners*. They also found that value-based essentialism has some of the same hallmarks as biological essentialism. For example, both forms of essentialism tend to elicit a belief in deep similarity among category members, as well as a distinction between appearance and reality.

It is plausible that judgments concerning true humanness are driven by value-based essentialism, for people might think that embodying certain moral values is essential to being a true human. This may well explain why people tend to deny that evil individuals are true humans; however, it is not obvious how it would accommodate the hypothesis, described above, according to which the morally fallible person is perceived as more of a true human than the morally perfect person.

One possibility is that people will regard the fallible person as more of a true human than the infallible person because they will perceive the former as embodying essential human values to a greater degree than the latter. For instance, people might regard the goal of overcoming one's imperfections and irrational impulses as an essential human value: one that the infallible character does not possess. Another possibility is that when judging true humanness, people distinguish between being committed to certain values versus behaving in ways that are consistent with these values. For example, people may 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).

It is also worth considering the simple hypothesis that people regard the ability to experience emotions as central to true humanness. If so, people might think that the morally fallible agent is more of a true human than the morally infallible agent, simply because they perceive the former as more capable of experiencing emotions than the latter.

1.3.2. Teleological essentialism and true humanness

Judgments concerning true humanness might also be driven by teleological essentialism: the belief that having a certain purpose is essential for membership in the given category. There is evidence that people are teleological essentialists about a wide array of categories, including biological, social, and artifactual categories (Bloom, 1996, 2007; Kelemen, 1999; Kelemen & Rosset, 2009; Piaget, 2017; Rose & Nichols, 2020; cf. Neufeld 2021, 2022). People might regard those individuals who *strive* to do the morally right thing, despite their fallibility, as embodying the essential purpose of being a true human.

1.4. Two dimensions of normative humanness?

Finally, it is worth exploring the possibility that *in addition to* the concept of true humanness, people also possess a Platonic-essentialist concept of humanness. As Kteily and Landry (2022, pp. 231–233) point out, people tend to think of humans as sitting atop a hierarchy of beings, those lower down in the hierarchy being seen as less valuable and less worthy of moral consideration (for some extensive discussions of dehumanization and hierarchical thinking, see Smith, 2011, 2014, 2020). Thus, it is plausible that people possess a Platonic concept of humanness: one that is distinct from the concept of true humanness.

Following Kteily and Landry (2022, pp. 231–233), I will construe Platonic essentialism about a given category, *C*, as having the following features: Membership in *C* is seen as graded in the sense that *C* is associated with an ideal, such that if *X* is represented as resembling this ideal to a greater extent than *Y*, then *X* is represented as more *C*-like than *Y*. Thus, to possess a *Platonic-essentialist* concept of humanness would be to (i) associate humanness with a representation of the ideal human (e.g. the morally perfect individual); and (ii) represent *X* as more humanlike than *Y* if *X* is perceived as closer to this ideal. In contrast, to be a *value-based* or *teleological*

essentialist about humanness would be to (i) associate humanness with certain values or purposes; and (ii) to represent *X* as more humanlike than *Y* only if *X* is perceived as embodying these values or purposes to a greater extent than *Y* (regardless of whether *X* is also perceived as more of an ideal human than *Y*).

To illustrate how Platonic essentialism can come apart from value-based and teleological essentialism, consider two individuals, Peter and Frank. They have the same values and purposes; however, Peter is more intelligent, more charming, and more cultured than Frank. People may regard someone like Peter as closer to the ideal human than someone like Frank; however, they may not regard either person as more of a true human than the other person, because they perceive Peter and Frank as possessing the same values and purposes.

Similarly, reconsider the morally perfect character, Michael, and the morally fallible character, William. People may think that there is a sense in which Michael is more humanlike than William because Michael is closer to the Platonic ideal of a human. At the same time, though, people may think that there is a sense in which William is more humanlike than Michael, because William is more of a "true human" than Michael. If so, this would suggest that we need to recognize two distinct dimensions of normative humanness: one corresponding to the concept of a true human; and the other corresponding to the concept of the ideal human.

1.5. Overview of experiments

Four experiments provided evidence that denials of true humanness are (i) distinct from mere dislike, as well as perceptions of ideal humanness; (ii) predicted by valued-based essentialism; and (iii) a unique predictor of intergroup hostility. More specifically, Experiment 1 examined whether judgments of true humanness are driven by value-based essentialism, teleological essentialism, or Platonic essentialism: it also examined whether denials of true humanness are

distinct from expressions of dislike. Experiments 2a and 2b examined the hypothesis that in addition to the concept of true humanness, people also have a distinctively Platonic concept of humanness. Experiment 3 examined whether hostility towards outgroup members is driven by judgments concerning true humanness, over and above judgments concerning ideal humanness; judgments concerning typical humanness; and like/dislike. The materials and data for each experiment are available at https://osf.io/bpr9h/.

2. Experiment 1

Experiment 1 had two aims. The first aim was to examine whether denials of true humanness are distinct from mere dislike, as well as perceptions of ideal humanness. The second aim was to examine whether normative dehumanization is predicted by value-based essentialism, as opposed to either Platonic or teleological essentialism.

Participants were presented with a vignette that described a character as either morally infallible, morally fallible, or evil. The purpose of manipulating perceptions of moral character in this way was to determine if judgments concerning ideal humanness can come apart from judgments concerning true humanness. I predicted that participants would rate the morally infallible character as more of an ideal human than both the morally fallible character and the evil character; however, I predicted that they would not rate the morally infallible character as more of a true human than the morally fallible character. If so, this would constitute evidence that people do not possess a Platonic concept of true humanness; and that when people deny that someone is a "true human," they are not merely expressing the belief that this individual is *less than ideal*.

Participants were also asked to rate how warm/favorable or cold/unfavorable they feel towards the target. The purpose of including this measure was to assess whether denials of true

humanness are distinct from mere dislike. Participants were also asked to rate the extent to which the target counts as a *typical* human. The purpose of eliciting perceptions of typical humanness was to examine the possibility that when people characterize someone as a "true human," they are merely expressing the belief that this individual is a typical human.

Finally, participants were asked to rate the extent to which the target is capable of experiencing emotions. The purpose of including this measure was to assess whether people regard emotional capacities as necessary for being a true human. For example, people might regard the morally infallible character as less of a true human than the morally fallible character, in part, because they see the infallible character as less capable of experiencing emotions.

2.1. 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).

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 the following eight questions, presented in random order. Each participant was asked to rate the extent to which William is a "true human" (0 = Not a true human, 100 = True human). The instructions 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 William is or isn't a true human in this deeper sense.

To assess whether beliefs about true humanness are predicted by value-based, Platonic, or teleological essentialism, participants were asked to indicate the extent to which they agree/disagree with the following statements (1 = Strongly disagree, 7 = Strongly agree):

"William is an ideal human"

"William embodies the true purpose of being human."

"William embodies the values that are an essential part of being human."

Participants were also asked to indicate the extent to which they agree with the statement, "William is a typical human" (1 = Strongly disagree, 7 = Strongly agree).

To assess whether beliefs about true humanness are predicted by either attributions of emotionality or mere like/dislike, participants were asked to specify 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.

2.2. Results and discussion

Table 1 displays the zero-order correlations among the variables. The mean ratings for *values*, *purpose*, *ideal humanness*, *typicality*, and *emotionality* (by condition) are displayed in Fig. 1; while the mean true humanness ratings (by condition) are displayed in Fig 2.

Table 1. Zero-order correlations from Experiment 1

	1	2	3	4	5	6	7
1. True Humanness		.48***	.42***	.29***	.50***	.46**	.39***

2. Values	.73***	.61***	.47***	.46**	.71***
3. Purpose		.66***	.44***	.40***	.72***
4. Ideal humanness			.30***	.10	.73***
5. Emotionality				.44***	.46***
6. Typicality					.29***
7. Feeling thermometer					

**p <.01

***p <.001

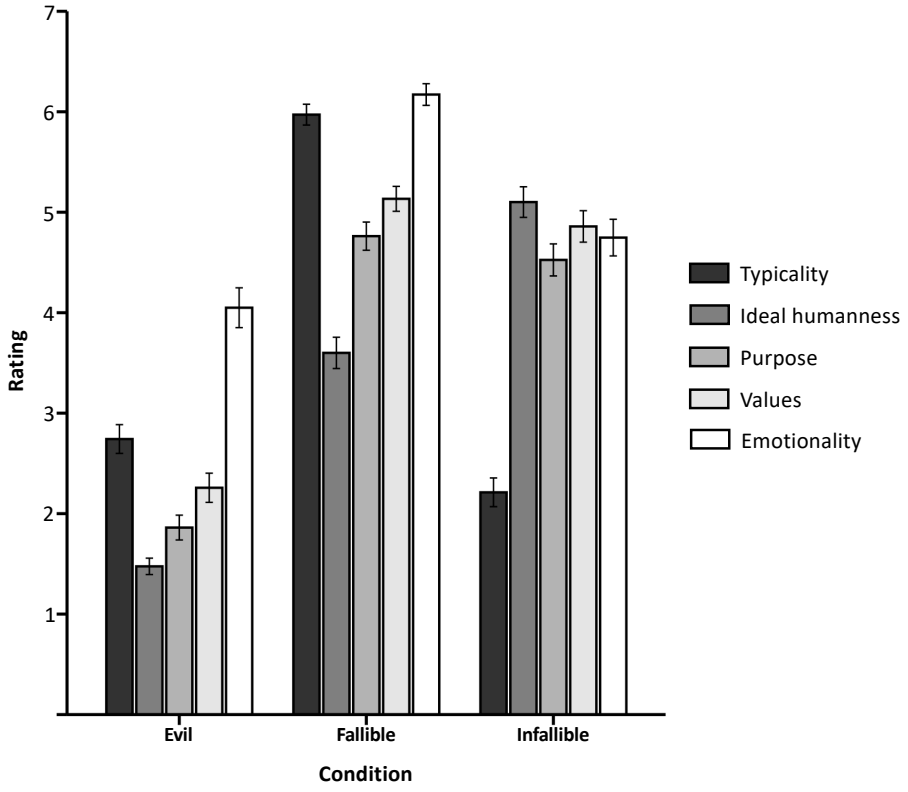


Fig. 1. Mean statement ratings by condition in Experiment 1 (error bars show SE mean).

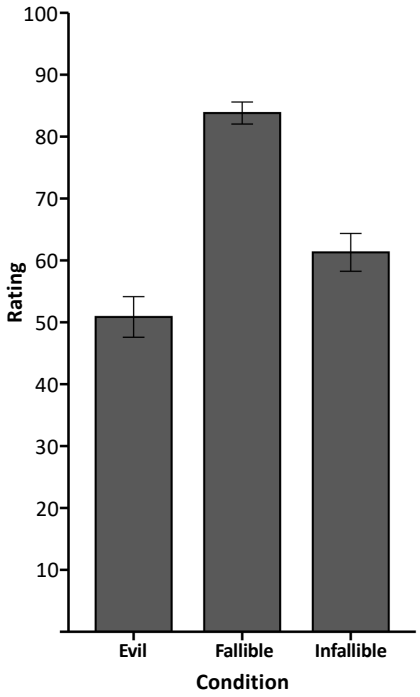


Fig 2. Mean true humanness ratings by condition in Experiment 1 (error bars show SE mean).

To examine which sorts of judgments predict true humanness ratings, I ran a regression model with *true humanness* as the dependent variable, and the following as independent variables: *Condition* (dummy coded, with *evil* as the reference group), *values*, *emotionality*, *ideal humanness*, *purpose*, *typicality*, and *feeling thermometer*. Judgments concerning *values* and *emotionality* were significant predictors of *true humanness* ratings (see Table 2). In particular, higher ratings for *values* and *emotionality* predicted higher ratings for *true humanness*. Ratings for *ideal humanness*, *purpose*, *typicality*, and *feeling thermometer* were not significant predictors of *true humanness*.

Table 2. Regression analysis predicting *true humanness* from *condition*, *values*, *emotionality, ideal humanness*, *purpose*, and *feeling thermometer* in Experiment 1.

Predictor	*ß*	95% CI
(Intercept)	0.10	[-0.14, 0.34]
Condition (fallible)	0.03	[-0.33, 0.39]
Condition (infallible)	-0.35	[-0.78, 0.07]
Values	0.20*	[0.04, 0.36]
Emotionality	0.25***	[0.13, 0.36]
Ideal humanness	0.10	[-0.06, 0.25]
Purpose	0.05	[-0.10, 0.20]
Feeling thermometer	0.06	[-0.12, 0.24]
Typicality	0.12	[-0.05, 0.29]

Note. Adjusted R^2 = 0.35, 95% CI [0.27, 0.43], F(8, 293) = 21.47, p <.001. ß indicates standardized regression weights.

* indicates p <.05

*** indicates p <.001

As predicted, the morally infallible character was judged as more ideal (M = 5.10, SD = 1.52) than the morally fallible character (M = 3.60, SD = 1.60), t(201.99) = 6.89, p <.001; whereas, the morally fallible character was judged as more of a true human (M = 83.8, SD = 18.2) than the morally infallible character (M = 61.3, SD = 30.4), t(154.6) = -6.32, p <.001. Moreover, participants had slightly more favorable feelings towards the morally infallible character (M = 66.80, SD = 22.9) than they did towards the morally fallible character (M = 59.3, SD = 19.4), t(192.5) = 2.51, p = 0.01.

These findings support the hypothesis 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. They also

support the hypothesis that people have a value-based concept of true humanness, as opposed to a teleological or Platonic concept.

Importantly, though, in addition to *values*, *emotionality* was also a significant predictor of true humanness ratings. One potential explanation of this finding is that people represent both the embodiment of certain moral values, and the ability to experience emotions, as central to true humanness (see Phillips, 2022, pp. 11–16, for some relevant findings). This is consistent with the fact that ratings for *emotionality* were higher for the morally fallible agent ($M = 6.17$, $SD = 1.10$) than the morally infallible agent ($M = 4.75$, $SD = 1.81$), $t(160) = -6.72$, $p < .001$; whereas, ratings for *values* were not significantly different when comparing the fallible agent ($M = 5.13$, $SD = 1.27$) to the infallible agent ($M = 4.86$, $SD = 1.56$), $t(189.34) = -1.37$, $p = 0.17$.

Another potential explanation for why *emotionality* was a significant predictor of true humanness judgments is that people distinguish between embodying (or implementing) essential human values versus having an emotional commitment to them. More specifically, even though participants thought that the morally infallible agent "embodies" essential human values to the same extent as the morally fallible agent, they may have perceived the fallible agent as more of a true human because they thought of him as embodying these values in a specific way: namely, by maintaining a deep, emotionally laden, commitment to them (I elaborate on this hypothesis below in section 6.2).

3. Experiment 2a

Experiment 1 provided evidence that attributions of true humanness are predicted by value-based essentialism, as opposed to Platonic or teleological essentialism. However, people might in fact possess *two* normative concepts of humanness: the value-based concept of a true human; and a Platonic-essentialist concept of the ideal human. If so, this would suggest that normative

dehumanization comes in two varieties. One variety would involve seeing *X* as less humanlike than *Y* so long *X* is represented as embodying (or possessing) essential human values to a lesser degree than *Y*. The second variety would involve seeing *X* as less humanlike than *Y* so long as *X* is represented as being further away from the ideal human than *Y*.

To investigate this hypothesis, I presented participants with a vignette describing two characters side-by-side: a morally infallible character, and a morally fallible one. Participants in one condition were asked to rate a statement asserting that there is a sense in which the infallible character is more "humanlike" than the fallible one; while participants in the other condition were asked to rate a statement asserting that there is a sense in which the fallible character is more "humanlike" than the infallible one. If, in addition to the concept of a *true human*, people also have a distinctively Platonic-essentialist concept of humanness, we should predict that participants will tend to agree with both statements. However, if people only tend to possess the value-based concept of a true human, they should only tend to agree with the statement asserting that there is a sense in which the fallible character is more humanlike than the infallible one.

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 (the aim was at least 100 participants per condition).

Each participant was presented with the following vignette:

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.

The order of information about Michael and William was counterbalanced across participants. Participants were randomly assigned to one of two conditions. Those in 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 in 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 was also asked to rate the extent to which Michael and William are true humans, as well as the extent to which they are ideal humans (these items were identical to those used in Experiment 1). Each participant also completed a comprehension check.

2.2. Results and discussion

The mean true and ideal humanness ratings are shown in Fig. 3. 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. 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.

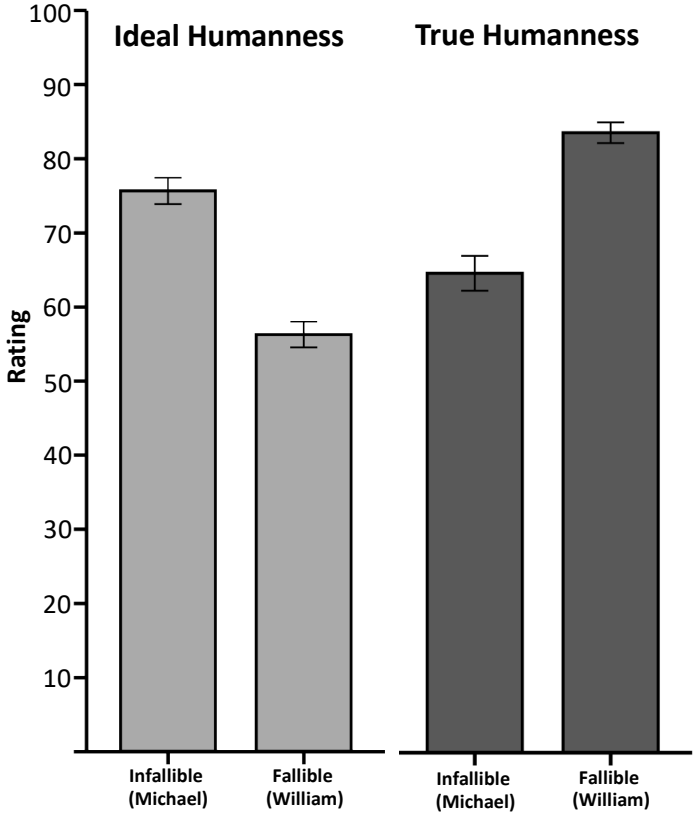


Fig. 3. Mean true humanness and ideal humanness ratings in Experiment 2a (error bars show SE mean).

The mean ratings for each condition (*infallible-more-humanlike* vs *fallible-more-humanlike*) are shown in Figure 4. Participants tended to agree 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); however, they tended to disagree 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.

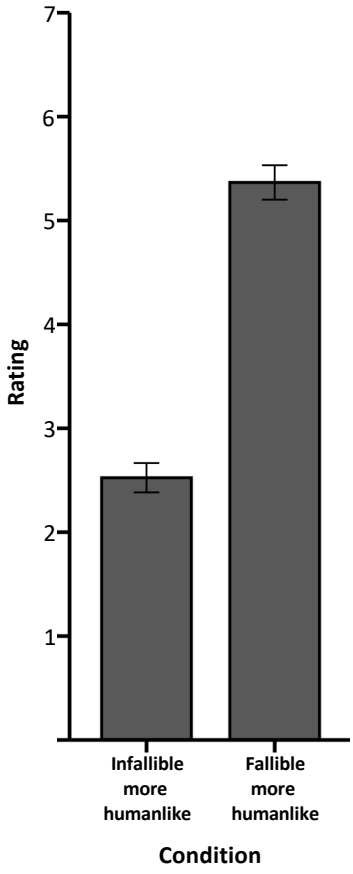


Fig. 4. Mean ratings by condition in Experiment 2a (error bars show SE mean).

To further examine whether participants were deploying a distinctively Platonic-essentialist concept of humanness (in addition to their value-based concept of true humanness), I ran two linear regression models. In the first model, the dependent variable was the judgment as to whether there is a sense in which the fallible agent is more humanlike than the infallible agent; while in the second model, the dependent variable was the judgment as to whether there is a sense in which the infallible agent is more humanlike than the fallible agent. In both models, the predictors were participants' *ideal humanness* and *true humanness scores*. Each participant's *ideal humanness score* was calculated by subtracting their ideal humanness rating for the infallible agent (Michael) from their ideal humanness rating for the fallible agent (William). This

meant that a positive *ideal humanness score* indicated that the participant regarded the infallible agent, Michael, as more of an ideal human than the fallible agent, William; whereas, a negative score indicated the opposite. Each participant's *true humanness score* was calculated in the same way. This meant that a positive *true humanness score* indicated that the participant regarded the infallible agent, Michael, as more of a true human than the fallible agent, William; whereas, a negative score indicated the opposite.

The results of each regression are displayed in Table 3. Lower true humanness scores significantly predicted higher agreement with the statement that there is a sense in which the fallible agent is more humanlike than the infallible agent. Given the way that participants' true humanness scores were calculated, this means that the more likely participants were to see the fallible agent as more of a true human than the infallible agent, the more likely they were to agree that there is a sense in which the fallible agent is more humanlike than the infallible agent.

Both higher true humanness scores, and higher ideal humanness scores, significantly predicted higher agreement with the statement that there is a sense in which the infallible agent is more humanlike than the fallible agent. Given the way that participants' true and ideal humanness scores were calculated, this means that the more likely participants were to see the infallible agent as more of a true, or ideal, human than the fallible agent, the more likely they were to agree that there is a sense in which the infallible agent is more humanlike than the fallible agent. However, the mean rating for this statement was significantly below the midpoint ($M = 2.52$, $SD = 1.43$), $t(102) = -10.45$, $p < .001$. In contrast, the mean rating for the statement asserting that the fallible agent is more humanlike the infallible agent was significantly above the midpoint ($M = 5.37$, $SD = 1.67$), $t(100) = 8.24$, $p < .001$ (Figure 4).

Table 3. Regression models predicting statement-ratings in Experiment 2a.

	Fallible agent more humanlike[a]		Infallible agent more humanlike[b]	
Predictor	*ß*	95% CI	*ß*	95% CI
True humanness score	-0.48**	[-0.67, -0.30]	0.24**	[0.06, 0.42]
Ideal humanness score	-0.09	[-0.28, 0.09]	0.35***	[0.17, 0.53]

Note. ß indicates standardized regression weights.

[a]Adjusted R^2 = 0.26, 95% CI [0.12, 0.40], F(2, 97) = 18.42, p <.001

[b]Adjusted R^2 = 0.22, 95% CI [0.08, 0.36], F(2, 98) = 15.42, p <.001

* indicates p <.05

** indicates <.01

*** indicates p <.001

Together, these findings fail to support the hypothesis that, in addition to the value-based concept of true humanness, people also deploy a distinctively Platonic-essentialist concept of humanness. At best, these findings provide evidence for the following, weaker, hypotheses: (1) the more likely someone is to regard true humans as morally infallible (as opposed to morally fallible), the more likely they are to deploy a distinctively Platonic-essentialist concept of humanness; and (2) the more likely someone is to regard the ideal human as morally infallible (as opposed to morally fallible), the more likely they are to deploy a distinctively Platonic-essentialist concept of humanness.

3. Experiment 2b

The purpose of Experiment 2b was to further examine the hypothesis that people do not possess a distinctively Platonic-essentialist concept of humanness (in addition to their value-based concept of true 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 Experiment 2a, 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, as Experiments 1 and 2a suggest, people have a non-Platonic concept of true humanness, they should agree that there is a sense in which the true human is more humanlike than the ideal human. And if, as Experiment 2a suggests, people tend not to possess a distinctively Platonic-essentialist concept of humanness, they should disagree with the statement that there is a sense in which the ideal human is more humanlike than the true human.

3.1. Materials and methods

The experiment was pre-registered through OSF at https://osf.io/3szb6. One hundred and fifteen participants (58 female, 57 male; M_{age} = 27.0 years) were 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. In Experiment 2a, *statement* (*fallible-more-humanlike* vs *infallible-more-humanlike*) had a large effect on ratings (d = 1.83). For Experiment 2b, 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.

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. 5. Consistent with the findings from Experiment 2a, participants tended to agree 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); however, they tended to disagree 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), $t(93.03) = -12.85$, $p < .001$.

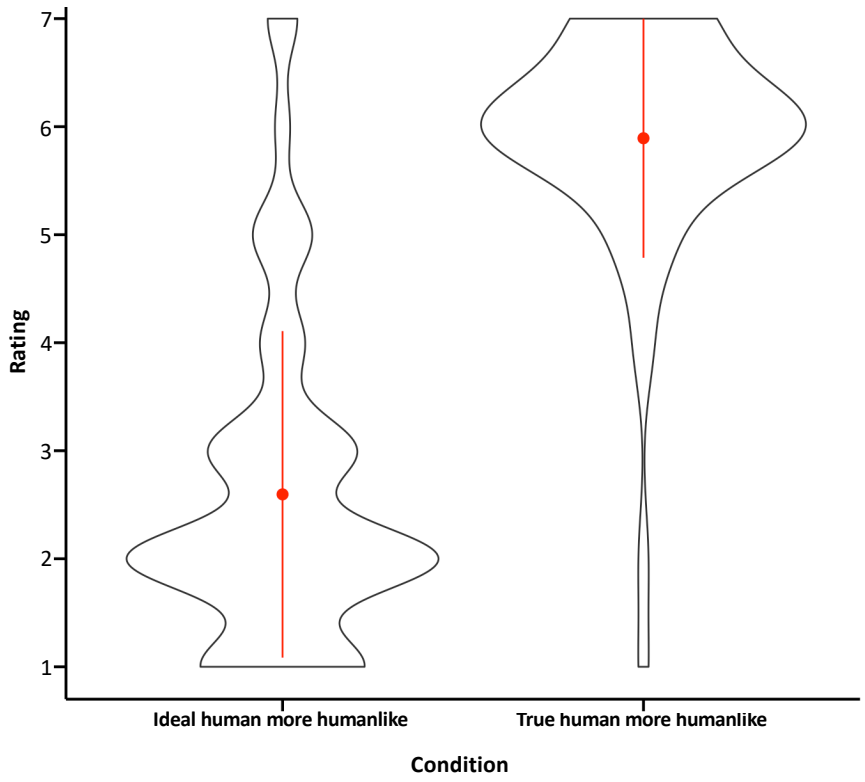


Fig. 5. Violin plots showing the distribution of ratings for each statement. Each plot also displays the mean ratings (red dots), and the standard deviations (red vertical lines).

These findings converge with those of Experiment 2a. Once again, they suggest that people tend not to deploy a distinctively Platonic-essentialist concept of humanness, over and above the value-based concept of true humanness. That is to say, these findings suggest that while people do perceive some individuals as more ideal humans than others, they do not see the former as more humanlike than the latter just in virtue of the fact that they are more ideal humans. In contrast, these findings suggest that when people see person *A* as more of a true human than person *B*, they thereby perceive *A* as more humanlike than *B*.

5. Experiment 3

Experiments 1, 2a, and 2b suggest that normative dehumanization has the following key features:

(1) It is grounded in the ordinary concept of true humanness.

(2) It is distinct from dislike; judgments concerning ideal humanness; and judgments concerning typical humanness.

(3) It is predicted by value-based essentialism, but not by Platonic or teleological essentialism.

(4) People tend not to deploy a distinctively Platonic-essentialist concept of humanness (in addition to their value-based concept of true humanness).

However, even if people tend not to deploy a distinctively Platonic-essentialist concept of humanness, it remains highly likely that judgments concerning ideal humanness predict intergroup hostility. Thus, to show that normative dehumanization is an important predictor of intergroup hostility, it will be important to control for various other sorts of judgments, including denials of ideal humanness.

As a first step towards examining this issue, I presented participants with various social groups. Participants in the *high morality* condition were presented with groups that people tend to rate as highly moral (e.g. firefighters); while those in the *low morality* condition were presented with groups that people tend to regard as relatively immoral (e.g. billionaires). Participants were asked to rate the average member of each group in terms of true humanness; ideal humanness; like/dislike; and human typicality. They were also asked to rate whether the average member of each group deserves support or opposition of various kinds. I predicted that

denials of true humanness would predict opposition over support, even when controlling for denials of ideal humanness; dislike; and perceptions of atypical humanness.

5.1. Materials and methods

The experiment was pre-registered through OSF at https://osf.io/zb9ed. Two hundred and twenty participants (110 female, 110 male; M_{age} = 32.2 years) were recruited from Prolific in exchange for $0.60. Twenty participants were excluded from the final analysis because they failed the attention check. This brought the final sample size down to 200 (the aim was 100 participants per condition).

Participants were randomly assigned to either the *high morality* condition or the *low morality* condition. Those in the *high morality* condition were presented with the following six groups, which were all rated as morally good by participants in a pilot study: nurses; doctors; grade school teachers; aid workers; firefighters; and veterinarians. Participants in the *low morality* condition were presented with the following six groups, which were all rated as relatively immoral by participants in the pilot study: telemarketers; billionaires; lawyers; journalists; celebrities; and Americans.[1]

Each participant was 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 Experiment 1. 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 = Ideal human); *typicality* (0 = Atypical human, 100 = Typical human); and *feeling thermometer* (0

[1] In the pilot study, I also included social groups that people tend to see as extremely immoral: serial killers; psychopaths; pedophiles; White supremacists; and school shooters. However, participants gave the members of these groups very low ratings for true and ideal humanness, which led to floor effects. For this reason, these sorts of groups were not included in Experiment 3.

= Very cold, 100 = Very warm). The item measuring *support/opposition* read as follows (0 = Deserves opposition, 100 = Deserves support):

> Using the sliders below, indicate the extent to which each group deserves support versus opposition. Support can include things like helping the group financially or emotionally, and including them in our social circle. Opposition can include things like denying the group financial or emotional help, excluding them from our social circle, and punishing them.

As a manipulation check, participants were also asked, "How morally good or morally bad is the average member of each group?" (0 = Very morally bad, 100 = Very morally good). Participants also completed an attention check.

5.2. Results and discussion

Results from the manipulation check suggested that judgments concerning moral character were successfully manipulated. Participants in the *high morality* condition tended to rate the average member of the target groups as morally good (M = 79.0, SD = 18.9); while those in the *low morality* condition tended to rate them as somewhat immoral (45.0, SD = 20.6), $t(1177.5) = 29.66$, $p < .001$.

Table 4 displays the zero-order correlations. The mean ratings for each item (by condition) are displayed in Fig. 6.

Table 4. Zero-order correlations from Experiment 3

Correlations	1	2	3	4	5
1. True Humanness		.53***	.48***	.55***	.56***
2. Ideal Humanness			.46***	.75***	.70***

3. Typical humanness	.48***	.52***
4. Feeling thermometer		.75***
5. Support/opposition		

***p <.001

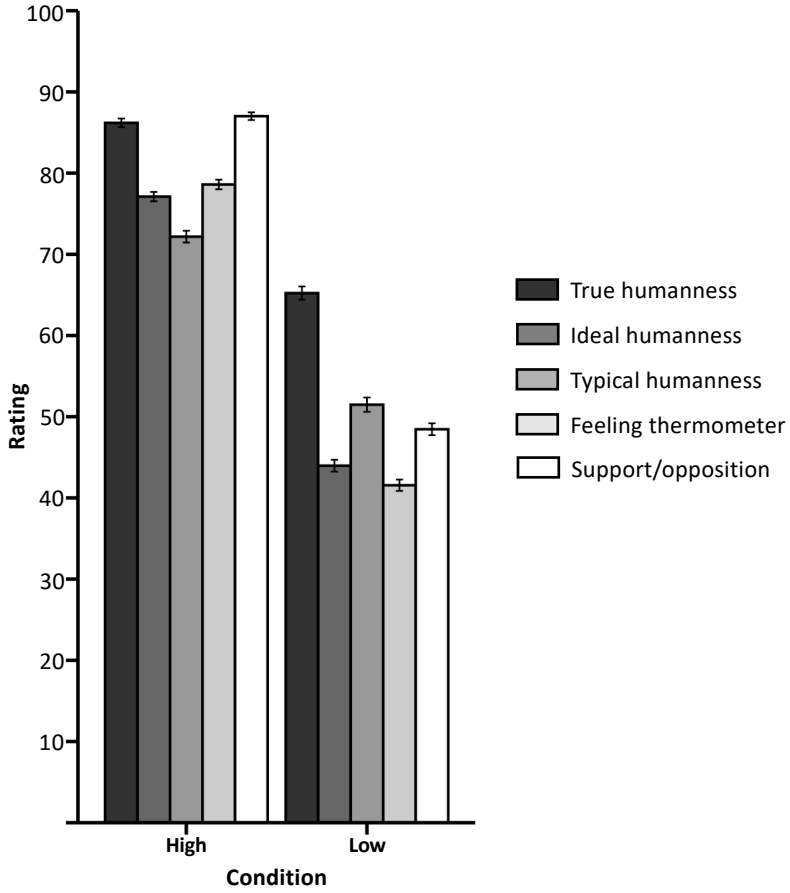


Fig. 6. Mean ratings (by condition) in Experiment 3 (error bars show SE mean).

To examine which sorts of judgments predict opposition (over support), a linear mixed effects model was conducted in R, with the package *lme4* (Bates et al., 2015). *Condition* (dummy coded, with *high morality* as the reference group); *true humanness*; *ideal humanness*; *typical*

humanness; and *feeling thermometer* were specified as fixed effects; while *participant* and *group* were specified as random effects. The model was expressed in *lme4* syntax as follows:

lmer(support_opposition ~ Condition + true_humanness + ideal_humanness + typicality + feeling_thermometer + (1|Group) + (1|Participant))

The model revealed that judgments concerning *true humanness*, *ideal humanness, typical humanness*, and *feeling thermometer* were all significant predictors of *support/opposition*, as was *condition* (see Table 5). As expected, participants in the *high morality* condition were more likely to express support for members of the target groups (M = 87.0, SD = 16.5); whereas those in the *low morality* condition were more likely to express opposition (M = 48.5, SD = 25.3). Lower ratings for true, ideal, and typical humanness all predicted opposition (over support); as did lower ratings on the feeling thermometer.

Table 5. Linear mixed effects model predicting *support/opposition* from *condition*, *true humanness*, *ideal humanness*, *typicality*, and *feeling thermometer* in Experiment 3.

Predictor	*ß*	95% CI
(Intercept)	0.28**	[0.12, 0.45]
Condition (low morality)	-0.57***	[-0.81, -0.33]
True humanness	0.12***	[0.08, 0.17]
Ideal humanness	0.20***	[0.15, 0.25]
Typical humanness	0.13***	[0.09, 0.17]
Feeling thermometer	0.26***	[0.21, 0.31]

Note. Marginal $R^2 = 0.67$ (Conditional $R^2 = 0.83$). *ß* indicates standardized regression weights.

* indicates p <.05

** indicates p <.01

** indicates p <.001

These findings provide evidence that normative dehumanization is a unique predictor of intergroup hostility, over and above denials of *ideal* humanness; denials of *typical* humanness; and dislike. Together with the results of the previous experiments, this supports the general hypothesis that when outgroup members are perceived as immoral, they are perceived as less than fully "human," in the normative sense of the term; and are more vulnerable to harm as a result.

6. General Discussion

The findings reported here support three main hypotheses: (1) People have a value-based concept of true humanness; (2) denials of true humanness are distinct from mere dislike, as well as denials of ideal and typical humanness; and (3) denials of true humanness are an important predictor of intergroup hostility. Experiment 1 provided evidence that denials of true humanness are predicted by value-based essentialism, as opposed to Platonic or teleological essentialism. Experiment 1 also provided evidence that denials of true humanness are distinct from expressions of dislike, as well as judgments concerning ideal and typical humanness. Experiments 2a and 2b provided evidence that people tend not to deploy a Platonic-essentialist concept of humanness, over and above the value-based concept of true humanness. Finally, the findings from Experiment 3 suggest that judgments concerning true humanness are a unique predictor of intergroup hostility, over and above judgments concerning ideal humanness; judgments concerning typical humanness; and dislike.

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). In particular, 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 unfavorable 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, Phillips (2022) found that people will not deny that someone is a true human just because they perceive them as engaging 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. This suggest 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" commitment to these values, even if one fails to implement them from time to time. Similarly, Experiment 1 provides evidence that people view this "deep" commitment as an emotionally laden one. Together, then, these findings suggest that not all acts of moral condemnation count as normative dehumanization. Instead, normative dehumanization appears to be limited to those forms of condemnation in which the target is perceived as lacking a deep, emotionally laden, commitment to moral values.

One possibility that the experiments 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.

Experiments 2a and 2b provide indirect evidence against this sort of view. In both experiments, 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 indirect 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. True humanness and perceptions of emotionality

Another issue that warrants further attention is the role that attributions of *emotionality* play in driving normative dehumanization. In Experiment 1, the more likely participants were to see the target as having emotional capacities, the more likely they were to agree that the target is a true

human. As was outlined above, there are two potential ways to explain this finding. According to one explanation, people represent both the embodiment of certain values and the ability to experience emotions as central to true humanness. According to the second explanation, people distinguish between "embodying" (or "displaying") essential human values versus having an emotional commitment to them, where the latter is regarded as more central to true humanness than the former. If so, participants may have rated 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 essential human values. In other words, they may have thought that the infallible character is less of a true human because he implements essential human values in a relatively "robotic" and emotionless way.

To examine this second hypothesis more directly, future studies could include items that distinguish between embodying (or displaying) essential human values versus having an emotional commitment to them. One such item could ask participants whether the target "displays essential human values," while another item could ask participants whether the target "feels committed to essential human values."

This distinction between embodying certain values and being emotionally committed to them may generalize to other dual character concepts. For instance, consider two working artists, Amber and Kenna. Both are equally resistant to "selling out" and pursuing inauthentic work. However, Amber is a much more successful artist than Kenna. What's more, Kenna has to struggle a great deal to express her ideas in a way that feels authentic; whereas, Amber expresses her ideas in a way that is relatively effortless. Intuitively, both are "true artists." However, people may perceive Kenna as more of a true artist than Amber, because embodying the values of a true

artist requires Kenna to put in much more effort, which, in turn, requires more of an emotional commitment to artistic values on her part.

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

The findings from all four experiments suggest that people do not have a Platonic concept of true humanness. To illustrate, participants in Experiment 1 tended to view the morally infallible character as more of an ideal human than the morally fallible character; however, they viewed the latter as more of a true human than the former.

There are reasons to think that dual character concepts are non-Platonic *in general*. For instance, suppose Marie and Jane are equally committed to implementing the scientific method in a rigorous manner, and are therefore perceived as "true scientists." However, suppose that Marie is much more successful than Jane in uncovering novel phenomena. People may agree that Marie is more of an *ideal scientist* than Jane, but they may not think that Marie is more of a *true scientist* than Jane, given that they are equally committed to scientific values.

Experiments 2a and 2b also suggest that people do not have a distinctively Platonic concept of humanness *over and above* their value-based concept of true humanness. When presented with a character, Michael, who was portrayed as more of an ideal human than another character, William, but as less of a true human, participants tended to agree that there is a sense in which William is more humanlike than Michael; however, they tended to disagree that there is a sense in which Michael is more humanlike than William.

It is worth emphasizing that these findings do not entail that people lack the concept of the *ideal human*, nor do they imply that beliefs about ideal humanness do not have a key role to play in driving hostility towards outgroup members. What these findings *do* suggest is that people tend not to deploy a concept of humanness that meets both of the following criteria: (i) this

concept is associated with an ideal (i.e. the ideal human); and (ii) membership in the relevant category is seen as graded in the sense that if *X* resembles this ideal to a greater extent than *Y*, then *X* is represented as more humanlike than *Y*. To put it in simpler terms, people clearly have a concept of the ideal human, but they do not appear to perceive certain individuals as more humanlike than others just in virtue of their relative proximity to this ideal.

This may have important implications for other measures of 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 (for further discussion, 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 human: instead, they may just be expressing the belief that the target is not an *ideal* human. Nonetheless, when the target is perceived as highly immoral, participants might not just be expressing the belief that this person is a *less-than-ideal* human: they may also be expressing the belief that this person is not a true human.

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 tend to regard these sorts of traits as reflecting the *ideal* human. If so, when people deny that someone possess 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 human. 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. Intergroup hostility

The findings from Experiment 3 provide evidence that normative dehumanization predicts intergroup hostility, over and above judgments concerning ideal humanness; judgments concerning typical humanness; and like/dislike. Nonetheless, Experiment 3 has various limitations that could be addressed in future research.

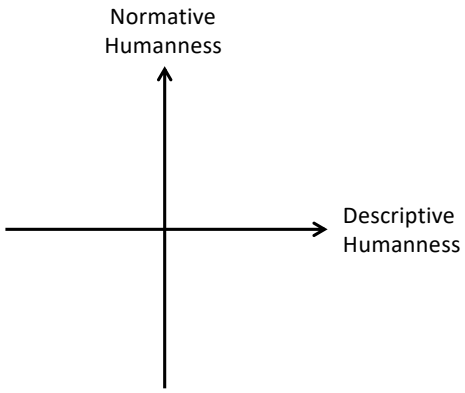
First, it is worth noting that in Experiment 1, the morally infallible agent was perceived as less of a true human than the morally fallible agent; however, participants had slightly more favorable feelings towards the infallible agent. This suggests that in certain cases, normative dehumanization is *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).

Second, in Experiment 3, participants' beliefs about the targets' biological (i.e. descriptive) humanity were not measured directly. Judgments concerning "typical humanness" were elicited, though, and these judgments were a significant predictor of hostility. People may regard the "typical human" as someone who is human in the descriptive (biological) sense. However, this hypothesis needs to be examined more directly in future research.

Third, Experiment 3 did not utilize a measure of *pure* hostility: instead, it utilized a measure of support vs opposition. Moreover, this measure lumped together different forms of hostility,

such as punishment and social exclusion. It is likely that descriptive and normative dehumanization give rise to distinct patterns of hostility. For example, they might fuel different forms of punishment (e.g. retributivist versus restorative); different forms of violence (e.g. 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).

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



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 (see Phillips, 2022, p. 18; 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 people have a concept of true humanness that is grounded in value-based essentialism. I also found evidence that denials of true humanness are a unique predictor of intergroup hostility; over and above dislike, as well as judgments concerning ideal and typical humanness. Future work on dehumanization could benefit from focusing on the distinction

between descriptive and normative dehumanization. As was suggested 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.

References

Anonymous (2015, October 22). *There are too many subhumans that don't care about other people let alone an animal. A true human being cares*. [Comment]. Facebook. https://www.facebook.com/AnimalEquality/photos/the-question-is-not-can-they-reason-nor-can-they-talk-but-can-they-suffer-jeremy/10153226140144077/

Anonymous [@Mr_F_Jordan]. (2016, April 3). *@AmnestyIreland @MadameRamotswe @gov access to abortion is not a human right. Real humans don't kill humans.* [Tweet]. Twitter. https://mobile.twitter.com/amnestyireland/status/716623271723212801

Bailey, A. H., Knobe, J., & Newman, G. E. (2021). Value-based essentialism: Essentialist beliefs about social groups with shared values. *Journal of Experimental Psychology: General, 150*(10), 1994–2014. https://doi.org/10.1037/xge0000822

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. (1996). Intention, history, and artifact concepts. *Cognition*, 60(1), 1–29.

Bloom, P. (2007). Religion is natural. *Developmental Science*, 10, 147–151.

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.

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

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.

Kelemen, D. (1999). Why are rocks pointy? Children's preference for teleological explanations of the natural world. *Developmental Psychology*, 35, 1440–1452.

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition*, 111, 138–143.

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

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

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

Neufeld, E. (2021). Against teleological essentialism. *Cognitive Science*, 45(4), e12961. https://doi.org/10.1111/cogs.12961

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

Piaget, J. (2017). *The child's conception of physical causality*. New York: Routledge.

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.

Rose, D., & Nichols, S. (2020). Teleological essentialism: Generalized. *Cognitive Science*, 44(3), e12818.

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.

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

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 evalutation from trait-based dehumanization. *Cognition*, 231, 105329. DOI: 10.1016/j.cognition.2022.105329