Seven challenges for the dehumanisation hypothesis

Harriet Over
University of York

In press at Perspectives on Psychological Science

Acknowledgements: I thank Paul Bloom, Rich Cook, Cade McCall, Florence Enock, David Over and three anonymous reviewers for valuable comments on an earlier draft. This research was supported by the European Research Council under the European Union's Horizon 2020 Programme, grant number ERC-STG-755719.

Address for correspondence
Dr. Harriet Over
Department of Psychology
University of York
YO10 5DD
Email: harriet.over@york.ac.uk

Abstract

Propaganda often compares members of stigmatised outgroups to non-human entities such as rats, lice and snakes. Drawing on these horrifying descriptions, the dehumanisation hypothesis proposes that outgroup members are viewed as 'less than human', and that being viewed as less than human renders them vulnerable to harm. I offer seven challenges to the dehumanisation hypothesis. I argue that, even in supposedly prototypical examples of extreme dehumanisation, outgroup members are not treated in a similar way to non-human entities. Furthermore, although outgroup members may be denied some human qualities and states, they are attributed others. I also argue that there is reason to doubt the hypothesised causal connection between being viewed as less than human and being at risk of harm – some non-human organisms are treated with great care, and some groups are harmed because of how their uniquely human qualities are perceived. I close by offering an alternative account of why outgroup members are sometimes referred to as nonhuman entities.

Seven challenges for the dehumanisation hypothesis

Between 1941 and 1945, the Nazis murdered up to 6 million of Europe's Jewish population (Kershaw, 2001). In the spring of 1994, Rwandan Hutus murdered close to 1 million Tutsis (Tirrell, 2012). In the early 2000s, American soldiers tortured and murdered Iraqi prisoners of war at Abu Ghraib (Apel, 2005). Victims of these atrocities, and others like them, were often compared to non-human entities - to rats, lice, dogs and cockroaches (Smith, 2011; 2014).

Research on dehumanisation seeks to offer a causal explanation for the apparent association between describing victims of atrocities as non-human and inflicting harm upon them. Research in this field can be broadly summarised as two related hypotheses 1) Victims of intergroup harm are perceived to be similar to non-human entities 2) As a result, natural inhibitions against causing them harm are eroded leading, in extreme cases, to genocide and torture. I refer to these two inter-related claims as 'the dehumanisation hypothesis'.

The dehumanisation hypothesis is ambitious in that it seeks to explain a wide range of phenomena by appealing to the same psychological construct. Blatant dehumanisation, in which outgroup members are removed from the human category, is thought to play a causal role in genocide and torture (Smith, 2011; 2014; 2016; Viki, Osgood, & Phillips, 2013). More subtle forms of dehumanisation, in which outgroup members are viewed as 'less human' than the ingroup, are thought to play a causal role in less extreme acts of harm such as withholding help from individuals in need and feeling reduced concern for their well-being (Andrighetto, Baldissarri, Lattanzio, Loughnan, & Volpato, 2014; Cuddy, Rock, & Norton, 2007; Haslam 2006; Leyens et al., 2007).

Since its conception, the dehumanisation hypothesis has gained considerable prominence in social neuroscience (Harris, 2017; Harris & Fiske, 2006; 2009; 2011), social psychology (Haslam, 2006; Leyens, Demoulin, Vaes, Gaunt, & Paladino, 2007) and philosophy (Smith, 2011; 2014; 2016). Indeed, the hypothesis has become so prominent that the idea certain groups suffer dehumanisation has been described as something of a truism (Smith, 2016).

I will argue that, while the dehumanisation hypothesis is prima facie reasonable, and indeed intuitively compelling, it does not withstand scrutiny. I review the major formulations of the dehumanisation hypothesis from

neuroscience, psychology and philosophy. Combining across behavioural data, brain data, and the historical evidence, I suggest that outgroup members are not represented in a similar way to non-human entities. Rather, they are likely attributed characteristics that are typical of humans but anti-social in character such as jealousy, spite and cunning (Appiah 2008; Bloom, 2017; Lang, 2010; Manne, 2016). Furthermore, rather than protecting them from harm, being viewed as having these distinctly human attributes may put them at greater risk of harm (Bloom 2017; Gopnik, 2006; Lang, 2010; Manne, 2016; 2018).

Variants of the dehumanisation hypothesis

The dehumanisation hypothesis is not represented by a single theory but rather by a family of theories from neuroscience, psychology and philosophy (Harris & Fiske, 2006; 2011; Haslam, 2006; Haslam & Loughnan, 2014; Leyens et al., 2007; Smith, 2011; 2016; Vaes, Leyens, Paladino, & Miranda, 2012). These theories share key underlying assumptions but differ in the particular way in which they characterise the construct of dehumanisation. That is to say, each theory offers a somewhat different characterisation of what is thought to be 'missing' when a person or group is considered less than human. Below I briefly review the most prominent contemporary theories from philosophy (Smith, 2011; 2016), neuroscience (Harris & Fiske, 2006; 2011) and psychology (Haslam, 2006; Leyens et al., 2007).

The philosopher Smith (2011) offers a characterisation of dehumanisation that centres around essentialism. According to this theory, humans are those to whom we attribute a 'human essence'. Those to whom we attribute a 'subhuman essence' are dehumanised. Smith draws the majority of his evidence from historical documents, including analyses of propaganda, in which target groups are described as dangerous or disgusting animals such as rats and lice. In his more recent writings, Smith (2016) argues that dehumanisation involves simultaneously categorising a group as human in appearance but subhuman in essence. According to this view, simultaneously categorising someone as human and sub-human gives rise to a feeling of 'uncanniness' or creepiness (Smith, 2016).

In contrast, Harris and Fiske (2006; 2011) offer a characterisation of dehumanisation that focuses on mental state attribution. According to this theory, when we consider a social group to be human, we attribute mental states to them. When we consider a social group to be 'less than human', we either do not attribute mental states to them or attribute fewer mental states to them (Harris & Fiske, 2006; 2011). Harris and Fiske (2006; 2011) argue that groups perceived to be low in warmth and competence, such as drug addicts and the homeless, are particularly likely to be dehumanised. To the extent that they are dehumanised, they will elicit disgust. Harris and Fiske (2006) provide neuroscientific data to support their characterisation - participants display less activity in brain regions associated with mentalising, more specifically the medial prefrontal cortex, when viewing images of homeless individuals and people addicted to drugs compared to when viewing images of their own group. Convergent evidence comes from behavioural work demonstrating that adults use fewer mental state words when describing the daily activities of individuals from these groups (Harris & Fiske, 2011) and data suggesting that participants have more stringent criteria for perceiving a mind behind an artificial face when that face belongs to an outgroup (Hackel, Looser, & van Bavel, 2014).

Leyens and colleagues (2007) explicitly departed from the tradition of seeking to explain extreme intergroup harm such as genocide and torture, instead focusing on more subtle manifestations of intergroup biases within contemporary Western society. They conducted an informal survey in which they asked participants what attributes they thought of as uniquely human. Focusing on a subset of their participants' responses that referenced emotions, Leyens and colleagues argued that groups are subtly dehumanised to the extent that they are thought to be lacking in secondary, or complex, emotions such as pride, guilt and remorse. In support of their view, they found that individuals typically attribute secondary emotions more strongly to their ingroup than to outgroups (Demoulin, et al., 2004). In order to distinguish their work from research on more blatant forms of dehumanisation, they termed this hypothesised psychological process 'infrahumanisation'.

Building on the work of Leyens and colleagues, Haslam and colleagues (Haslam, Bain, Douge, Lee and Bastian, 2005; Haslam, 2006; Haslam & Loughnan,

2014) sought to characterise dehumanisation by first understanding folk perceptions of what it means to be human. Haslam et al. (2005) asked participants to complete two tasks: First, rating a list of attributes for the extent to which they applied to humans and not to other species and, second, rating a list of attributes for the extent to which they were characteristic of humans. On the basis of their answers to these two questions, Haslam inferred there were two forms of humanness. Uniquely human attributes were those participants listed as distinguishing humans from other species and consisted of civility, refinement, moral sensibility, rationality and maturity. Human nature attributes were those participants listed as characteristic of humans and consisted of emotional responsiveness, interpersonal warmth, openness, agency and emotional depth. Haslam (2006) postulated two corresponding forms of dehumanisation: Animalistic dehumanisation, where individuals or groups are thought to possess fewer uniquely human attributes, and mechanistic dehumanisation where individuals or groups are thought to possess fewer human nature attributes.

The relationship between dehumanisation and intergroup harm

Much of the interest in dehumanisation stems from the claim that it plays a causal role in intergroup harm. According to the dehumanisation hypothesis, humans are naturally inclined to empathise with, and care for, each other. When these barriers against harm are removed, or eroded, individuals are at greater risk of discrimination (Smith, 2011; Harris & Fiske, 2011). Smith (2011; 2016) describes dehumanisation as "a psychological lubricant for the machinery of violence" arguing that it plays "a significant role in facilitating and motivating episodes of genocide, war, slavery, and other forms of mass violence". This reasoning is echoed by Harris and Fiske (2011) who argue that "dehumanized perception, a failure to spontaneously consider the mind another person, may be a psychological mechanism facilitating inhumane acts like torture". Similarly, Haslam and Loughnan (2014) argue that "a major impetus for the study of dehumanization is to understand its profoundly negative consequences. Dehumanization of enemies, victims, and colonized peoples has been associated with pogroms, atrocities, and exploitation". Haslam and Loughnan (2014) further

elaborate on this hypothesised causal connection by arguing that dehumanisation results in reductions in prosocial behaviour, commission of antisocial acts, and disinhibition of violence.

Evidence consistent with this proposed causal relation has been drawn from analyses of historical documents. Analysis of propaganda and other historical documents shows that regimes that commit genocide and other atrocities often compare their victims to non-human entities such as rats, lice and cockroaches. These examples are widely cited in psychological research on dehumanisation (e.g., Harris & Fiske, 2011; Haslam & Loughnan, 2014) but have been most comprehensively studied philosophers (Smith, 2011; Tirrell, 2012). Smith (2011), for example, focuses primarily on examples from Nazi Germany and American slavery. He points to a plethora of cases in which victims of these regimes were compared to, or even described as, subhuman creatures. Furthermore, when interviewed after atrocities they have committed, perpetrators of mass violence periodically report that they did not view their victims as human (Hatzfeld, 2003).

Further evidence consistent with the claim that dehumanisation is causally related to harm comes from lab-based research. This research typically shows that measures of dehumanisation correlate with willingness to endorse harm. For example, Kteily and colleagues showed that the extent to which American participants endorsed the claim that Arabs "seem less highly evolved" than do Americans predicted their endorsement of discrimination against Arabs (Kteily, Bruneau, Waytz, & Cotterill, 2015; see also Kteily & Bruneau, 2017). In related work, Goff, Eberhardt, Williams and Jackson (2008) found that White participants implicitly associate African Americans with apes and that participants who have been primed with ape-related words are more likely to condone police violence against a suspect but only when that suspect is African American. Other research on the attribution of human qualities has shown that participants' tendency to deny labourers in sweatshops mental experiences such as the capacity to form plans, as well as to feel emotions such as love and pain, correlates with their willingness to endorse the use of sweatshop labour (Rai, Valdesolo, & Graham, 2017).

Seven challenges for the dehumanisation hypothesis

Challenge 1: Comparisons to non-human entities are not reserved for outgroups

A key source of evidence in favour of the dehumanisation hypothesis comes from real world examples of situations in which members of certain outgroups have been compared to non-human entities (Smith, 2011; 2014; 2016; Tirrell, 2012). While intuitively compelling, these examples alone are not sufficient to conclude that outgroup members are viewed as more similar to non-human entities than are ingroup members. If using historical associations between groups and animals as one source of evidence for the dehumanisation hypothesis, it is crucial to search for disconfirmatory as well as confirmatory cases (see Nickerson, 1998, for a review of research on the confirmation bias). Surveying real-world examples more broadly, it is clear that comparisons to non-human entities are not always used as a way to insult or demean (Haslam, Loughnan, & Sun, 2011). Comparisons to animals can be used to compliment an individual, and even to highlight some of their prototypically human virtues. For example, the epithet 'Lionheart' was intended to emphasise the bravery of Richard the first in battle. While comparing a person to a monkey can sometimes be deeply offensive, using 'little monkey' as a term of endearment to describe a toddler might emphasise that they are charming and mischievous. Similarly, comparing an athlete to a machine might emphasise their perseverance and skill. Closely related to this, social groups often invoke non-human entities to refer to themselves. For example, sports teams often have non-human entities as their emblems and might refer to themselves as the lions, bulls, the blades, the gunners or the eagles (See Figure 1).

It might be possible to counter that whereas comparisons to some non-human entities are complimentary, comparisons to others are, for some as yet unspecified reason, dehumanising. However, ingroup members sometimes compare themselves to the same supposedly 'disgusting and dangerous' animals to which outgroups are often compared. For example, the Tutsis in Rwanda were often compared to snakes by propagandists (Tirrell, 2012) but the American revolutionaries often compared *themselves* to snakes. The Gadsden

flag of the American Revolution depicts the American people as a rattlesnake ready to bite the British Empire (Rankin, 1954, see Figure 1). Taken together, these examples demonstrate that comparisons to animals are not, in and of themselves, problematic nor do they necessarily reflect a deep difference in how members of ingroups and outgroups are represented. The challenge for the dehumanisation hypothesis is to explain why comparisons to non-human entities are sometimes taken as evidence for dehumanisation and sometimes not.

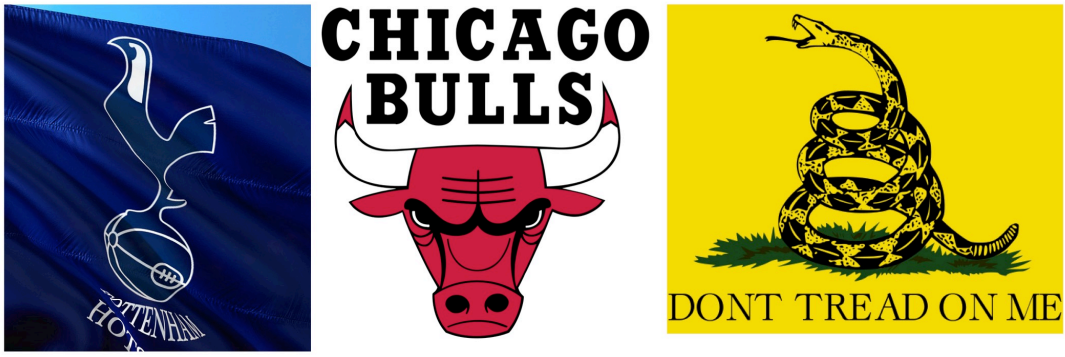

Figure 1. Groups often refer to themselves in terms of non-human entities. Tottenham Hotspur Emblem, Chicago Bulls Emblem, The Gadsden Flag.

Challenge 2: Outgroup members are often described in ways that only apply to humans

The dehumanisation hypothesis proposes that, to the extent they are dehumanised, outgroup members are perceived in a similar way to non-human entities, most commonly animals or automata (Haslam 2006; Haslam & Loughnan, 2014; Haslam, Loughnan, & Sun, 2011; Smith, 2011). Although there may be occasions on which outgroup members are described in ways that are equivalent to how animals and automata are described, these cases are much less common than they first appear. A careful reading of propaganda and hate speech reveals that target groups are often described in ways that only apply to humans (Bloom 2017; Manne, 2016). For example, Nazi propaganda often referred to Jewish people as criminals, murderers, enemies and traitors (Keen, 1992). These terms are readily applied to humans but make little sense if applied to animals or other non-human entities (Manne, 2016). A rat or an automaton cannot be a criminal or a traitor. The use of these descriptors thus

suggests that the Jewish population were at least implicitly represented as human.

It might be possible to counter this critique by arguing that outgroups are sometimes dehumanised, or dehumanised by some writers, and discriminated against in different ways at other times and by other writers. However, this idea does not fit with the historical evidence either - a variety of metaphors often appear within the same piece of propaganda. For example, in the piece 'The Jewish World Plague' Hermann Esser (1939) describes Jews as weeds, parasites and worms but also as swindlers, thieves, beggars and deceivers. In the piece, 'To know the Jew is to understand the meaning of war', distributed by the Nazi Propaganda Office in 1944, Jewish people are described as parasites, mistletoe, and an infection but also as deadly foes, slanderers and desirous of world domination.

As Manne (2016) and Appiah (2008) have pointed out, even in the act of referring to a group as lice or vermin propagandists reveal an implicit recognition of the difference between their targets and the animals to which they are compared. There is no sense in consistently reminding a rat that it is, in fact, a rat (Manne, 2016). Rather, the power of the metaphor comes from the recognition that the two entities are different (Bloom, 2017). If victims of intergroup harm are not viewed as less than human even in these supposedly prototypical cases, then the onus is on proponents of the dehumanisation hypothesis to explain when outgroups are dehumanised.

Challenge 3: Being associated with a non-human entity is not equivalent to being seen as similar to that non-human entity

Psychologists have typically relied more heavily on evidence from lab-based research than on evidence from the historical record. For example, data showing that White Americans implicitly associate African Americans with apes in lab-based tasks is often cited as an example of dehumanisation (Goff et al., 2008; 2014). While these implicit associations are clearly deeply problematic they are not convincing evidence that African Americans are perceived as less human than are White Americans or as akin to apes. Associative connections between two stimuli do not necessarily suggest that they will be viewed as equivalent or

as similar to each other. To borrow an example from the animal learning literature, a rat may come to associate a grey square with food through repeated presentation of the square in conjunction with food (Rescorla & Wagner, 1972). However, the rat does not come to think of the grey square as being food or similar to food (and would not try to eat it). Relatedly, cultural pairings may lead to associations between African Americans and apes, which reveal themselves in non-verbal tasks in the lab but that is not, on its own, evidence that participants in those studies viewed African Americans as similar to apes.

Challenge 4: Outgroup members may be denied some mental states but they are frequently attributed others

In some theories, dehumanisation is characterised as a "failure to consider the mind of another person" (Harris & Fiske, 2011; see also Hackel, et al., 2014; Harris & Fiske, 2006; 2001; Leyens et al., 2007; Rai, et al., 2017; Waytz, Gray, Epley, & Wegner, 2010). In the seminal study advocating for this view, Harris and Fiske (2006) provided evidence that, when presented with images of homeless individuals and drug addicts, participants display less activity in brain regions associated with mentalising.

This characterisation of dehumanisation does not appear to accurately characterise how outgroup members are perceived. There are certainly situations in which the thoughts and emotions of outgroup members are not sufficiently salient to dominant majorities (Harris & Fiske, 2011). However, I predict that outgroup members are not thought to lack all mental states or even to hold mental states to a lesser extent. While outgroup members may be denied some mental states, they are likely attributed others (Hackel, et al., 2014). To analyse Harris and Fiske's own example a little more closely, drug addicts are often criticised *because* of how observers represent their mental states - they are perceived as greedy and lacking in self control (Corrigan, Kuwabara, & O'Shaughnessy, 2009). Further evidence that this is the case can be drawn from historical record. Even in supposedly prototypical examples of blatant dehumanisation, target groups are described in terms of mental states. Propaganda often references the mental states of its targets in order to generate hatred against them. In Nazi propaganda, for example, Jewish people were often

attacked for their supposed cunning, malice and scheming against the Nazi regime (Bywerk. 2004; Keen 1992; Kershaw, 2001).

Furthermore, it is not clear why possessing certain mental states to a lesser extent should be conceptualised as dehumanisation. Equally problematic for the dehumanisation hypothesis are cases where ingroup members are denied some mental states or afforded them to a lesser extent. For example, ingroup members are typically judged to be less deceptive, sneaky and ruthless than are outgroup members (Dunham, 2018). We are thus left with a peculiar situation in which biases in mental state attribution are sometimes characterised as dehumanisation and sometimes not.

Challenge 5: Outgroup members are granted some uniquely human attributes

Haslam and colleagues argue that groups can be dehumanised in one of two respects. They can fall victim to animalistic dehumanisation in which they are thought to possess attributes like civility, refinement and rationality to a lesser extent than do the ingroup. Alternatively, they can fall victim to mechanistic dehumanisation, in which they are thought to possess attributes like emotional responsiveness, interpersonal warmth and agency to a lesser extent than do the ingroup (Haslam et al., 2005; Haslam, 2006; Haslam & Loughnan, 2014).

On closer inspection, Haslam and colleagues' characterisation of dehumanisation appears incomplete. In particular, it omits anti-social attributes like jealously, spite, dishonesty and disloyalty. These anti-social attributes are not salient in Haslam's characterisation of the concept 'human' and yet they only make sense when applied to humans. It would be extremely unusual, and most likely inappropriate, to describe an animal or a machine as disloyal or spiteful, for example (Manne, 2016; 2018).

I propose that this problem arises because of the way in which Haslam and colleagues sought to identify participants' lay concept of humanness. Haslam and colleagues assumed that they could characterise the lay concept of humanness by asking participants two questions: which from among a list of attributes they felt distinguished humans from other species and which attributes they thought were characteristic of humans. This approach is at odds with decades of

research in the cognitive psychology of categorisation (Barsalou, 1989; Bellezza, 1984; McNamara & Sternberg, 1983; Medin, 1989). Research in this area demonstrates that the attributes that appear typical of a category will vary depending on the context (Smith & Medin, 1981). In this case, the attributes that appear typical of the category human will differ depending on the comparison point. As a thought experiment, imagine that instead of asking his participants what distinguishes humans from other species, Haslam and colleagues had asked their participants what distinguishes humans from another non-human category - angels (see Figure 2). I predict that more anti-social qualities such as greed, jealousy, laziness and spite would have been listed as typical of humans in this context. Somewhat different attributes again would be salient to participants if the comparison point was zombies, robots, rats or dolphins. It is no coincidence that Haslam et al. asked their participants two questions, and found evidence for two characterisations of the concept human. Had they asked their participants a third question, for example how humans differ from angels, they may well have found evidence for three forms. This leaves us with a puzzle whereby being perceived as somewhat lacking civility and warmth is thought to constitute dehumanisation, whereas being seen as somewhat lacking in spite and jealousy, also attributes unique to humans, is not thought to constitute dehumanisation.

When we consider these uniquely human but anti-social attributes as part of what it means to be human, it is no longer clear that outgroups are perceived as less human than are ingroups. Outgroup members are often thought to possess anti-social attributes such as cunning and spite. Indeed, they are thought to possess them to a greater extent than are members of the ingroup (Fiske, Cuddy, Glick, & Xu, 2002).

Figure 2. The salience of particular attributes depends on the comparison point. Attributes listed when comparing humans to gorillas will be different from those listed when comparing humans to angels.

Challenge 6: Groups are often persecuted because of their perceived humanity

The dehumanisation hypothesis is based on the assumption that being perceived as human tends to offer protection from harm. In this view, humans naturally empathise with and care for each other (Harris, 2017; Harris & Fiske, 2011; Smith 2011). When a group is perceived to be less human, or less than human, some of these natural responses of care and empathy are thought to be eroded rendering the group more vulnerable to harm (Harris, 2017; Smith, 2011; Haslam, 2006; Haslam & Loughnan, 2014).

This confidence in humans' desire to protect and care for each other is partly misplaced (Lang, 2010). Although humans do often care for each other, certain groups may be targeted *because* they are believed to possess uniquely human mental states and attributes (Bloom 2017; Manne, 2018). It seems relatively uncontroversial to predict that perceiving a group to have human qualities such as spite and cunning will make them appear a threat. A threat that, in extreme cases, will need to be controlled or destroyed (Keen, 1992). Similarly, believing a group exists in human-specific social relations to a dominant majority such as being enemies, traitors, rapists or criminals, will likely make them appear morally responsible for their actions (Brown & Webb, 2007). This is one reason why violence against target groups can take on a moral quality (Rai & Fiske, 2014; Rai et al., 2017). Whereas eradicating vermin might be seen as desirable, the lynching or mass murder of humans can be framed as a moral

crusade (Dray 2003; Keen, 1992; Rai & Fiske, 2014). Consistent with this observation, lynching of African Americans was often presented as exacting justice upon the guilty (Dray, 2003). Similarly, propaganda in Nazi Germany often described the German population as the victim of Jewish aggression and plots (Keen, 1992). In lab-based research, Rai and colleagues (2017) have shown that reflecting on the mental states of a potential target increases morally motivated violence against them.

Related to this, Manne (2018) has pointed out that victims of atrocities are humiliated and tortured *because* their abusers at least implicitly recognise their humanity. Whereas rats might be killed, they are not forced to simulate sex acts with other rats nor are they forced to watch their fellow rats die. Nor would there be any sense in publicly humiliating or murdering a rat in order to serve as a warning to other rats. Yet human victims often suffer these indignities (Bloom, 2017; Brown & Webb, 2007; Dray, 2003). Thus, even in these supposedly prototypically dehumanising actions, perpetrators implicitly acknowledge the difference between their victims and non-human entities in the type of harm that they inflict upon them (Bloom, 2017; Manne, 2016; 2018).

Challenge 7: Being seen as less than human is not necessarily a risk factor for harm

Equally problematic for the dehumanisation hypothesis are cases where individuals are thought to lack characteristically human qualities and yet are not subjected to harm. Consider the example of a baby. As Smith (2011) has observed, babies are not typically considered to have complex thoughts and beliefs. Nor are they thought to possess typically human attributes such as refinement, civility and maturity. Yet, they are treated with the utmost care and devotion by their parents and caregivers. Indeed, treating young children as more mature and rational than they are would most likely be harmful to them (Goff, Jackson, Di Leonie, Culotta, & DiTomasso, 2014).

Even when organisms are clearly viewed as outside of, or indeed beneath, the human category they are not necessarily at risk of harm. Although many animals are treated badly by humans, some are treated with care. People donate substantial funds to conservation efforts focused on animals like the Great Panda

for example. Even allegedly dangerous and disgusting animals such as snakes are protected by conservation efforts. Likewise, family pets, while typically thought less important than humans, are usually treasured and protected by their human families (see Figure 3).

Figure 3. Entities viewed as 'less than human' are sometimes treated with care.

In certain circumstances, animals are actually treated with greater care than are humans. In the UK, the Royal Society for the Prevention of Cruelty to Animals (the RSPCA) receives substantially more donations than Shelter, the best known charity for the homeless (Annual Report, 2017/2017; Trustee's reports and accounts, 2018). On an individual level, Hitler was responsible for the murder of millions of innocent people and yet he adored his pet dogs (Kershaw, 2001). Objects too are sometimes treated with great care. The campaign to save Notre Dame after the 2019 fire raised hundreds of millions of euros (as reported in the Guardian, 2019). These examples suggest that it is not being perceived as less than human per se that puts a group at risk of harm.

Is it possible to explain the prevalence of non-human metaphors in propaganda without recourse to the concept of dehumanisation?

I have outlined seven challenges for the dehumanisation hypothesis. I have argued, that while there are clearly differences in how members of different social groups are perceived, these differences do not seem best characterised as a psychological process of 'dehumanisation'. This leaves us with an open question, however. If outgroup members are not really perceived as less than human, then how are we to explain the apparent prevalence of comparisons to

non-human entities in propaganda and other forms of hate speech? Below I sketch an alternative theoretical account of why comparisons to non-human entities are sometimes associated with harm.

Propagandists use comparisons to certain non-human entities in order to persuade audiences to view their target group in a negative light. By consistently pairing representations of a target group with negatively valenced stimuli such as disease and lice, propagandists encourage observers to negatively evaluate that group. Importantly, this mechanism is not specific to references to sub-human categories. Rather, conditioning can take place by pairing the group with any negative concept including words that evoke human-specific social relations such as enemy, rapist, criminal, and traitor (Siegel & Allan, 1996).

Referring to a target group as certain types of non-human entity also licenses particular inferences about that group and their supposed traits (Tirrell, 2012). For example, by evoking the idea that a target group are like cockroaches, a propagandist might convey the idea that they disgusting and low status. By evoking the idea that a target group are like monkeys, a propagandist might convey the idea that they are stupid and unsophisticated. Metaphors of this type can also license particular behaviours (Tirrell, 2012). For example, whereas comparing a target group to oxen might imply that they ought to be enslaved, comparing a target group to cockroaches might imply that they ought to be exterminated. Again, this mechanism is not specific to subhuman categories. Referring to a target group as 'the enemy' licenses the inference that they are a threat and need to be defeated.

To the extent that comparisons to non-human entities are more common in propaganda and hate speech than they are in other forms of communication, it may be because they are especially effective ways of conveying multiple negative messages simultaneously. For example, whereas using the insult 'enemy' conveys the idea that the target group is a threat, using the insult 'louse' not only conveys that the group is a threat, but also that the group is contaminating, low status and homogenous in character.

Non-human metaphors can also be used to communicate negative messages to the target group themselves - they can be used to humiliate and threaten (Bloom, 2017; Lang, 2010; Manne, 2016). In these cases, it is not the

message that the group is less than human per se that is problematic. Rather, these non-human comparisons are offensive to the extent that they imply the target group is low status, disgusting and threatening. In Western culture, it would be bizarre but not particularly offensive to repeatedly refer to a group as dolphins or pandas (both of which are considered 'less than human'), whereas it would be deeply offensive to refer to a group as rats. The insult 'rat' is offensive because of learned associations between rats, disease, contamination and threat.

It is worth pointing out that comparisons to non-human entities are not always used intentionally to persuade or humiliate. If listeners repeatedly hear a target group compared to negatively valenced animals, like rats or snakes, then they will come to associate that group with those particular entities (Devine, 1989). Once this association is in place, when listeners hear the target group mentioned, images of the associated entity will be automatically activated (Goff, et al., 2008). Automatic associations between outgroups and animals, or indeed other negatively valenced categories, can leak in to communication outside of conscious awareness. For example, an individual who automatically associates a particular group with apes might use terms like 'jungle', 'wild' and 'creature' when describing that group even when they have no intention of exacerbating animosity towards the group in question.

Importantly, comparisons to non-human entities are not always used to criticise, humiliate and demean. They can also be invoked to compliment a group and extol their virtuous intentions and attributes. For example, consistently referring to a group as lions might activate associated inferences about courage and dominance. Referring to a group as angels might emphasise the ways in which their mental states and attributes are worthy of respect and admiration. Again, according to my perspective, references to non-human entities are not the only way to convey these messages, but they are one effective means by which to do so.

Directions for future research

The seven challenges outlined here bring much of what we thought we knew about dehumanisation into question. These challenges also suggest a number of priorities for future research that will enable the field to distinguish between the

dehumanisation hypothesis and the alternative I have outlined here. First, it will be crucial to conduct systematic content analyses of historical documents in order to establish the incidence of non-human comparisons in propaganda and other forms of hate speech. In conducting these content analyses, it will be of paramount importance to avoid the confirmation bias whereby the historical record is only searched for examples that provide positive evidence for the dehumanisation hypothesis (that is, cases where target groups are referred to as non-human entities). The most problematic cases for the dehumanisation hypothesis include passages where target groups are described in terms of uniquely human attributes, mental states and emotions.

Another priority for future research is to re-assess how the construct of dehumanisation is operationalized. Thus far, much psychological research on dehumanisation has focused on the attribution of broadly positive and prosocial qualities such as civility, rationality, warmth, pride and guilt (Haslam, 2006; Leyens et al., 2007). Future research must incorporate insights from context dependent categorisation in order to understand whether adults also associate anti-social attributes such as jealousy, spite, cunning, greed and dishonesty with the concept of human.

If it transpires that participants do associate anti-social attributes such as these with the concept human, then further empirical studies can determine whether outgroup members are thought to possess all uniquely human attributes to a lesser extent (as predicted by the dehumanisation hypothesis) or only those which are prosocial in nature. Furthermore, future work can reassess the hypothesised causal relationship between the attribution of uniquely human qualities and intergroup harm by evaluating whether describing a potential victim in terms of uniquely human but anti-social attributes places them at greater or lesser risk of harm.

Conclusion

According to the dehumanisation hypothesis, when a dominant majority describes an outgroup as ‘rats’ or ‘lice’, they do so because they believe that group to be less than human. This argument has an elegant simplicity and an intuitive appeal. Not surprisingly, therefore, it has become extremely prominent

in social psychology, social neuroscience and certain areas of philosophy (Harris & Fiske, 2006; 2009; Haslam, 2006; Haslam & Loughnan, 2014; Leyens et al., 2007; Smith, 2011; 2014; 2016). Indeed, it is all but accepted as fact that certain groups suffer dehumanisation (Smith, 2016).

I have argued that the explanatory power of this hypothesis is considerably more restricted that it first appears. While there may be some cases in which outgroup members are genuinely believed to be less than human, there is not yet convincing evidence that this is a common phenomenon. Comparisons to non-human entities are not restricted to outgroups, nor are outgroups regularly described in similar ways to non-human entities. While members of outgroups are often perceived to be lacking in some human qualities and attributes, they may well be thought to possess other human attributes to a greater extent than do the ingroup (Bloom, 2018; Lang, 2010; Manne, 2016). Added to this, there are other plausible explanations for why comparisons to non-human entities may be common in propaganda and other forms of hate speech. It remains for future empirical research to determine the relative merits of the dehumanisation hypothesis and the alternative view I have suggested here.

If supported by future empirical research, the argument I have advanced has important implications our understanding of how to bring about social change. Inspired by the dehumanisation hypothesis, researchers have shown an increasing interest in developing interventions to reduce intergroup harm that focus around 'humanising' outgroups (Albarello & Rubini, 2012; Gaunt, 2009). My critique suggests that other routes to reducing intergroup harm may prove more effective. Encouraging dominant majorities to reflect on the humanity of outgroups could even backfire in certain circumstances. For example, if it leads to increased focus on the supposedly anti-social mental states and attributes of the group in question.

References

Albarello, F., & Rubini, M. (2012). Reducing dehumanisation outcomes towards blacks: The role of multiple categorisation and of human identity. *European Journal of Social Psychology, 42,* 875-882. doi: https://doi.org/10.1002/ejsp.1902

Allport, G.W. (1954). *The nature of prejudice.* Reading, Massachutsetts: Addison-Wesley.

Andrighetto, L., Baldissarri, C., Lattanzio, S., Loughnan, S., & Volpato, C. (2014). Human-itarian aid? Two forms of dehumanization and willingness to help after natural disasters. *The British Journal of Social Psychology, 53*, 573-584. doi: 10.1111/bjso.12066

Annual report and, Shelter (2017/2018), Retrieved from https://england.shelter.org.uk/__data/assets/pdf_file/0013/1405030/OBR-1521_Annual_Report_DIGITAL_FINAL.pdf

Apel, D. (2005). Torture culture: Lynching photographs and the images of Abu Ghraib. *Art Journal, 64,* 88-100. doi: https://doi.org/10.1080/00043249.2005.10791174

Appiah, K.A. (2008). *Experiments in ethics.* Cambridge, MA: Harvard University Press.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3*, 193-209. doi: 10.1207/s15327957pspr0303_3

Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education, 31,* 101–119.

Barsalou, L. W. (1989). Intra-concept similarity and its implications for inter-concept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76-121). Cambridge, England: Cambridge University Press.

Bellezza, E S. (1984). Reliability of retrieval from semantic memory: Noun meanings. *Bulletin of the Psychonomic Society, 22,* 377-380.

Bloom, P. (2017). The root of all cruelty? *The New Yorker,* November issue.

Brown, D., & Webb, C. (2007). *Race in the American South: From slavery to civil rights.* Edinburgh, UK: Edinburgh University Press.

Bytwerk, R.L. (2004). *Bending spines. The propagandas of Nazi Germany and the German Democratic Republic.* Michigan, USA: Michigan State University Press.

Bytwerk, R.L. (2006). The argument for genocide in Nazi propaganda. *Quarterly Journal of Speech, 91,* 37-62. doi: https://doi.org/10.1080/00335630500157516

Cathedral fire under control after spire and root destroyed. *The Guardian, 16th April, 2019,* Retrieved from https://www.theguardian.com/world/live/2019/apr/15/notre-dame-cathedral-fire-paris-france-landmark-live-news?page=with:block-5cb50e018f08bc7376aedec3#block-5cb50e018f08bc7376aedec3

Corrigan, P. W., Kuwabara, S. A., & O'Shaughnessy, J. (2009). Public stigma of mental illness and drug addiction. *Journal of Social Work, 9,* 139–147. doi: https://doi.org/10.1177/1468017308101818

Cuddy, A. J. C., Rock, M. S., & Norton, M. I. (2007). Aid in the aftermath of hurricane Katrina: Inferences of secondary emotions and intergroup helping. *Group Processes & Intergroup Relations, 10*, 107-118. doi: 10.1177/1368430207071344

Demoulin, S., Rodriguez, R. T., Rodriguez, A. P., Vaes, J., Paladino, M. P., Gaunt, R., Cortes, B. P., & Leyens, J. P. (2004). Emotional prejudice can lead to infra-humanization. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology, 15,* 259–296.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5-18.

Dray, P. *At the hands of persons unkown: The lynching of Black America.* New York: Modern Library.

Dunham, Y. (2018). Mere membership. *Trends in Cognitive Sciences, 22,* 780-793. https://doi.org/10.1016/j.tics.2018.06.004

Esser, H. (1939). *Die jüdische Weltpest.* Translated by Bytwerk, R. Retrieved from https://research.calvin.edu/german-propaganda-archive/esser.htm, 30th September 2019.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from

perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878-902. doi: http://dx.doi.org/10.1037/0022-3514.82.6.878

Gaunt, R. (2009). Superordinate categorization as a moderator of mutual infrahumanisation. *Group Processes and Intergroup Relations, 12,* 731– 746. DOI: 10.1177/1368430209343297.

Goff, P. A., Eberhardt, J. L., Williams, M. J., & Jackson, M. C. (2008). Not yet human: implicit knowledge, historical dehumanization, and contemporary consequences. *Journal of Personality and Social Psychology, 94*, 292-306. doi: 10.1037/0022-3514.94.2.292

Goff, P.A., Jackson, M.C., Di Leone, B.A.L., Culotta, C.M., DiTomasso, N.A. (2014). The essence of innocence: consequences of dehumanizing black children. *Journal of Personality and Social Psychology, 106,* 526-45. doi: http://dx.doi.org/10.1037/a0035663

Gopnik, A. (2006). Headless horseman: The reign of terror revisited. *The New Yorker,* June issue.

Hackel, L. M., Looser, C. E., & Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: The role of social identity, collective identification, and intergroup threat. *Journal of Experimental Social Psychology, 52*, 15-23. doi: 10.1016/j.jesp.2013.12.001

Harris, L. T. (2017). *Invisible mind: Flexible social cognition and dehumanisation.* Cambridge, MA: MIT Press.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science, 17*, 847-853. doi: 10.1111/j.1467-9280.2006.01793.x

Harris, L. T., & Fiske, S. T. (2009). Social neuroscience evidence for dehumanised perception. *European Review of Social Psychology, 20*, 192-231. doi: 10.1080/10463280902954988

Harris, L. T., & Fiske, S. T. (2011). Dehumanized perception: A psychological means to facilitate atrocities, torture, and genocide? *Zeitschrift für Psychologie/Journal of Psychology, 219*, 175-181. doi: 10.1027/2151-2604/a000065

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10*, 252-264. doi: 10.1207/s15327957pspr1003_4

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. doi: 10.1037/0022-3514.89.6.937

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology, 65*, 399-423. doi: 10.1146/annurev-psych-010213-115045.

Haslam, N., Loughnan, S., & Sun, P. (2011). Beastly: What makes animal metaphors offensive? *Journal of Language and Social Psychology, 30*, 311–325. doi: https://doi.org/10.1177/0261927X11407168

Hatzfeld, J. (2003). *Machete season: The killers in Rwanda speak.* New York: Picador.

Keen, S. (1992). *Faces of the enemy: Reflections of the hostile imagination.* San Fransisco: Harper.

Kershaw, I. (2001). *Hitler 1936-1945: Nemesis.* Norton and Company, New York.

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,* 487-494. doi:https://doi.org/10.1177/0963721417708230

Kteily, N. S., 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*, 901-931. doi:http://dx.doi.org/10.1037/pspp0000048

Lang, J. (2010). Questioning dehuumanisation: Intrsubujective dimensions of violence in the Nazi concentration and death camps. *Holocaust and Genocide Studies, 24,* 225-246.

Leyens, J. P., Demoulin, S., Vaes, J., Gaunt, R., & Paladino, M. P. (2007). Infra-humanization: The wall of group differences. *Social Issues and Policy Review, 1*, 139-172. doi: 10.1111/j.1751-2409.2007.00006.x

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

Manne, K. (2016). Humanism: A critique. *Social Theory and Practice, 42,* 389-415. doi: 10.5840/soctheorpract201642221

McNamara, T. P., & Sternberg, R. J. (1983). Mental models of word meaning. *Journal of Verbal Learning and Verbal Behavior, 22*, 449- 474.

Medin, D. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469-1481.

Nickerson, R.S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 157-220.

Rai, T.S., & Fiske, A.P. (2014). *Virtuous violence.* Cambridge, UK: Cambridge University Press.

Rai, T.S., Valdesolo, P., & Graham, S. (2017). Dehumanisation increases instrumental violence, but not moral violence. *Proceeedings of the National Academy of Sciences, 114,* 8511-8516. doi: https://doi.org/10.1073/pnas.1705238114

Rankin. H.F. (1954). The naval flag of the American Revolution. *The William and Mary Quarterly, 11,* 339-353.

Reichspropagandaleitung (1944). Den Juden kennen heißt den Sinn des Krieges verstehen. Translated by Bytwerk, R. Retrieved from https://research.calvin.edu/german-propaganda-archive/sprech44a.htm, 30th September, 2019.

Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement, in A.H. Black & W.F. Prokasy (Eds) *Classical Conditioning II: Current research and theory*. pp. 64–99. New York: Appleton-Century-Crofts.

Siegel, S., & Allan, L. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin and Review, 3,* 314-321.

Smith, D.L. (2011). *Less than human: Why we demean, enslave, and exterminate others.* New York: Macmillan.

Smith, D.L. (2014). Dehumanization, essentialism, and moral psychology. *Philosophy Compass, 9*, 814–824. doi: https://doi.org/10.1111/phc3.12174

Smith, D.L. (2016). Paradoxes of dehumanisation. *Social Theory and Practice, 42,* 416-443. doi: 10.5840/soctheorpract201642222

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Tirrell, L. (2012). Genocidal language games. Om *Speech and harm: Controversies over free spech.* I. Maitra & M.K. McGowan (Eds), Oxford University Press, NY. pp.174-221.

Trustees reports' and accounts (2018), Retrieved from https://yougov.co.uk/ratings/politics/popularity/charities-organisations/all

Vaes, J., Leyens, J. P., Paladino, M. P., & Miranda, M. P. (2012). We are human, they are not: Driving forces behind outgroup dehumanisation and the humanisation of the ingroup. *European Review of Social Psychology, 23*, 64-106. doi: 10.1080/10463283.2012.665250

Viki, G. T., Osgood, D., & Phillips, S. (2013). Dehumanization and self-reported proclivity to torture prisoners of war. *Journal of Experimental Social Psychology, 49*, 325-328. doi: 10.1016/j.jesp.2012.11.00

Waytz, A., Gray, K., Epley, N., & Wegner, D. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences, 14*, 383–388. doi: 10.1016/j.tics.2010.05.006