

‘They’re Not True Humans’: Beliefs About Moral Character Drive Denials of Humanity

Forthcoming in *Cognitive Science*

Ben Phillips

Arizona State University

bspilli@asu.edu

Abstract. A puzzling feature of paradigmatic cases of dehumanization is that the perpetrators often attribute uniquely human traits to their victims. This has become known as the “paradox of dehumanization.” We address the paradox by arguing that the perpetrators think of their victims as human in one sense, while denying that they are human in another sense. We do so by providing evidence that people harbor a dual character concept of humanity. Research has found that dual character concepts have two independent sets of criteria for their application, one of which is descriptive and one of which is normative. Across four experiments, we found evidence that people deploy a descriptive criterion according to which being human is a matter of being a *Homo sapiens*; as well as a normative criterion according to which being human is a matter of possessing a deep-seated commitment to do the morally right thing. Importantly, we found that people are willing to affirm that someone is human in the descriptive sense, while denying that they are human in the normative sense, and vice versa. In addition to providing a solution to the paradox of dehumanization, these findings suggest that perceptions of moral character have a central role to play in driving dehumanization.

* Pre-registrations, data, and code can be found here: <https://osf.io/updtg/>

1. Introduction

In 1993, a mob attacked the Roma people living in the small village of Hadareni, Romania.

Thirteen houses were torched, and three Roma were killed. In an interview, one member of the mob, Maria, said that “it would have been better if we had burnt more of the people, not just the houses.” She then insisted that she was not guilty of murder because “Gypsies are not really people, you see” (Bridge, 1993). As Smith (2020, p. 146) points out, one thing that is puzzling about these remarks is that Maria initially referred to her victims as “people,” before going on to characterize them as “not really people.” In explaining why she thought of her victims in this way, Maria provided a moralistic justification, claiming that the Roma are “criminals” who are “always killing each other.”

Similar examples can be found in the writings of Nazi propagandists. On some occasions, Goebbels explicitly denied that Jewish people are human. For example, during a speech to the Nazi Party Congress in 1935, he described Bolshevism as “the declaration of war by Jewish-led international subhumans against culture itself” (United States Holocaust Memorial Museum, n.d.). On other occasions, though, Goebbels explicitly categorized Jewish people as “human.” For example, he is quoted as saying, “Of course, the Jew is a human being too. None of us ever doubted it” (Eitzen, 1936).

These apparently inconsistent categorizations reflect a common theme in paradigmatic cases of dehumanization. On the one hand, the perpetrators deny that their victims are real humans; but on the other hand, they sometimes refer to them as “humans,” and frequently treat them as (bad) moral agents, with an array of traits that we do not normally attribute to nonhuman animals. For example, the Nazis frequently characterized their victims as criminals, traitors, and conspirators. These are not the sorts of traits that we think of nonhuman animals as possessing. This puzzling

feature of paradigmatic cases has become known as “the paradox of dehumanization” (see Smith, 2016).

There are three main ways to resolve the paradox. According to one approach, the perpetrators in these paradigmatic cases do, in fact, recognize the humanity of their victims: when they explicitly refer to them as “subhuman,” they are merely engaging in rhetoric (e.g., see Appiah, 2008, p. 144; Bloom, 2017; Kuper, 1981, pp. 86, 92; Lang, 2010; Manne, 2016, 2018, chapter 5; Over, 2020; and for some relevant empirical work, see Enock et al., 2021; Rai et al., 2017). An alternative approach is to claim that the perpetrators harbor inconsistent representations of their victims, simultaneously representing them as both human and subhuman (Smith, 2016, 2020). The third approach is to claim that the perpetrators are willing to affirm the humanity of their victims, while at the same time denying it, because they are invoking distinct senses of the term “humanity” (de Ruiter, 2021; Kronfeldner, draft; Steizinger, 2018; see also Kronfeldner, 2018, Chapters 2 & 10). In what follows, we defend a version of this third approach by showing that people tend to associate distinct sets of application conditions with the term “humanity”: one set concerns membership in the species *Homo sapiens*, while the other set concerns normative criteria.

1.1. Current models and measures of dehumanization

Influential models of dehumanization tend not to address the possibility that people harbor a concept of humanity with more than one set of application conditions. For example, according to Haslam and colleagues’ model, there are two modes of dehumanization (Haslam, 2006; Haslam et al., 2005). *Animalistic dehumanization* occurs when we represent someone as possessing uniquely human traits to a lesser extent than certain other people (e.g. one’s fellow ingroup members). *Mechanistic dehumanization* occurs when we represent someone as possessing

human-typical traits to a lesser extent than certain other people. Importantly, Haslam and colleagues' findings do not show that when we engage in one form of dehumanization, but not the other, we are representing the target as human in one sense, while denying that they are human in another sense. For example, suppose I perceive the members of some outgroup as possessing uniquely human traits to a lesser extent than the members of my ingroup; whereas, I do not perceive them as possessing human-*typical* traits to a lesser extent than the members of my ingroup. In other words, suppose I engage in animalistic, but not mechanistic, dehumanization. It does not follow that I am deploying a concept of humanity with two distinct sets of application conditions (for a relevant discussion, see Smith, 2014, pp. 817–820). Rather, Haslam and colleagues' model is compatible with the claim that I am deploying a concept of humanity with a *single* set of application conditions: one that happens to invoke both human-typical and uniquely human traits.

Similarly, consider Kteily and colleagues' (2015) measure of “blatant dehumanization.” Participants are shown the Ascent of Man image, which depicts a series of ape-like creatures through to a modern human. They are then asked to rate the members of various social groups in terms of how “human-like” and “evolved” they seem. In utilizing the Ascent scale, participants may be deploying a concept of humanity with a unique set of application conditions. However, it is also possible that they are deploying two sets of application conditions: one involving the species sense; and one involving the normative sense. By itself, the Ascent measure does not provide us with the sort of evidence that could adjudicate between these two hypotheses.

Finally, consider Smith's (2011, 2014, 2016, 2020) influential theory of “subhumanization” according to which categorical denials of humanity occur when the perpetrator thinks of his victims as lacking “the human essence”: a property that the dehumanizer regards as immutable,

inherited, and inductively rich. Smith's theory is compatible with the claim that dehumanizers deploy more than one set of criteria for being human. However, his theory does not entail this claim. In fact, as was mentioned above, Smith adopts the alternative approach according to which subhumanizers harbor inconsistent representations of their victims' humanity (Smith, 2016).

1.2. HUMANITY as a dual character concept

Why think that people harbor distinct sets of criteria for being human, and what sorts of features might these criteria invoke? To answer these questions, consider dual character concepts. In a series of experiments, Knobe, Prasada, and Newman (2013) found that for various social concepts, people tend to deploy two distinct sets of criteria for their application, one of which is always normative. To illustrate, consider one of their examples. George is employed at Ameritech to run experiments and analyze the data. However, he does not care at all about finding answers to the questions he studies: instead, he just goes through the motions, carrying out his duties to get a paycheck. Knobe and colleagues found that people tend to agree with the following claim, which they refer to as the *ultimately non-member* statement:

There's a sense in which George is clearly a scientist, but ultimately, if you think about what it really means to be a scientist, you'd have to say that there is a sense in which he is not a scientist at all.

Now consider another individual, Gerry, who has never received formal scientific training. Nonetheless, Gerry is deeply committed to systematically revising his beliefs in light of empirical evidence. Knobe and colleagues found that people tend to agree with the following claim, which they refer to as the *ultimately member* statement:

There's a sense in which Gerry is clearly not a scientist, but ultimately, if you think about what it really means to be a scientist, you'd have to say that there is a sense in which he is a true scientist after all.

Knobe and colleagues tested participants on various other candidates for dual character concepts, such as ARTIST, TEACHER, MUSICIAN, and FRIEND. They also tested people on various control concepts (e.g. DOORMAN, MECHANIC), as well as various natural-kind concepts (e.g. TIGER, ZEBRA). Participants were much more likely to endorse both the *ultimately non-member* statement and the *ultimately member* statement when presented with one of the dual character concepts. As Knobe and colleagues point out, in simultaneously endorsing both statements, participants appear to be deploying two distinct sets of criteria for the given concept's application. One set of criteria involves descriptive features; whereas, the other set appears to involve normative features. For example, in the case of SCIENTIST, the descriptive criteria involve working in a lab, running experiments, etc., while the normative criteria involve having a desire to systematically revise one's beliefs in light of empirical evidence (for further discussion of the normative dimension, see Del Pinal and Reuter, 2017; Leslie, 2015). In the case of TEACHER, the descriptive criteria involve working in a classroom, instructing students, etc., while the normative criteria involve having a desire to impart knowledge to others.

There are a number of reasons to expect that people harbor a dual character concept of humanity. Anecdotally, people appear to use the phrases "real human" and "true human" in ways that fit the profile of a dual character concept. In a recently published book, entitled, *The True Humanity in Us: 52 Great Acts in a Year* (Laraiyetan, 2021), the author implores readers to "make the world a better place through acts of kindness, caring, generosity, love, and by promoting peace." In 2015, a columnist for the Lakota County Times, Vi Waln, made the

following remark about those responsible for the construction of the Keystone XL pipeline: “A genuine human being will always put Mother Earth first. A real human being doesn’t sell out the entity which gives them life” (Waln, 2015). In one online discussion forum, participants were discussing the question “What do you think is a true human being?” The majority of answers focused on adherence to moral norms, such as respecting others, being altruistic, and having empathy towards others. One commenter answered by noting that human beings are “moral creatures” who “instinctively know that there is such a thing as right and wrong” (Quora, n.d.). Finally, on the occasion of Nelson Mandela’s death, a condolence book was posted on the UN website. Some commented on Mandela’s moral character, characterizing him as a “true human” (United Nations: Nelson Mandela International Day, n.d.). In all these cases, people appear to be deploying normative, not biological, criteria when applying the terms “real human” and “true human.”

A second reason to expect that people possess a dual character concept of humanity comes from research into the “true self.” This research suggests that people tend to think of humans as having a deep-seated commitment to do the morally right thing (see Strohminger et al., 2017). For instance, Newman et al. (2014, 2015) found that when people regard someone as behaving morally, they tend to see the behavior as a product of the agent’s “true self”: the person they are “deep down.” Importantly, they also found that when people regard someone as behaving immorally, they tend to see the behavior as a product of the agent’s “surface self.” Two other key findings are relevant for our purposes. First, Christy et al. (2017) found that people regard true good selves as essence-like, having features such as immutability and inductive value. Second, people do not just attribute true good selves to ingroup members: in fact, they are willing to attribute them to political opponents (Newman et al., 2014), and even the members of

threatening outgroups (De Freitas and Cikara, 2018). Together, these findings suggest that people harbor a conception of humanity according to which all humans have a deep-seated commitment to do the morally right thing. Thus, people may deny that someone is a “true (or real) human” when they perceive them as essentially bad, i.e. as lacking a true good self.

In what follows, we assess the hypothesis that people harbor a dual character concept of humanity. In doing so, we examine whether people deploy a descriptive criterion according to which being human is a matter of being a *Homo sapiens*; as well as a normative criterion according to which being human is a matter of possessing a deep-seated commitment to do the morally right thing. If our hypothesis is correct, there should be cases in which people categorize an individual as human in the species sense, but deny that they are human in the normative sense, and vice versa.

1.3. The present studies

Four pre-registered experiments (<https://osf.io/updtg/>) provided converging evidence for our main hypothesis. Experiment 1 examined whether people are willing to categorize someone as human in the species sense, but deny that they are human in the normative sense, and vice versa. Experiment 2 was designed to show that people will deny that someone is human, in the normative sense, when they regard them as lacking a true good self. Experiments 3 and 4 examined whether the normative criteria that people associate with being a true human concern purely moral features, or, whether certain non-moral features play a role as well.

2. Experiment 1: Humanity as a dual character concept

The aim of the first experiment was to assess whether people recognize two independent sets of criteria for being human: one based on species membership; and the other based on moral features. If people do recognize both sets of criteria, there should be cases in which they are

willing to affirm someone's humanity in the species sense, but deny it in the normative sense; as well as cases in which they are willing to affirm it in the normative sense, but deny it in the species sense.

2.1. Methods

Four hundred and twenty participants (251 females, 169 males; $M_{\text{age}} = 25.0$ years) were recruited from Prolific in exchange for \$0.65. The sample size was based on a target of at least 100 participants per cell. Seven participants were excluded from the final analysis as a result of failing either the comprehension check or the attention check. This brought the final sample size down to 413.

Participants were randomly assigned to one of four vignettes. Each one described a character as either good or evil, and as either a *Homo sapiens* ("Jim") or an alien ("Xanthon"). Participants were then asked to rate two statements, presented in random order, in terms of how weird or natural they sounded (1 = Sounds very weird, 7 = Sounds very natural). The *ultimately member* statement read as follows:

There is a biological sense in which [Jim/Xanthon] is clearly not human, but ultimately, if you think about what it really means to be human, you'd have to say that there is a deeper sense in which [Jim/Xanthon] is a true human after all.

The *ultimately non-member* statement read as follows:

There is a biological sense in which [Jim/Xanthon] is clearly human, but ultimately, if you think about what it really means to be human, you'd have to say that there is a deeper sense in which [Jim/Xanthon] is not a true human after all.¹

We reasoned that if participants harbor a dual character concept of humanity, they should rate the *ultimately non-member* statement, but not the *ultimately member* statement, as natural-sounding when the target is portrayed as an evil *Homo sapiens*. The vignette in question read as follows:

Jim is a member of our species (i.e. a *Homo sapiens*). He is a psychopathic serial killer who, deep down, sees nothing wrong with actions, such as torture, that cause his victims to suffer immensely. He does the best he can to harm others, no matter who they are, or where they come from. In fact, Jim frequently performs random acts of cruelty on innocent people. For example, on one occasion, Jim abducted and killed a young girl simply because he could. He usually tortures his victims. Those who have encountered Jim describe him as evil and bloodthirsty, with a deep-seated commitment to making others suffer.

We also reasoned that if participants harbor a dual character concept of humanity, they should rate the *ultimately member* statement, but not the *ultimately non-member* statement, as natural-sounding when the target is portrayed as an alien who is committed to doing the morally right thing. The vignette in question read as follows:

¹ The qualifier “biological” was included in both statements to forestall an issue concerning people’s interpretation of the phrase “true human.” There is evidence that phrases of the form “X is not a true F” are ambiguous between two readings. According to the reading elaborated on above, this phrase means that X does not satisfy the normative criteria associated with being an F. According to the other reading, it means that X does not belong to the natural kind associated with Fs (e.g., see Tobia et al., 2020). If the folk regard the category *Homo sapiens* as a natural kind, then they may regard the phrase “true human” as ambiguous between the natural-kind reading and the normative reading (for some analogous cases involving race and gender concepts, see Bailey et al., 2021; Guo et al., 2021). To preempt this potential ambiguity, we utilized statements that contrast the claim that the target is/isn’t a “true human” with the claim that the target isn’t/is human in the “biological sense.”

Imagine that Xanthon is a member of an alien species. He has a strong sense of right and wrong, and he is deeply committed to doing the right thing by others. He does the best he can to help others in need, no matter who they are, or where they come from. In fact, Xanthon often performs random acts of kindness in which he sacrifices his own needs for the sake of helping others. For example, on one occasion, Xanthon donated one of his organs to save someone's life. He is also known to go without food to ensure that those around him have something to eat. Those who have met Xanthon describe him as generous and kind, with a deep-seated commitment to his moral principles.

Finally, we predicted that participants would rate both statements as weird-sounding when the target was portrayed as either a good *Homo sapiens* or as an evil alien. The relevant vignettes read as follows:

Jim is a member of our species (i.e. a *Homo sapiens*). He has a strong sense of right and wrong, and he is deeply committed to doing the right thing by others. He does the best he can to help others in need, no matter who they are, or where they come from. In fact, Jim often performs random acts of kindness in which he sacrifices his own needs for the sake of helping others. For example, on one occasion, Jim donated one of his organs to save someone's life. He is also known to go without food to ensure that those around him have something to eat. Those who have met Jim describe him as generous and kind, with a deep-seated commitment to his moral principles.

Imagine that Xanthon is a member of an alien species. He is a psychopathic serial killer who, deep down, sees nothing wrong with actions, such as torture, that cause his victims to suffer immensely. He does the best he can to harm others, no matter who they are, or where they

come from. In fact, Xanthon frequently performs random acts of cruelty on innocent individuals. For example, on one occasion, Xanthon abducted and killed a young female simply because he could. He usually tortures his victims. Those who have encountered Xanthon describe him as evil and bloodthirsty, with a deep-seated commitment to making others suffer.

In addition to rating the two statements, participants also completed a comprehension check and an attention check.

2.2. Results and discussion

A $2 \times 2 \times 2$ mixed ANOVA was performed, with moral character (good, evil) and species (alien, *Homo sapiens*) as between-subjects factors; statement (ultimately member, ultimately non-member) as a within-subjects factor; and participants' ratings as the dependent variable. The mean ratings are shown in Fig. 1 (information about the distribution of responses can be found in Supplementary Material S1). As predicted, there was a significant interaction between moral character and species on participants' ratings ($F(1,409) = 22.72, p < .001, \eta_p^2 = .053$). Analyses of simple main effects revealed that this was due to a statistically significant interaction between moral character and species for both the *ultimately non-member* statement ($F(1,409) = 13.5, p < .001, \eta_p^2 = .032$), and the *ultimately member* statement ($F(1,409) = 9.47, p = 0.002, \eta_p^2 = .023$). As predicted, this was due to higher ratings for the *ultimately non-member* statement when the target was portrayed as an evil *Homo sapiens* ($M = 4.41, SD = 1.63$) as opposed to an evil alien ($M = 3.28, SD = 1.77$), $t(204) = -4.78, p < .001$; and higher ratings for the *ultimately member* statement when the target was portrayed as a good alien ($M = 4.71, SD = 1.59$) as opposed to a

good *Homo sapiens* ($M = 3.43$, $SD = 1.81$), $t(200) = 5.37$, $p < .001$. The three-way interaction was not significant.

These findings support the hypothesis that people harbor a dual character concept of humanity. Participants tended to think that it is relatively natural to deny that an “evil” *Homo sapiens* is a “true human,” even if he counts as “human” in another sense. They also tended to think that it is relatively natural to affirm that a morally good alien is a “true human,” even if the alien does not count as “human” in another sense. This is the pattern that we would expect from a dual character concept, with two distinct sets of application conditions. Participants appeared to be deploying one criterion according to which being human is a matter of being a *Homo sapiens*; as well as an independent criterion according to which being human is a matter of having a commitment to do the morally right thing.²

² These findings converge with those of a follow-up experiment in which we removed the qualifier “biological” from both the *ultimately member* statement and the *ultimately non-member* statement (see Supplementary Material S2). Additionally, before running the experiment reported above, we ran a variant in which both moral and non-moral traits were attributed to targets (see Supplementary Material S3). The results of this experiment also converge with those reported above.

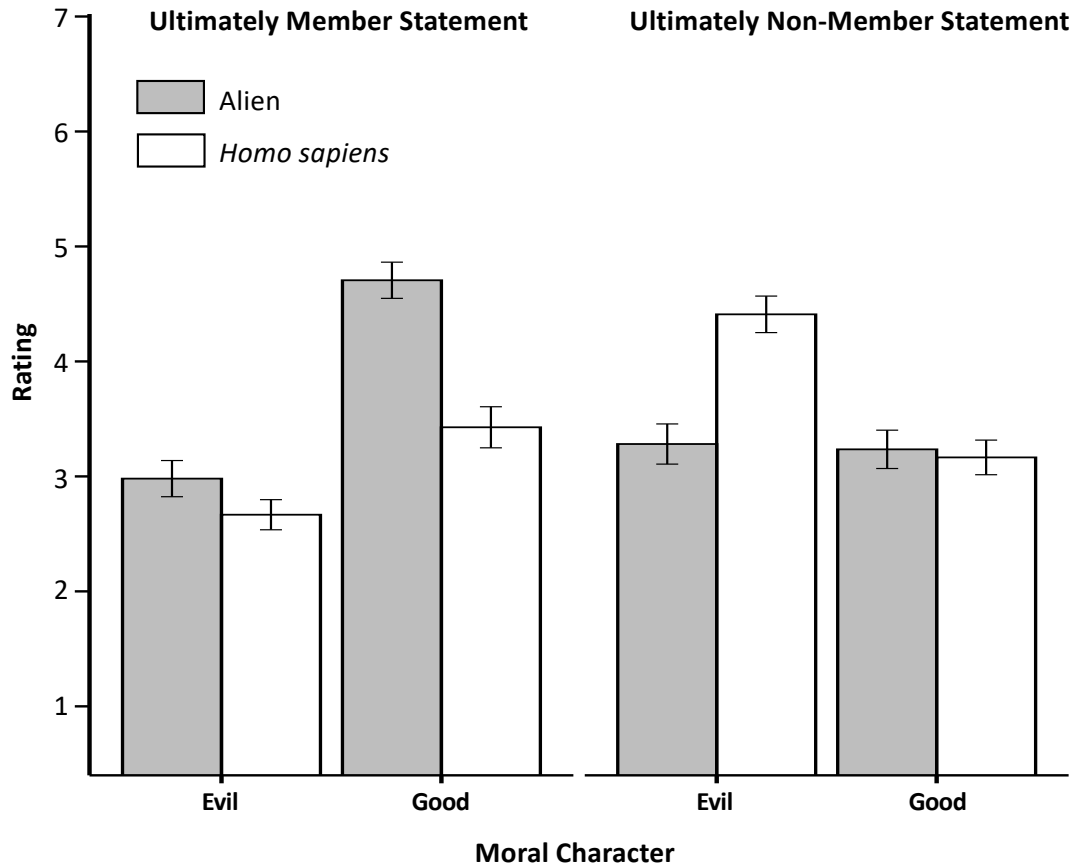


Figure 1. Mean ratings by condition for Experiment 1 (error bars show SE mean). According to the *ultimately member* statement, the individual in question is a “true human,” despite not being human in the biological sense. According to the *ultimately non-member* statement, the individual in question is not a “true human,” despite being human in the biological sense.

3. Experiment 2: The mediating role of beliefs about the true self

The results of Experiment 1 provide evidence that people harbor a dual character concept of humanity. Importantly, they suggest that people are willing to affirm someone’s humanity in the species sense, while denying it in the normative sense. Nonetheless, the conditions under which people are willing to deny someone’s humanity, in the normative sense, remain somewhat unclear. It is possible that people are willing to do so whenever they perceive someone as engaging in highly immoral behavior. However, research into the true self has shown that people

tend to regard others as possessing true *good* selves, even when they perceive them as behaving immorally (Strohming et al., 2017). It is thus doubtful that people will deny someone's humanity just because they see them as behaving in a highly immoral manner. Instead, research into the true self suggests that in most cases, people will regard an immoral individual as categorically human, in the normative sense, because they see them as harboring a deep-seated commitment to do the morally right thing: it is just that they are failing to manifest this aspect of their true self. Denials of humanity are thus more likely to occur when the target individual is seen as lacking a true good self altogether. Experiment 2 was designed to test this hypothesis.

3.1. Methods

Four hundred and twenty participants (176 females, 243 males, 1 did not specify; $M_{\text{age}} = 26.5$ years) were recruited from Prolific in exchange for \$1.25. The target sample size was at least 400 participants. Eight participants were excluded from the final analysis as a result of failing either the comprehension check or the attention check. This brought the final sample size down to 412.

Participants were randomly assigned to one of two conditions. In the *evil* condition, participants read the following vignette:

Jim is a psychopathic serial killer. He is fundamentally evil, and completely lacks empathy for the suffering of his victims. In fact, he very much enjoys killing innocent people. In general, Jim has no concern at all about the rights of others. One evening, Jim shoots and kills a gas station attendant simply because he feels like it.

In the *non-evil* condition, participants read the following vignette:

Jim is a teenager who has never done anything illegal before. In general, Jim has a strong sense of right and wrong. One evening, he is hanging out with some new friends. Some of

them decide to rob a gas station. Jim feels pressured to go along with it, and ends up taking part in the robbery. During the robbery, a gas station attendant lunges at Jim. In a panic, Jim shoots and kills the gas station attendant.

Participants were then asked to rate the extent to which they agree or disagree (1 = Strongly disagree, 7 = Strongly agree) with two statements. First, they rated the *true self* statement, which read as follows:

Jim was drawn towards killing the gas station attendant by his true self—the person he really is deep down.

Participants were then asked to rate an *ultimately non-member* statement, which read as follows:

There is a sense in which Jim is clearly human, but ultimately, if you think about what it really means to be human, you'd have to say that there is a sense in which Jim is not a true human after all.

Participants also completed a comprehension check and an attention check.

We predicted that the effect of the target's motivation (evil, non-evil) on participants' ratings of the *ultimately non-member* statement would be mediated by their beliefs about the target's true self.

3.2. Results and discussion

As Fig. 2 illustrates, participants were more likely to agree that the evil individual is not a true human ($M = 4.57$, $SD = 1.81$) than they were to agree that the non-evil individual is not a true human ($M = 2.84$, $SD = 1.50$), $t(398) = 10.57$, $p < .001$ (the distribution of responses for each condition are presented in Supplementary Material S4).

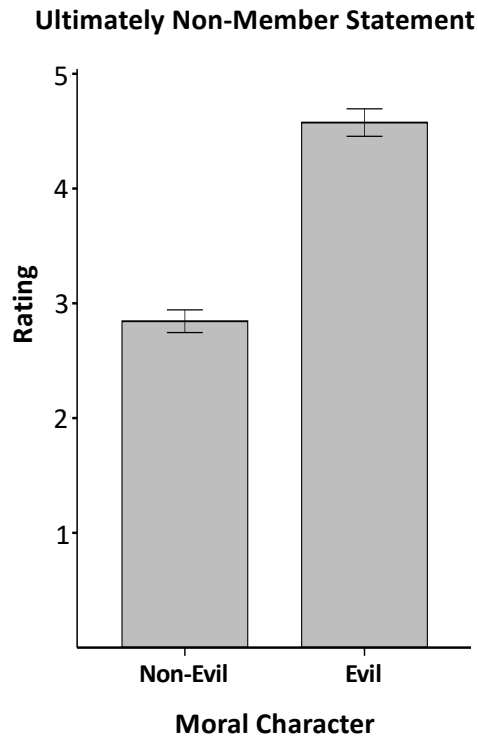


Figure 2. Mean ratings of *ultimately non-member* statement in Experiment 2 (error bars show SE mean).

Participants were also more likely to agree that the actions of the evil individual reflected his true self ($M = 4.90$, $SD = 1.72$) than they were to agree that the actions of the non-evil individual reflected his true self ($M = 2.66$, $SD = 1.41$), $t(397) = -14.43$, $p < .001$ (see Fig. 3).

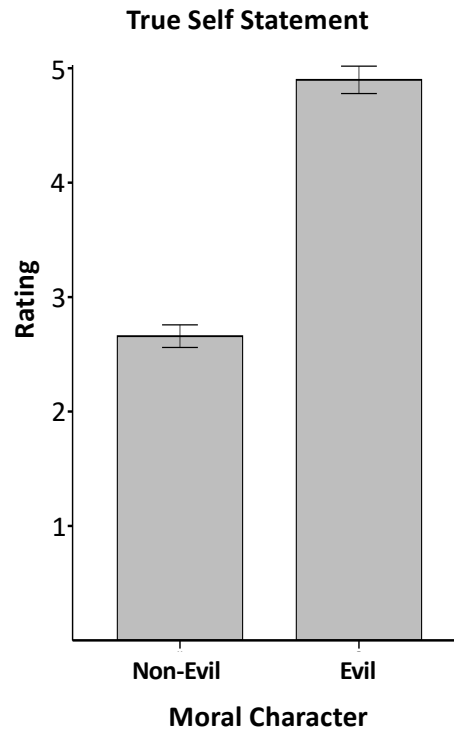


Figure 3. Mean ratings of *true self* statement in Experiment 2 (error bars show SE mean).

As predicted, the effect of the target’s motivation (evil, non-evil) on judgments concerning the target’s humanity was partially mediated by beliefs about the target’s true self (see Fig. 4). The unstandardized regression coefficient between the target’s motivation and beliefs about the target’s true self was statistically significant, as was the coefficient between beliefs about the target’s true self and judgments concerning the target’s humanity. We used bootstrapping procedures to test the significance of this indirect effect. Unstandardized indirect effects were computed for each of 1000 bootstrapped samples. The 95% confidence interval was computed by determining the indirect effects at the 2.5th and 97.5th percentiles. The bootstrapped unstandardized indirect effect was .42 (95% CI = .18 to .70). Thus, the indirect effect was statistically significant ($p < .001$).

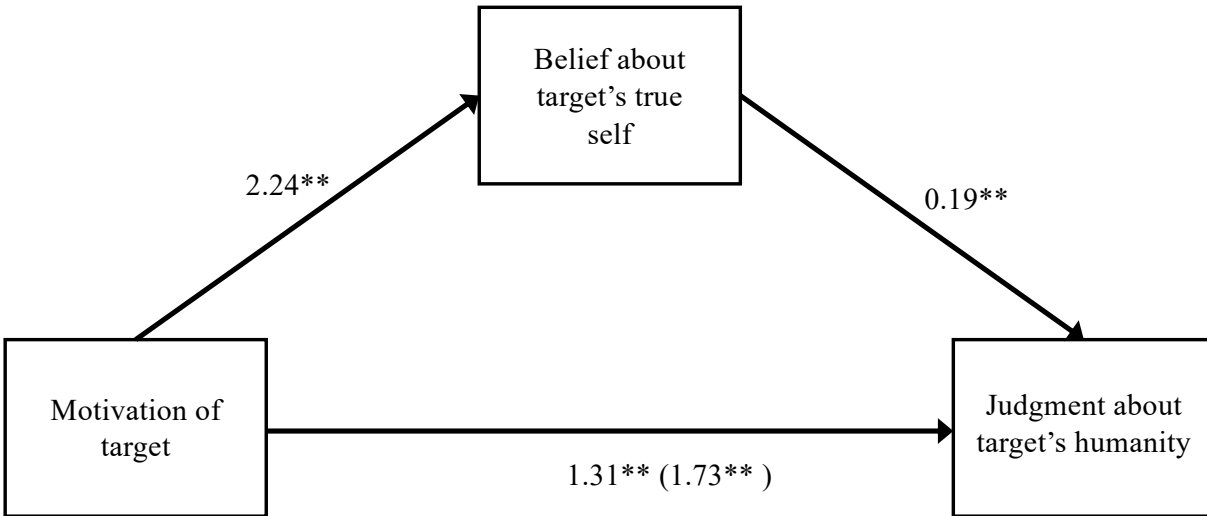


Figure 4. Mediation results from Experiment 2. **p < .001.

4. Experiment 3: True humanity and non-moral traits

Experiments 1 and 2 provide evidence that people harbor a dual character concept of humanity. Moreover, they suggest that people are willing to deny that someone is a true human when they see them as lacking a true good self. However, we did not directly test whether people regard *non-moral* traits as central to being a true human. In Experiment 3, we addressed this issue by presenting participants with characters who lack various non-moral traits. We predicted that participants would rate an *ultimately non-member* statement as relatively natural-sounding when the target was described as lacking a deep-seated commitment to do the morally right thing; whereas, they would rate it as relatively weird-sounding when the target was only described as lacking some non-moral trait.

4.1. Methods

Seven hundred and twenty participants (310 females, 408 males, 2 did not specify; $M_{\text{age}} = 26.1$ years) were recruited from Prolific in exchange for \$0.72. The sample size was based on a target of at least 100 participants per cell. Twenty participants were excluded from the final analysis as a result of failing either the comprehension check or the attention check. This brought the final sample size down to 700.

Participants were randomly assigned to one of seven conditions. In each case, participants read about an individual, Jim, who was described as completely lacking some trait. We included some traits that people tend to rate highly in terms of human uniqueness (see Haslam, 2006). These traits included moral sensibility; intelligence; rationality; open-mindedness; and culturedness. We also included two traits—warmth and emotionality—that people tend to rate highly in terms of human-typicality (see Haslam, 2006). For example, in the *antisocial* condition, a character called Jim was described as “completely reserved and antisocial” with “no desire to be around others” (see Appendix A for the complete set of vignettes). Having read the assigned vignette, participants were asked to rate the same *ultimately non-member* statement that was used in Experiment 2 (1 = Sounds very weird, 7 = Sounds very natural). Participants also completed a comprehension check and an attention check.

4.2. Results and discussion

A one-way ANOVA was conducted to compare the effect of vignette on participants' ratings. The mean ratings for each condition are shown in Fig. 5 (information about the distribution of responses is presented in Supplementary Material S5). As predicted, the effect of vignette on participants' ratings was significant, $F(6, 693) = 11.23, p < .001, \eta_p^2 = .089$. Post hoc comparisons using the Tukey HSD test are shown in Table 1. The difference between the mean ratings for the

evil ($M = 4.03$, $SD = 1.52$) and *emotionless* conditions ($M = 3.83$, $SD = 1.43$) was not significant ($p = 0.97$). However, the difference between the mean ratings for the *evil* condition and each of the other conditions was statistically significant (see Table 1).

These results are consistent with the claim that people regard a commitment to do the morally right thing as relatively central to being human, in the normative sense. Interestingly, though, the mean rating for the *ultimately non-member* statement was only at the midpoint in the *evil* condition ($M = 4.03$, $SD = 1.52$). In Experiments 1 and 2, the mean ratings for this statement were well above the midpoint in the corresponding conditions ($M = 4.41$, $SD = 1.63$ in Experiment 1; and $M = 4.57$, $SD = 1.81$ in Experiment 2). One potential explanation for this difference concerns the degree to which the target's true self was emphasized in each case. In Experiment 1, the evil character was described as possessing a "deep-seated commitment to making others suffer." The vignette also stated that "deep down," this character sees nothing wrong with actions that cause his victims to suffer. In Experiment 2, participants were asked to reflect on the evil character's true self before rating the *ultimately non-member* statement. Moreover, Experiment 2 provided evidence that the judgment that someone is not a true human is mediated by the judgment that they possess a true bad self. In Experiment 3, however, participants were not asked to reflect on the target's true self, nor was the target described in terms of who they are "deep down." Thus, it is possible that participants were less likely to deny that the evil character is a true human in Experiment 3, because they were less likely to reflect on the target's true self (see sect. 6.2 for further discussion of this issue).

Given that the difference between the mean ratings for the *evil* and *emotionless* conditions was not significant, the role of emotionality in driving denials of true humanity needs to be examined in more detail. There are three main possibilities to consider. People may believe that

to be a good moral agent, one must experience certain emotions, such as guilt, shame, and empathy. People may also believe that if an agent lacks any commitment to do the morally right thing, they must not have emotional capacities. Finally, it is possible that the effects of moral character and emotionality on judgments concerning true humanity are independent of one another. In other words, people may regard both good moral character and emotional capacities as central to being human, in the normative sense. Experiment 4 was designed to address this issue.

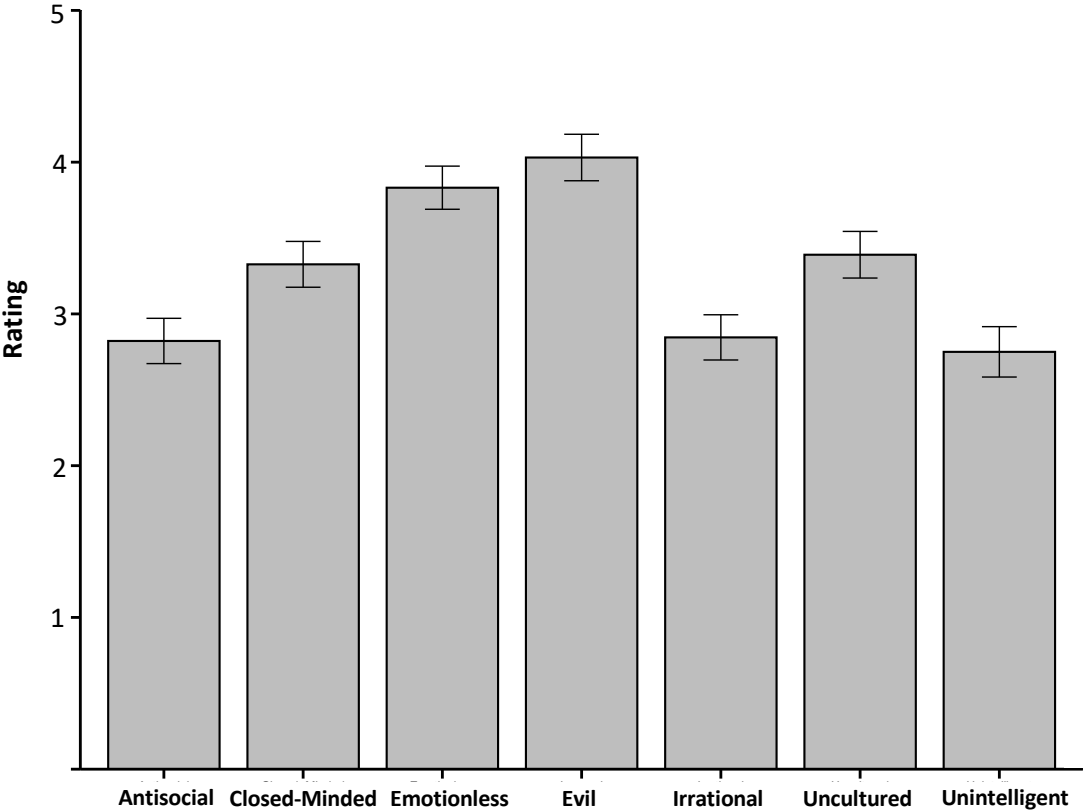


Figure 5. Mean ratings of *ultimately non-member* statement in Experiment 3 (error bars show SE mean).

Table 1. Results of post hoc Tukey HSD test.

	Evil	Emotionless	Uncultured	Closed-Minded	Irrational	Antisocial	Unintelligent
Evil		0.20	0.64*	0.70*	1.18**	1.21**	1.28**
Emotionless			0.44	0.50	0.98**	1.01**	1.08**
Uncultured				0.06	0.54	0.57	0.64*
Closed-Minded					0.48	0.50	0.57
Irrational						0.03	0.09
Antisocial							0.07
Unintelligent							

Note. Differences between means shown. *p <.05, **p <.001. All values are Bonferroni-adjusted.

5. Experiment 4: Moral character versus emotionality

The aim of Experiment 4 was to further examine whether people regard emotionality as central to being a true human. We did so by presenting participants with characters who vary with respect to both emotionality (emotional, emotionless) and moral character (good, evil). We then asked participants to rate an *ultimately non-member* statement about the given character. If people regard emotionality as central to being human, in the normative sense, they will be more likely to rate the statement as natural-sounding when the character is described as emotionless, regardless of whether he is described as good or as evil.

5.1. Methods

Four hundred and forty participants (295 females, 145 males, $M_{age} = 26.8$ years) were recruited from Prolific in exchange for \$0.50. The sample size was based on a target of at least 100 participants per cell. Thirteen participants were excluded from the final analysis as a result of failing either the comprehension check or the attention check. This brought the final sample size down to 427.

Participants were randomly assigned to one of 4 conditions (good/emotional, good/emotionless, evil/emotional, evil/emotionless). In each case, participants read about an individual, Jim, who was described as either possessing or lacking emotional capacities, and as either possessing or lacking morals. Importantly, in the condition in which Jim was described as having emotional capacities, while lacking morals, he was only described as possessing *non-moral* emotions. The vignette read as follows:

Jim is completely lacking in morals. He is a fundamentally evil, psychopathic, serial killer who sees nothing wrong with actions that cause his victims to suffer immensely. Jim feels a range of emotions. For example, he feels joy, sadness, pride, and nostalgia. After he abducted and killed a 3-year-old girl, Jim experienced satisfaction and pride.

We counterbalanced the order in which the pieces of information regarding the character's emotionality and moral character were conveyed across participants. This resulted in 8 vignettes (see Appendix B for the full set). As with Experiment 3, participants were then asked to rate an *ultimately non-member* statement in terms of how weird or natural it sounded to them.

Participants also completed a comprehension check and an attention check.

5.2 Results and discussion

A 2×2 ANOVA was conducted, with moral character (good, evil) and emotionality (emotional, emotionless) as between-subjects factors; and participants' ratings as the dependent variable. The mean ratings are shown in Fig. 6 (the distribution of responses for each condition is presented in Supplementary Material S6). Participants gave higher ratings in the *evil* conditions than they did in the *good* conditions ($F(1,423) = 32.49, p < .001, \eta_p^2 = .071$). Participants also gave higher ratings in the *emotionless* conditions than they did in the *emotional* conditions ($F(1,423) = 4.47,$

$p = 0.04$, $\eta_p^2 = .010$). The interaction between moral character and emotionality was not significant ($p = 0.053$).

Together with the findings from Experiment 3, these results suggest that in addition to morally good traits, people may regard emotionality as somewhat central to being a true human. Moreover, these results are consistent with the claim that people do not just regard emotions as indicators of true humanity because they see them as necessary for being a good moral agent: rather, beliefs about emotionality may have an independent effect on people's intuitions about true humanity. Nonetheless, given that the interaction between emotionality and moral character was close to significant, this is an issue that warrants further attention in subsequent research.

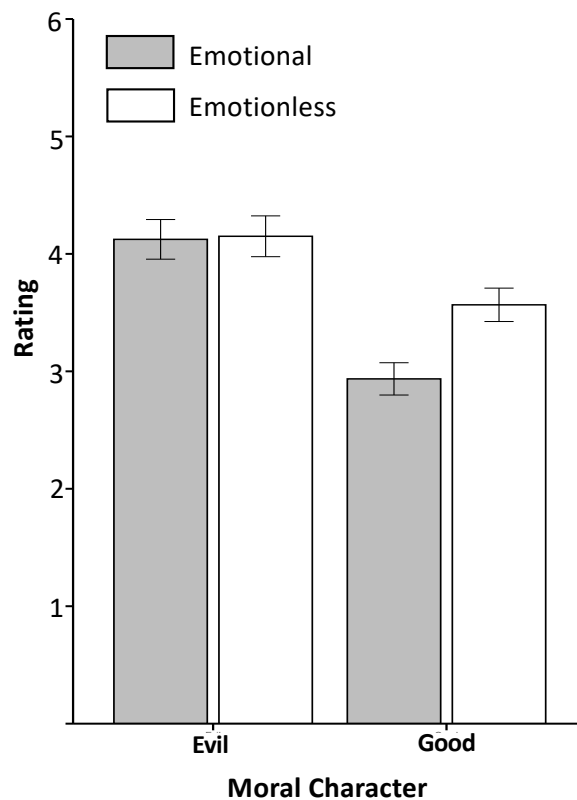


Figure 6. Mean ratings of *ultimately non-member* statement in Experiment 4 (error bars show SE mean).

6. General Discussion

The experiments reported above provide evidence that people tend to deny that someone is a true human when they perceive them as essentially bad. Experiment 1 provided evidence that people associate two distinct sets of criteria with the concept HUMAN. According to the descriptive criterion, being human is a matter of being a *Homo sapiens*. According to the normative criteria, being human is largely a matter of having a commitment to do the morally right thing.

Experiment 2 extended these findings by providing evidence that people are willing to deny that someone is a true human if they see them as lacking a true good self. Experiments 3 and 4 provided evidence that in addition to good moral character, people may regard emotionality as central to being a true human.

6.1. *The paradox of dehumanization*

The findings reported here provide a solution to the paradox of dehumanization. Recall that in paradigmatic cases of dehumanization, such as the Holocaust, the perpetrators tend to attribute uniquely human traits to their victims. For example, the Nazis frequently characterized Jewish people as criminals, traitors, and enemies (see Herf, 2008). They also treated them as bad moral agents, and subjected them to severe forms of punishment and humiliation (see Gutman & Berenbaum, 1998). Criminality, treachery, enmity, and moral agency are not capacities that we tend to attribute to nonhuman animals. Thus, can we really say that the Nazis thought of their victims as nonhuman?

In responding to this paradox, some theorists have suggested that the perpetrators in these paradigmatic cases do not, in fact, think of their victims as nonhuman (Appiah, 2008, p. 144; Bloom, 2017; Kuper, 1981, pp. 86, 92; Lang, 2010; Manne, 2016, 2018, chapter 5; Over, 2020). Other theorists have suggested that the perpetrators harbor inconsistent representations of their

victims, simultaneously thinking of them as both human and subhuman (Smith, 2016, 2020). Our findings suggest a third possibility: namely, that the perpetrators harbor a dual character concept of humanity, categorizing their victims as human in one sense, but denying that they are human in another sense. For example, Nazi leaders frequently characterized their victims as enemies, criminals, and moral deviants. However, when categorizing their victims as such, the Nazis may have been thinking of them as nonhuman in the normative sense, while recognizing them as human in the species sense (Steizinger, 2018, provides a compelling analysis of Nazi ideology and comes to a similar conclusion; see also de Ruiter, 2021; Kronfeldner, draft, 2018).³ This squares away with the fact that when the Nazis likened Jewish people to animals, such as rats, they frequently adopted a moralizing tone. For example, in an antisemitic book entitled *The Eternal Jew* (Nachfolger, 1937), Jewish neighborhoods in Berlin were described as “breeding grounds of criminal and political vermin.” Similarly, when the Nazis referred to Jews as “subhumans,” they often characterized them as bad moral agents. For example, as was mentioned above, Goebbels described Bolshevism as “the declaration of war by Jewish-led international subhumans against culture itself.” Similarly, in one 1943 Nazi pamphlet, Marxist values are described as appealing to subhumans, while liberalist values are described as “allowing the triumph of subhumans” (Anonymous, 1943, chapter 1).⁴

³ There is an ongoing debate about the psychology of the Nazi perpetrators (e.g., see Arendt, 1963; Haslam & Reicher, 2007, 2008; Lang, 2010; Steizinger, 2018). Arendt (1963) observed the trial of Adolf Eichmann—a central figure in the implementation of the “Final Solution.” She concluded that his actions were sustained by “sheer thoughtlessness” (Arendt, 1963, pp. 287–288). Arendt goes on to argue that “desk murderers” like Eichmann had banal, rather than monstrous, motives (e.g. they were simply “following orders”). For a critical discussion of Arendt’s thesis, including her interpretation of the Eichmann case, see Haslam and Reicher (2007, 2008). Debates of this sort show that we should be wary of overapplying the model of dehumanization defended above. It is plausible that not all participants in paradigmatic cases of dehumanization think of their victims as evil individuals, who lack true humanity—some may act on banal intentions of the sort identified by Arendt. Moreover, as Kronfeldner (draft) points out, we shouldn’t assume that each and every instance of the paradox is amenable to the same solution.

⁴ In Experiment 1, we found that people are willing to affirm someone’s humanity in the descriptive sense, while denying it in the normative sense, *and vice versa*. However, this double dissociation is not necessary for resolving the paradox of dehumanization. To resolve the paradox, it suffices to show that people are willing to affirm someone’s humanity in the descriptive sense, while denying it in the normative sense. In other words, even if people

Our findings also provide an explanation of why the perpetrators in these paradigmatic cases sometimes *explicitly* affirm and deny that their victims are human. For example, recall that despite characterizing Jewish people as “subhumans,” Goebbels once remarked that Jewish people are “human beings,” and that “None of us ever doubted it” (Eitzen, 1936). Instead of flatly contradicting himself, it is possible that Goebbels was categorizing Jews as human in the species sense, but denying that they are human in the normative sense. Similarly, when Maria initially referred to the Roma as “people,” before characterizing them as “not really people,” she may have been tacitly endorsing an *ultimately non-member* statement to the effect that the Roma are human in the species sense, but not the normative sense. This is evidenced by the fact that when justifying her claim that the Roma are “not really people,” Maria claimed that they have severe moral shortcomings, referring to them as criminals and perpetual killers.⁵

In applying our model of dehumanization to the paradigmatic cases, some additional questions remain. For example, even though the Nazis regarded Jewish people as evil, it is plausible that they regarded them as empathetic *towards one another*.⁶ This suggests that people may regard the tendency to empathize with *other humans*—not just the members of one’s ingroup—as central to true humanity. More generally, it is likely that people do not regard each moral trait as equally central to true humanity. For example, people may regard a commitment to respect the rights of others as more central to true humanity than generosity and kindness. Alternatively, people may see the latter as more central than the former. Relatedly, there may be individual and cross-cultural differences concerning which moral traits are regarded as the most

think that someone cannot be a true human without also being a *Homo sapiens*, the paradox is still resolvable, so long as they think that someone can be a *Homo sapiens* without being a true human. Thank you to Kevin Reuter for bringing this point to our attention.

⁵ Maria did not use the term “humans.” However, “people” is often used as a synonym for “humans.” Moreover, in clarifying her claim that the Roma are “not really people,” Maria also used the term “subhuman.”

⁶ Thank you to Kevin Reuter for raising this issue.

central to true humanity (on the general issue of cross-cultural differences in concepts of humanity, see Steizinger, 2018, p. 16; Kronfeldner, 2018, Chapter 2).

6.2. The role of beliefs about evil and the true self

Another issue that warrants further examination concerns the roles that beliefs about evil and the true self play in driving denials of true humanity. Experiment 2 provided evidence that when people deny that someone is a true human, they do so because they perceive them as lacking a true good self. In other words, they do so because they perceive the given individual as essentially bad. Previous research has revealed that people have a very robust tendency to see others as possessing true good selves. For example, people tend to see political opponents (Newman et al., 2014, 2015), and even threatening outgroup members (De Freitas and Cikara, 2018), as possessing true good selves. Our findings suggest that people are willing to attribute true *bad* selves to certain individuals. This gives rise to an obvious question: At what point do people suppress their tendency to see others as essentially good, instead, representing an individual as essentially bad, and thus nonhuman?

To address this question, additional research into beliefs about evil will be important. Extant research has uncovered individual differences in beliefs about the existence of evil (Bastian et al., 2015; Webster and Saucier, 2013; see also Heiphetz, 2020; Martin and Heiphetz, in press). These differences may, in turn, give rise to individual differences in the tendency towards dehumanization. For instance, suppose *A* has a stronger tendency to believe in the existence of evil individuals than *B* does. Presumably, this means that *A* is more likely to deny that someone is a true human than *B*, because *A* is more likely to regard certain individuals as lacking true good selves.

Future research into beliefs about the true self may also reveal methods for preempting dehumanization. As was mentioned above, De Freitas and Cikara (2018) found that threatening outgroup members are often seen as having true good selves. Importantly, De Freitas and Cikara also found that when people explicitly represent threatening outgroup members as having true good selves, this reduces their biases against them. By focusing on outgroup members' true good selves, people may also increase their tendency to explicitly represent them as human, in the normative sense. In other words, focusing on outgroup members' true good selves may preempt "dehumanization by omission": a form of dehumanization in which one fails to categorize another individual as human, rather than categorizing them as nonhuman (see Waytz and Schroeder, 2014).

On the other hand, our findings suggest that if one perceives an outgroup member as evil, then reflecting on this individual's true self will actually *enhance* one's tendency to dehumanize them. This might explain why the mean ratings for the *ultimately non-member* statement were at the midpoint in Experiments 3 and 4, but well above the midpoint in Experiments 1 and 2. The former did not prompt participants to reflect on the target's true self, whereas, the latter did. Most notably, in Experiment 2, participants were asked to reflect on the target's true self before rating the *ultimately non-member* statement. Subsequent research on this issue should examine whether priming concepts associated with the true self enhances the tendency to deny that "evil" individuals are true humans. Nazi propagandists consistently utilized phrases implying that Jewish people deliberately conceal their true bad selves. In a very typical example of this, one Nazi propagandist claimed that the "fable of the decent Jew ... is a shameless lie," and that Jewish people wear "masks" that hide their "true intentions" and "their very nature"

(Oberlindober, 1940). Phrases of this sort may have enhanced dehumanizing beliefs about Jewish people by priming the audience's concept of the true self.

Finally, subsequent research should examine the extent to which beliefs about true humanity concern factors that go beyond the true *good* self. Experiments 3 and 4 provided evidence that in addition to good moral traits, people may regard emotional capacities as somewhat central to true humanity. In determining why this might be the case, there are various possibilities to consider. First, there is evidence that people regard emotional memories as somewhat central to the self, even if they regard moral traits as more central (Strohmingner & Nichols, 2014). This suggests that people may regard emotional capacities as somewhat central to true humanity because they regard them as somewhat central to the self.

Second, research suggests that in addition to morally good behavior, people regard feelings of fulfilment as important for leading a meaningful life (e.g., see Prinzing et al., 2021). It is possible that beliefs about what it means to be a true human are informed by beliefs about what it takes to lead a meaningful human life. If so, this would provide a unifying explanation as to why both moral goodness and emotionality are regarded as central to being a true human.

Third, there is research suggesting that attributions of emotion are constrained by normative considerations. Based on a series of studies, Phillips et al. (2011, 2014, 2017) conclude that according to the folk, feeling good is insufficient for being happy: one must be morally good as well. More recently, Díaz and Reuter (2020) found that people are less likely to attribute happiness to someone who feels good about something that is morally bad than someone who feels good about something that is morally good. It is possible that the tendency to represent someone as a true human is enhanced when they are perceived as experiencing certain emotions, such as happiness, joy, or even pride, because feeling positive about morally good outcomes—

one of the hallmarks of a true human—is seen as a prerequisite for experiencing these emotions in the first place.

Finally, Del Pinal and Reuter (2017) have provided compelling evidence that the normative dimension of a dual character concept represents the commitment to some idealized function (see also Reuter et al., 2020). For example, on this view, people will regard *S* as a “true artist” only if they believe that *S* is *committed* to producing deep works of art. Our findings suggest that people attribute true humanity to someone when they perceive them as having a deep-seated commitment to morally good behavior. It is possible that people also regard certain emotions as necessary for being a true human because they see them as a general requirement for maintaining commitments, including the commitment to behave morally.

6.3. Consequences for extant models and measures of dehumanization

Our findings have implications for extant models of dehumanization. Recall that according to Haslam’s (2006) model, animalistic dehumanization occurs when we represent another individual as possessing uniquely human traits to a lesser extent than certain others (e.g. our fellow ingroup members). Haslam claims that when this occurs, one tacitly represents the given individual as *less human-like* than members of the relevant comparison group (see also Leyens et al., 2000, 2001). Clearly, though, we do not represent all uniquely human traits as human-typical. For example, playing poker is a uniquely human capacity. However, hardly anyone would regard it as a human-typical trait. In other words, if someone does not possess the ability to play poker, hardly anyone would represent them—tacitly or otherwise—as less human-like than someone who can. Thus, there is a question of which uniquely human capacities are thought of as highly typical of humans, and which ones are not. The findings reported here suggest that moral traits—in particular, those centered on having a deep-seated commitment to do the morally right thing—

will be seen as highly typical of true humans. This means that perception of moral character may have a larger role to play in determining how (truly) human-like someone seems than the perception of those non-moral traits that people regard as uniquely human, such as rationality and culturedness (for a relevant discussion, see Schwartz & Struch, 1989).

Similarly, consider blatant dehumanization. In contrast to tacit forms of dehumanization, the blatant variety occurs when one explicitly represents another individual as less human-like than someone else. In a series of studies, Kteily and colleagues (2015) have developed a measure of blatant dehumanization that utilizes the Ascent of Man image, which depicts a series of ape-like creatures through to a modern human. Accompanying the image is a caption which says that some people seem more “evolved” and “human-like” than others. Participants are then asked to rate the members of various groups in terms of how “evolved” they seem. Kteily and colleagues have found that the Ascent measure is a unique predictor of hostility towards outgroups. Moreover, they have uncovered evidence that it does not just measure outgroup dislike (Bruneau et al., 2018; Kteily and Bruneau, 2017). Nonetheless, which dehumanizing beliefs people are expressing when they utilize the Ascent scale remains an open question (see Kteily and Bruneau 2017, p. 491).

When a participant gives the members of an outgroup a relatively low Ascent rating, they may be expressing the belief that these outgroup members are relatively “unevolved” in the normative sense. For example, paradigmatically immoral outgroups, such as racists and sexists, tend to receive the lowest Ascent ratings (see Heiphetz and Craig, in press). Participants may perceive racists and sexists as less human-like and “evolved” than ingroup members because they regard the former as immoral individuals who are failing to manifest their true good selves. Another belief that participants may be expressing when they give relatively low Ascent ratings to

outgroup members is that they are less “evolved” in the species sense. For example, Williams and Eberhardt (2008) found that people with biological conceptions of race are more likely to accept racial inequalities, perceiving them as natural and unlikely to change. People with this conception of race may view the members of racial outgroups as relatively “unevolved” in the species sense. Of course, they may also view them as relatively “unevolved” in the normative sense. To disentangle these different forms of blatant dehumanization, novel measures will be required.

6.4. Conclusion

We have presented evidence that people harbor a dual character concept of humanity. In examining the normative dimension, we showed that people tend to regard both moral character and emotionality as central to true humanity. We also showed that people tend to believe that being a true human requires a true self that is committed to doing the morally right thing. Finally, we argued that our account provides a solution to the paradox of dehumanization; and that it has important implications for extant models of dehumanization, including the measures that are utilized to support them.

Appendix

A.1. Vignettes used in Experiment 3

Jim is completely lacking in morals. He is an evil, psychopathic, serial killer who completely lacks empathy for the suffering of his victims. In fact, he very much enjoys killing innocent people. In general, Jim has no concern at all about the rights of others.

Jim is completely irrational. He never makes sensible decisions about even the most basic day-to-day activities. Instead, his decisions are the result of irrational urges, and random thoughts that

occur to him on a whim. Jim cannot go about his daily life without 24-hour assistance from others.

Jim is completely lacking in intellectual ability. For example, he is unable to solve problems that are easy for a 2-year-old child. He is unable to read, nor can he do basic arithmetic. Jim cannot go about his daily life without 24-hour assistance from others.

Jim is completely reserved and antisocial. He has no desire to be around others. When he does interact with others, he is cold, awkward, and completely aloof. For example, when someone tries to converse with him, he walks away without even making eye-contact.

Jim is completely emotionless. He never feels the emotions that others do, such as joy, sadness, and fear. For example, when he got a promotion at work, Jim felt nothing—he did not feel even the slightest tinge of happiness.

Jim is completely closed-minded. He has no curiosity about the world around him. He is not open to new experiences, nor does he have any desire to learn about new ideas that might challenge his view of the world.

Jim is completely uncultured. He has no interest in music, art, or science. He is extremely vulgar in both the way he looks, and the way he talks. Jim is widely seen as a slob, whose main interests are eating and sleeping.

B.1. Vignettes used in Experiment 4

Jim is completely lacking in morals. He is a fundamentally evil, psychopathic, serial killer who sees nothing wrong with actions that cause his victims to suffer immensely. Jim does not

experience emotions. For example, he never feels joy, sadness, pride, or nostalgia. After he abducted and killed a 3-year-old girl, Jim did not experience any guilt or sadness.

Jim does not experience emotions. For example, he never feels joy, sadness, pride, or nostalgia. Jim is completely lacking in morals too. He is a fundamentally evil, psychopathic, serial killer who sees nothing wrong with actions that cause his victims to suffer immensely. Jim did not experience any guilt or sadness after he abducted and killed a 3-year-old girl.

Jim is completely lacking in morals. He is a fundamentally evil, psychopathic, serial killer who sees nothing wrong with actions that cause his victims to suffer immensely. Jim feels a range of emotions. For example, he feels joy, sadness, pride, and nostalgia. After he abducted and killed a 3-year-old girl, Jim experienced satisfaction and pride.

Jim feels a range of emotions. For example, he feels joy, sadness, pride, and nostalgia. However, Jim is completely lacking in morals. He is a fundamentally evil, psychopathic, serial killer who sees nothing wrong with actions that cause his victims to suffer immensely. Jim experienced satisfaction and pride after he abducted and killed a 3-year-old girl.

Jim is deeply committed to doing the right thing by others, and he tries to help those in need whenever he can. He often performs random acts of kindness in which he sacrifices his own needs for the sake of helping others. Jim does not experience emotions. For example, he never feels joy, sadness, pride, or nostalgia. After donating one of his kidneys to save the life of a 3-year-old girl, Jim did not experience any satisfaction or pride.

Jim does not experience emotions. For example, he never feels joy, sadness, pride, or nostalgia. Nonetheless, Jim is deeply committed to doing the right thing by others, and he tries to help

those in need whenever he can. He often performs random acts of kindness in which he sacrifices his own needs for the sake of helping others. Jim did not experience any satisfaction or pride after donating one of his kidneys to save the life of a 3-year-old girl.

Jim is deeply committed to doing the right thing by others, and he tries to help those in need whenever he can. He often performs random acts of kindness in which he sacrifices his own needs for the sake of helping others. Jim feels a range of emotions. For example, he feels joy, sadness, pride, and nostalgia. After donating one of his kidneys to save the life of a 3-year-old girl, Jim experienced satisfaction and pride.

Jim feels a range of emotions. For example, he feels joy, sadness, pride, and nostalgia. Jim is also deeply committed to doing the right thing by others, and he tries to help those in need whenever he can. He often performs random acts of kindness in which he sacrifices his own needs for the sake of helping others. Jim experienced satisfaction and pride after donating one of his kidneys to save the life of a 3-year-old girl.

Acknowledgements

I would like to thank Josh Knobe, Nour Kteily, Maria Kronfeldner, and Kevin Reuter for invaluable feedback. I am also grateful to Laura, Marg, and Rog for allowing me to probe their intuitions.

References

- Anonymous. (1943). *Der Reichsführer SS/SS-Hauptamt, Rassenpolitik*.
- Appiah, K. A. (2008). *Experiments in ethics*. Cambridge, MA.: Harvard University Press.
- Arendt, H. (1963). *Eichmann in Jerusalem: A report on the banality of evil*. New York: Penguin.

- 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*. doi: [10.1037/xge0000822](https://doi.org/10.1037/xge0000822)
- Bastian, B., Bain, P., Buhrmester, M. D., Gómez, Á., Vásquez, A., Knight, C. G., & Swann, Jr., W. B. (2015). Moral vitalism: Seeing good and evil as real, agentic forces. *Personality and Social Psychology Bulletin*, 41(8), 1069–1081. doi:[10.1177/0146167215589819](https://doi.org/10.1177/0146167215589819)
- Bloom, P. (2017, November 20). The root of all cruelty? *The New Yorker*.
<https://www.newyorker.com/magazine/2017/11/27/the-root-of-all-cruelty>
- Bridge, A. (1993, October 18). Romanians vent old hatreds against Gypsies: the villagers of Hadereni are defiant about their murder of vermin. *The Independent*.
<https://www.independent.co.uk/news/world/romanians-vent-old-hatreds-against-gypsies-the-villagers-of-hadareni-are-defiant-about-their-murder-of-vermin-adrian-bridge-reports-1511734.html>
- Bruneau, E., Jacoby, N., Kteily N., & Saxe, R. (2018). Denying humanity: The distinct neural correlates of blatant dehumanization. *Journal of Experimental Psychology: General*, 147(7), 1078–1093.
- De Freitas, J. L., & Cikara, M. (2018). Deep down my enemy is good: Thinking about the true self reduces intergroup bias. *Journal of Experimental Social Psychology*, 74, 307–316.
- Del Pinal, G., & Reuter, K. (2017). Dual character concepts in social cognition: Commitments and the normative dimension of conceptual representation. *Cognitive Science*, 41(S3), 477–501. <https://doi.org/10.1111/cogs.12456>
- De Ruiter, A. (2021). To be or not to be human: Resolving the paradox of dehumanisation. *European Journal of Political Theory*. <https://doi.org/10.1177/1474885120984605>

- Díaz, R., & Reuter, K. (2020). Feeling. The right way: Normative influences on people's use of emotion concepts. *Mind & Language*, 36(3), 451–470.
- Eitzen, K. H. (1936). Zehn knüppel wider die Judenknechte. *Unser Wille und Weg*, 6, 309–310.
- 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*. <https://doi.org/10.1016/j.cognition.2021.104682>
- Guo, C., Dweck, C. S., & Markman, E. M. (2021). Gender categories as dual-character concepts? *Cognitive Science*, 45(5), e12954. doi: 10.1111/cogs.12954.
- Gutman, Y., & Berenbaum, M. (1998). *Anatomy of the Auschwitz death camp*. Indiana University Press.
- Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, 10(3), 252–264.
- Haslam, N., Bain, P., Douge, L., Lee, M., & Bastian, B. (2005). More human than you: Attributing humanness to self and others. *Journal of Personality and Social Psychology*, 89, 937–950.
- Haslam, S. A., & Reicher, S. D. (2007). Beyond the banality of evil: Three dynamics of an interactionist social psychology of tyranny. *Personality and Social Psychology Bulletin*, 33(5), 615–622.
- Haslam, S. A., & Reicher, S. D. (2008). Questioning the banality of evil. *Psychologist*, 21(1), 16–19.
- Heiphetz, L. (2020) The development and consequences of moral essentialism. *Advances in Child Development and Behavior*, 59, 165–194.

- Heiphetz, L., & Craig, M. (in press). Dehumanization and perceptions of immoral behavior. *Oxford Studies in Experimental Philosophy*.
- Herf, J. (2008). *The Jewish enemy: Nazi propaganda during World War II and The Holocaust*. Harvard University Press.
- Knobe, J., Prasada, S., & Newman, G. E. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition*, 127(2), 242-57. doi: 10.1016/j.cognition.2013.01.005.
- Kronfeldner, M. (2018). *What's left of human nature? A post-essentialist, pluralist, and interactive account of a contested concept*. MIT Press.
- Kronfeldner, M. (draft). The mirage of a paradox of dehumanization, or: How to reclaim the reality of dehumanization.
- Kteily, N. S., & Bruneau, E. (2017). Darker demons of our nature: The need to (re)focus attention on blatant forms of dehumanization. *Current Directions in Psychological Science*, 26(6), 487–494.
- 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.
- Kuper, L. (1981). *Genocide: Its political use in the twentieth century*. New Haven: Yale University Press.
- Lang, J. (2010). Questioning dehumanization: Intersubjective dimensions of violence in the Nazi concentration and death camps. *Holocaust and Genocide Studies*, 24(2), 225–246.
- Laraiyetan, H. (2021). *The true humanity in us: 52 great acts in a year*. Westbow Press.

- Leslie, S.-J. (2015). “Hillary Clinton is the only man in the Obama administration”: Dual character concepts, generics and gender. *Analytic Philosophy*, 56(2), 111–141.
- Leyens, J.-P., Paladino, P. M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The emotional side of prejudice: The attribution of secondary emotions to ingroups and outgroups. *Personality and Social Psychology Review*, 4(2), 186–197.
- Leyens, J.-P., Rodriguez-Perez, A., Rodriguez-Torres, R., Gaunt, R., Paladino, M.-P., Vaes, J., & Demoulin, S. (2001). Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology*, 31(4), 395–411.
- Manne, K. (2016). Humanism: A critique. *Social Theory and Practice*, 42, 389–415.
- Manne, K. (2018). *Down girl*. Oxford: Oxford University Press.
- Martin, J., & Heiphetz, L. (In press). “Internally wicked”: Investigating how and why essentialism influences punitiveness and moral condemnation. *Cognitive Science*.
- Nachfolger, F. E. (1937). *The eternal Jew*. Munich: Zentralverlag der NSDAP.
- Newman, G. E., De Freitas, J., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science*, 39(1), 96–125.
- Newman, G., Knobe, J., & Bloom, P. (2014). Value judgments and the true self. *Personality and Social Psychology Bulletin*, 40, 203–216.
- Oberlindober, H. (1940). *Ein Vaterland, das allen gehört! Briefe an Zeitgenossen aus zwölf Kampffahren*. Munich: Zentralverlag der NSDAP.
- Over, H. (2020). Seven challenges for the dehumanization hypothesis. *Perspectives on Psychological Science*. <https://doi.org/10.1177/1745691620902133>

- Phillips, J., de Freitas, J., Mott, C., Gruber, J. & Knobe, J. (2017). True happiness: The role of morality in the folk concept of happiness. *Journal of Experimental Psychology: General*, 146(2), 165–181. <https://doi.org/10.1037/xge0000252>
- Phillips, J., Misenheimer, L. & Knobe, J. (2011). The ordinary concept of happiness (and others like it). *Emotion Review*, 71(3), 929–937. <https://doi.org/10.1177/1754073911402385>
- Phillips, J., Nyholm, S. & Liao, S.-Y. (2014). The good in happiness. In T. Lombrozo, J. Knobe & S. Nichols (Eds.), *Oxford studies in experimental philosophy*. Oxford: Oxford University Press.
- Prinzinger, M., de Freitas, J., & Fredrickson, L. (2021). The ordinary concept of a meaningful life: The role of subjective and objective factors in third-person attributions of meaning. *The Journal of Positive Psychology*. <https://doi.org/10.1080/17439760.2021.1897866>
- Quora. (n.d.). *Who do you think is a true human being?* <https://www.quora.com/Who-do-you-think-is-a-true-human-being>
- 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.
- Reuter, K., Löschke, J., & Betzler, M. (2020). What is a colleague? The descriptive and normative dimension of a dual character concept. *Philosophical Psychology*, 33(7), 997–1017.
- 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 University Press.
- Steizinger, J. (2018). The significance of dehumanization: Nazi ideology and its psychological consequences. *Politics, Religion and Ideology*, 19(2), 139–157.
- Strohming, N., Knobe, J., & Newman, G. (2017). The true self: A psychological concept distinct from the self. *Perspectives on Psychological Science*, 12(4), 551–560.
- Strohming, N., & Nichols, S. (2014). The essential moral self. *Cognition*, 131, 159–171.
- Tobia, K. P., Newman, G. E., & Knobe, J. (2020). Water is and is not H₂O. *Mind & Language*, 35, 183–208.
- United Nations. (n.d.) *Nelson Mandela International Day: 18 July*.
<https://www.un.org/en/events/mandeladay/tribute.shtml>
- United States Holocaust Memorial Museum. (n.d.). *Goebbels claims Jews will destroy culture*.
https://usc.ushmm.org/wlc/en/media_fi.php?ModuleId=0&MediaId=192
- Waln, V. (2015, November 19). Thank tribal activists for defeating Keystone XL permit. *Indian Country Today*. <https://www.indianz.com/News/2015/11/19/vi-waln-thank-tribal-activists.asp>
- Waytz, A., & Schroeder, J. (2014). Overlooking others: Dehumanization by commission and omission. *TPM-Testing, Psychometrics, Methodology in Applied Psychology*, 21(3), 251–266.

Webster, R. J., & Saucier, D. A. (2013). Angels and demons are among us: Assessing individual differences in belief in pure evil and belief in pure good. *Personality and Social Psychology Bulletin*, 39(11), 1455–1470. doi:[10.1177/0146167213496282](https://doi.org/10.1177/0146167213496282)

Williams, M. J., & Eberhardt, J. L. (2008). Biological conceptions of race and the motivation to cross racial boundaries. *Journal of Personality and Social Psychology*, 94(6), 1033–1047.