Rethinking Models of Subtle Dehumanization in Social Psychology

Robert Brennan

PhD

University of York

Psychology

March 2024

Abstract

Dehumanization research has grown in recent decades, often examining contexts of intergroup attitudes and harm. Several influential models of subtle or "everyday" dehumanization have been at the forefront of these developments. One such model, infrahumanization theory, suggests that we subtly dehumanize outgroup members by seeing them as experiencing uniquely human emotions (e.g., optimism, nostalgia) to a lesser extent than ingroup members. The dual model of dehumanization suggests we subtly dehumanize human targets by denying them character traits that distinguish humans from other animals (e.g., sophisticated, civilised) or machines (e.g. interpersonally warm, openminded). However, these models have recently been critiqued, particularly for depending on the perception of prosocial aspects of humanity and overlooking antisocial aspects. Thus, we cannot disentangle dehumanization from intergroup preference – associating ingroup members with more prosocial characteristics than outgroup members. This thesis re-examines models of subtle dehumanization and associated findings considering these critiques. The hypothesised causal link between trait-based dehumanization and harm is examined in Chapter 2, demonstrating harm to be predicted by attributing undesirable traits to the target rather than denying uniquely human traits. Infrahumanization as a post hoc justification for harm is explored in Chapter 3. Rather than the pattern infrahumanization theory predicts, harmed outgroup members were seen as feeling negative emotions when harmed, regardless of emotion humanness or sense of collective responsibility. Chapter 4 re-examines infrahumanization with more robust methods than typically used, finding intergroup preference better explains apparent evidence of infrahumanization. These findings support recent critiques, with no evidence of trait or emotion-based dehumanization found across 11 highly powered studies (N_{total} = 1,748). Implications of the confounds and replicability issues for future dehumanization research are discussed. This thesis provides a strong challenge for current conceptualisations of subtle dehumanization, emphasising the importance of robust theories and replicable findings concerning urgent social issues like intergroup harm.

List of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to thank Harriet Over for the opportunity to pursue this PhD and for her excellent mentorship over the last three years. I am grateful to Harriet for her insight, guidance, and encouragement, and I know her professionalism and strive for impactful research that addresses urgent social issues will continue to inspire me long after my time in the Social Origins lab group. I also thank Florence Enock for her support, guidance, and help throughout my PhD studies, especially in developing my R and experimental design skills. From the interview for my PhD position to submitting this thesis, it has been a pleasure working with Harriet and Florence, and I will miss our friendly discussions and shared value-driven approach to research. I would also like to thank everyone else in the Social Origins lab, particularly Adam Eggleston, Juliana Olivier, and Carl Bunce, for their support and encouragement. I want to thank each of my thesis advisors, Cade McCall and Bailey House, and my progression chair Aiden Horner, and Katie Slocombe for their valuable insight and feedback throughout my PhD.

A heartfelt thank you to all the friends I have made during my time in York: Akul, Alex, Anastasia, Ariana, Ben, Cameron, Charlotte K., Charlotte S-K, Chiu, Connie, Dan, Emily, Fran, Geri, Harrison, Jonathan, Mari, Mirela, Nick, Ursula, Vanessa, little Stripey, and my peers in the Department of Psychology. Whether feeding the many birds of the uni campus, virtual board games during lockdown, impromptu séances, moderately successful pub quizzes, survibing, gigs at Brudenell, or just one more, my time in York will be missed. Thank you to my friends from home and around the world who have been cheering me on over the last three years: Anastasia, Edit, Fanni, Genta, Hannah, Harry, Izzy, Johner, Jonathan, Karina, Katalin, Kira, Lete, Líbhan, Liliana, Lotan, Lucy, Mari, Nuran, Phil, Shane, And Tainara. A special thank you to Lucy for offering to proofread this thesis. Thank you also to my mentor-turned-friend, Anca Minescu, and everyone else at the Migration and Diversity Research Lab in Limerick and the Global MINDS community for their support.

Of course, I owe a huge thank you to my mother, Teresa, and my father, Conor, for their continued support of every kind. I hope to make each of them proud. I also thank my

family on both sides, as well as my siblings Niall, Leah, and Shane for their support (primarily half-listening to me ramble), and our floofy queens Mayah and Ziggy for their eager company on many head-clearing walkies.

Lastly, I could not have completed this PhD, nor many other things, without the love, support, and compromises made by my partner Árpi. Thank you, and I hope to be the pillar of strength you were for me during this PhD as you complete yours.

Author's Declaration

I, Robert Brennan, declare that this thesis is a presentation of original work, and I am the sole author, under the supervision of Professor Harriet Over (primary supervisor) and Dr Florence Enock. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. All research reported in this thesis was supported by the European Research Council under the European Union's Horizon 2020 Programme, grant number ERC- STG-755719 awarded to Harriet Over.

Robert Brennan wrote the complete original draft of this thesis. Harriet Over and Florence Enock each provided feedback on the drafts of all chapters.

Chapter 2: The empirical work presented in Chapter 2 was under review for publication in a peer-reviewed journal at the time of thesis submission:

Brennan, R.A., Enock, F.E., & Over, H. (2024). *Attribution of undesirable character traits, rather than trait-based dehumanization, predicts punishment decisions*. [Manuscript submitted for publication]. Department of Psychology, University of York.

All authors conceptualised and designed the experiments. Robert Brennan curated the data and conducted analyses with the support of Florence Enock and under the supervision of Harriet Over. Robert Brennan drafted the full original manuscript, and all authors reviewed and edited the final manuscript before submission.

Chapters 3 and 4: The empirical works presented in Chapters 3 and 4 of this thesis are each a collaborative project of Robert Brennan, Florence Enock, and Harriet Over. Robert Brennan, Florence Enock and Harriet Over conceptualised and designed all experiments. Robert Brennan curated the data and conducted all analyses with the support of Florence Enock and under the supervision of Harriet Over. Robert Brennan wrote the original draft of each chapter. Harriet Over and Florence Enock each provided feedback on the draft of each chapter.

Chapter 1: General Introduction

The horrors of war and mass genocide in the 1940s ushered in a reckoning, as Western philosophy and psychological research struggled to explain the extent of harm caused by what Reynolds and colleagues (2012) refer to as a “fatal flaw” in the human psyche: prejudice. The motivations behind humankind inflicting such harm on itself based on perceived differences in identity became a defining focus of social psychology (Farr, 1996; Gaines & Reed, 1995). As intergroup matters such as racism, colonialism, genocide, and the resurgence of right-wing populism continue to define today’s social world, understanding the facets of prejudice remains as imperative as ever to prevent its manifestation at both the group and individual levels.

Frequently associated with large-scale atrocities such as genocide, marginalisation, war, and slavery (Bar-Tal, 2000; Boucher, 2019; Kelman, 1973; Savage, 2013; Smith, 2011, 2020; Stanton, 1998), as well as individual-level discrimination such as misogyny (Fredrickson & Roberts, 2016; Manne, 2018; Rudman & Mescher, 2012), and hate crime (Firinci, 2018; Hodwitz & Massingale, 2021; Posselt, 2017), is the concept of *dehumanization*. The 21st century brought a particular focus on the construct of dehumanization in social psychology. Different theoretical frameworks attempted to clarify how dehumanization occurs, its socio-behavioural outcomes, and the functions it appears to serve. Whether it is to see another individual as less than human in an ontological sense (Kteily et al., 2015), as human but with lesser mental and socio-emotional capacity than oneself (Leyens et al., 2000, 2001; Waytz et al., 2010), or as something superficially human yet lacking essentially human aspects simultaneously (Haslam, 2006; Smith, 2011) is a debate that continues across many disciplines, particularly philosophy, psychology, political science, and sociology (Kronfeldner, 2021).

In this introduction chapter, I provide a broad overview of relevant social-psychological literature on dehumanization. Before focusing on dehumanization, I give an overview of the closely related constructs of prejudice, stereotyping, and discrimination in intergroup contexts.

I then summarise the most prominent social-psychological theories of dehumanization. Following this, I review the literature on an apparent link between dehumanization and harm. Critiques of current literature on dehumanization will then be discussed, followed by a review of evidence-based alternatives to dehumanization. I then outline the aims of this thesis and give a brief outline of my research reported in subsequent chapters.

1.1. Defining Constructs Related to Dehumanization

Several constructs in social psychology are often associated with dehumanization, in which the overlap can vary considerably. In this section, I briefly summarise related constructs and distinguish them from dehumanization.

Prejudice

Prejudice is a preconceived attitude that creates or maintains hierarchical status relations between groups (Dovidio et al., 2010). Rather than an individual being prejudiced or not, prejudice is better understood as existing along a continuum in which it can vary from implicit forms, such as maintaining social distance from a particular group (Williams, 1964), to explicit forms, such as openly expressing dislike (see also Billig, 2012). The natural, cognitive process of social categorisation, dividing the world into social groupings of “them and us”, is widely seen as the starting point in developing prejudicial beliefs (Kurzban et al., 2001; Tajfel & Wilkes, 1963; Waxman, 2010). Tajfel and Turner’s Social Identity Theory (1979) suggests that individuals make intergroup comparisons favouring their ingroup, incidentally encouraging the perceived inferiority of particular outgroups (see also Reicher et al., 2008). Brewer (1999) distinguishes between two aspects of this bias: ingroup love and outgroup hate. Contextual influences such as past experiences, intergroup contact, and socio-political intergroup relations are all seen as influences on one’s prejudicial attitudes and the norms around prejudice expression (see Durrheim et al., 2016; Gaines, 2012).

Stereotypes

Stereotypes are considered the cognitive component of prejudicial attitudes, defined by Dovidio et al. (2010) as subjectively positive or negative beliefs and associations about a

social group. Stereotypes are schematically formed to help us manoeuvre the social world around us by providing cognitive shortcuts. The content of a stereotype is primarily shaped by one's lived experiences, social interactions, and the socialisation process, whereby proximate ingroup members and institutions influence the presumptions, expectations, and associations one holds about their ingroup and outgroups (Brewer, 1999; Hamilton, 1981; see also: Bandura, 1999). The power of a negative stereotype in eliciting and sustaining prejudicial attitudes and social inequality has much empirical support. For example, ethnic or racial stereotypes have been linked to dire social problems such as shooter bias (Correll et al., 2007), and gender stereotypes are linked to many socio-economic and civil issues that affect all genders, but particularly women (for a review, see: Ellemers, 2018) and LGBTQ+ individuals (see: Seelau & Seelau, 2005).

Banaji et al. (1993) concluded that the mere awareness of a negative stereotype's content could prompt implicit, prejudicial-like associations of an outgroup regardless of one's agreement with the content. However, Devine (1989) stresses that this duality of stereotypes and personal beliefs allows the latter to mitigate the application of stereotypes in the form of prejudice.

Discrimination

Discrimination occurs when an outgroup stereotype is applied, and a prejudicial attitude is expressed or acted upon, which at the individual level can involve the poor treatment of someone due to their group membership. Dovidio and colleagues (2010) note that discrimination creates, maintains, and reinforces advantages for some groups over others, usually one's ingroup. The interplay of stereotypes, prejudice, and discrimination is considered cyclical, in which each component influences the others (see: Harding et al., 1969). The constructs or stages of this detrimental cognitive, affective, and behavioural interplay are linked to dehumanization across many contexts in the literature. As discrimination often involves behaviour involving the maltreatment, negligence, or harm of outgroup members, dehumanization is often applied in attempts to explain how such discrimination can occur (see Haslam & Loughnan, 2015; Kelman, 1973; Smith, 2011, 2020).

1.2. Accounts of Subtle Dehumanization in Social Psychology

Complex in its operationalisation due to its many definitions and contextual applications, dehumanization can be broadly defined as the perception, portrayal, or treatment of a human being as something that is not quite human (de Ruiter, 2021). Overt forms of dehumanization often involve the portrayal of an outgroup as animal-like or as an object through visual mediums and metaphorical language. Blatant or explicit dehumanization remains a central characterisation of the construct when discussing historical examples of dehumanization, such as the Jewish population being depicted as rodent-like in Nazi Germany or the Tutsi being referred to as cockroaches in the lead-up to the Rwandan genocide in 1994 (Smith, 2011; Tirrell, 2012). Kteily et al.'s (2015) 'Ascent of Human' measure (see Figure 1) was developed to capture one's supposed blatant dehumanization of an outgroup, which Kteily & Bruneau (2017) argue is particularly problematic due to it being a conscious denial of full humanity to other humans.

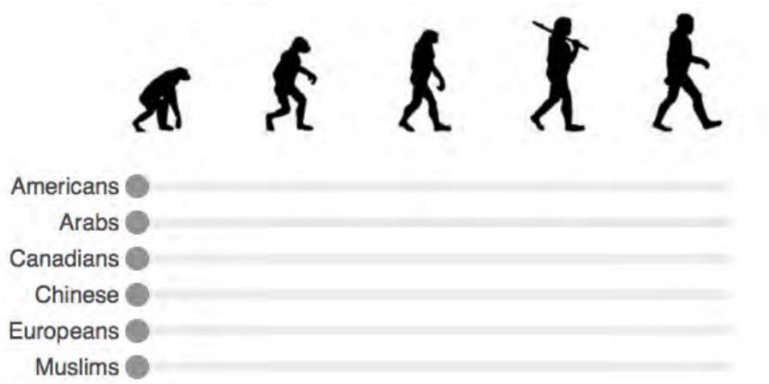

Figure 1. The 'Ascent of Human' measure of explicit dehumanization

Participants are asked to indicate how "evolved" they consider the average member of each group to be using the respective sliding scales. Responses range from 0 (least evolved) on the extreme left to 100 (most evolved) on the extreme right.

Note: Target groups used by Kteily et al. (2015) are displayed (copyright 2015 American Psychological Association).

Other forms of dehumanization in social psychology include perceived deficiencies someone has compared to an “ideal human” (Kteily & Landry, 2022) and perceptual dehumanization, defined as the activation of neural networks associated with the perception of nonhuman animals and objects when viewing an image of another person (Deska et al., 2018; Fincher & Tetlock, 2016). However, this thesis will not focus on these operationalisations of dehumanization. My doctoral work focused on *subtle* forms of dehumanization, which are those thought to occur in everyday perceptions people have of outgroup members without necessarily being conscious of this perception. In the coming section, I discuss three prominent characterisations of subtle dehumanization, which have garnered considerable attention in the field: infrahumanization theory, the dual model of dehumanization, and the mental state attribution account. Each of these models of subtle dehumanization share similar theoretical and methodological frameworks, and each provides a perspective on how perpetrators of harm perceive their victims.

Infrahumanization Theory

The first characterisation of subtle dehumanization I will discuss is infrahumanization theory. Initially proposed by Leyens and colleagues (2000, 2001), infrahumanization theory offers an operationalisation of dehumanization that emphasises variations in one’s perceived capacity of outgroup members to experience uniquely human emotions compared to ingroup members. Specifically, Leyens and colleagues suggested that the perceived humanness of outgroup members can be measured by examining the extent to which they can experience uniquely human emotions relative to ingroup members. Secondary emotions are those thought to be uniquely human, which we assume nonhuman animals do not experience, such as compassion, remorse, or optimism. Compared to *primary* emotions, those we share with nonhuman animals, such as joy, fear, and sadness, uniquely human emotions are more complex, internally sourced, less visible, and longer lasting (Demoulin et al., 2004; Ekman, 2008). Both kinds of emotions can be positively or negatively valenced, meaning they can be positive to experience (e.g., happiness, love) or negative to experience (e.g., loneliness, guilt). Secondary emotions are considered those that capture the “essence” of what it means to be

human relative to nonhuman animals. A proposed indicator of subtle dehumanization, Leyens et al. (2000) initially noted that individuals tend to perceive ingroup members as experiencing uniquely human emotions to a greater extent than outgroup members while attributing primary emotions equally.

This reservation against recognising uniquely human emotions in outgroup members to the same degree as ingroup members is at the centre of infrahumanization theory, suggesting the perception of an incomplete or *infrahuman* essence (Leyens et al., 2007; Leyens et al., 2003; Vaes et al., 2012). Relating to similar ingroup biases in social identity research (Hewstone & Ward, 1985; Turner & Tajfel, 1979), it has been suggested that individuals might view their ingroup as possessing more human essence than other groups (Cortes et al., 2005). Examined initially in the context of cross-regional ethnic groups (Canary Islanders and mainland Spaniards, Leyens et al., 2001), infrahumanization has since been examined in many intergroup contexts, including religious, ethnic, and regional differences in identity among others (Banton et al., 2020; Demoulin et al., 2009; Prati et al., 2016; for reviews see Haslam & Loughnan, 2014; Haslam & Stratemeyer, 2016).

The Dual Model of Dehumanization

First proposed by Haslam (2006), the dual model of dehumanization is built upon prior conjectures of infrahumanization research, particularly the causal nature of subtle dehumanization and the concept of humanness as a dimension of group perception (Haslam, 2021). The dual model is anchored in prior research on essentialist beliefs about what it means to be human in a descriptive sense (Haslam et al., 2004, 2005). At its core, the dual model relies on the assumption that to deny essentially human character traits to a person or group of people indicates a subtle form of dehumanization. The dual model also differentiates between two “senses” of what it means to be human and proposes two associated forms of dehumanization, as illustrated in Figure 2. The first sense of humanness is *human uniqueness*, consisting of traits humans possess that supposedly separate us from other species in the animal kingdom, such as civility, refinement, rationality, moral sensibility, and maturity. According to the dual model, when a target is denied the attribution of these uniquely human

traits, they are targets of *animalistic dehumanization*. They thus are perceived as uncultured, amoral, coarse, immature, or irrational. The second sense of humanness is *human nature*, composed of traits that supposedly separate us from non-sentient machines or automata, such as emotional responsiveness, interpersonal warmth, depth, cognitive openness, and individual agency. The dual model defines the denial of these traits being attributed to a human target as *mechanistic dehumanization*, whereby they are perceived as cold, superficial, passive, rigid, or emotionally inert.

The dual model and its differentiation between animalistic and mechanistic forms of dehumanization has received considerable empirical attention (Bain et al., 2009, 2012; Haslam & Loughnan, 2014). Subtle forms of intergroup dehumanization have supposedly been found using the dual model approach in an occupational context by Loughnan and Haslam (2007). Participants in their implicit association test were argued as animalistically dehumanizing artists and mechanistically dehumanizing businesspeople. When applying the dual model, claims of subtle dehumanization also exist when examining different ethnic and cultural groups. For example, Bain et al. (2009) concluded that ethnic Chinese people tended to be mechanistically dehumanised by Anglo-Australians, who themselves tended to be animalistically dehumanized by ethnic Chinese people. Regarding ethnonational identity, Andrighetto and colleagues (2014) determined that Italian nationals' reduced intention to support disaster relief for Haitian and Japanese victims of earthquakes was explained by the animalistic dehumanization of the former and the mechanistic dehumanization of the latter.

The consequences of being the target of trait-based dehumanization differ depending on whether it is animalistic or mechanistic dehumanization. The dual model (Haslam, 2006) claims that animalistic dehumanization promotes disgust and contempt towards the target, with their behaviour being rooted in desires and wants rather than cognitive states. Mechanistic dehumanization is said to elicit a lack of empathy and indifference in the perceiver towards the target, with the behaviour of the target tending to be explained in nonintentional, causal terms (p. 262).

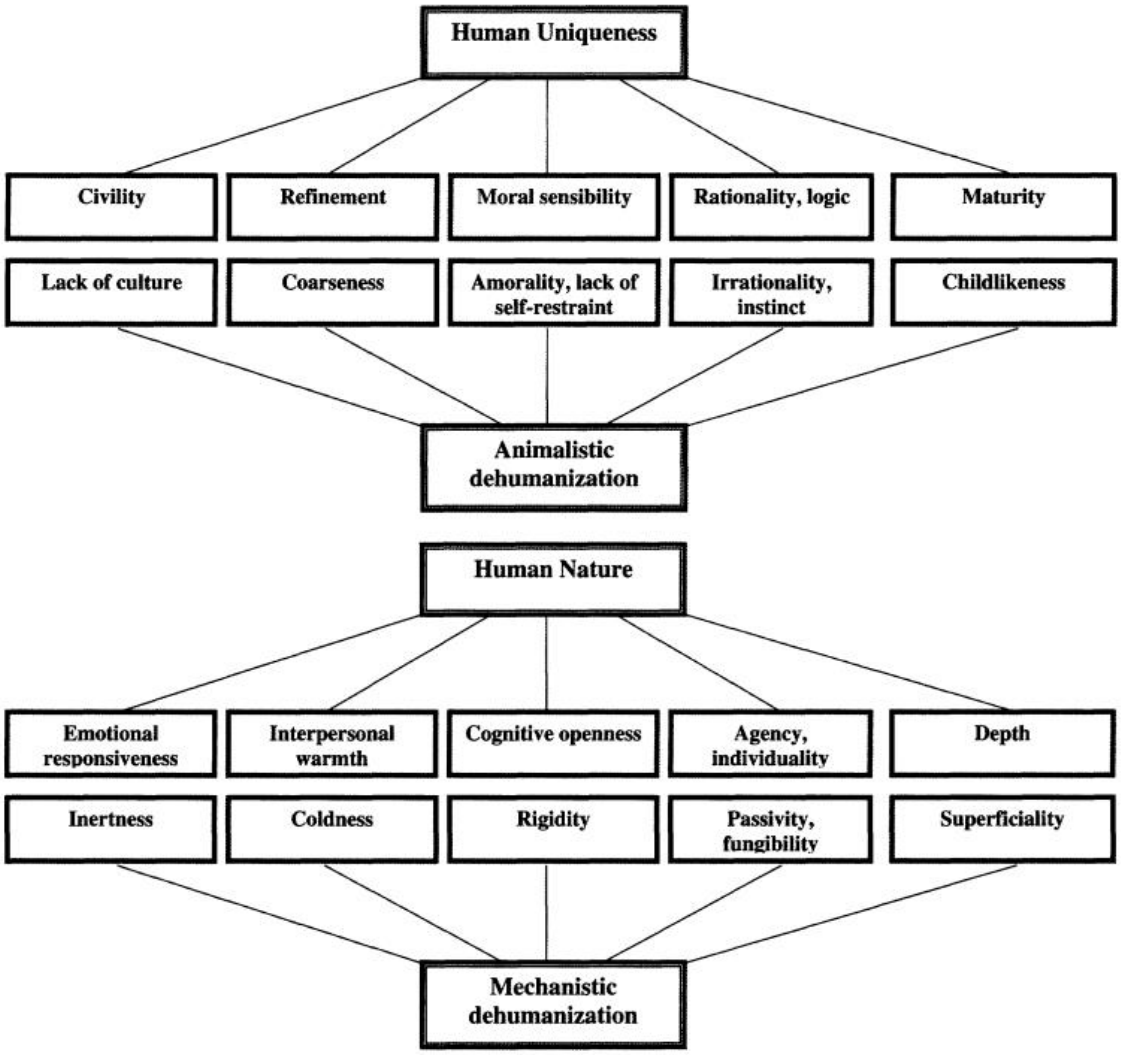

Figure 2. Conceptual illustration of the dual model of dehumanization

The figure is taken directly from the original publication (Haslam, 2006; Copyright 2006, Lawrence Erlbaum Associates, Inc)

Mental State Attribution

Another prominent framework of subtle dehumanization is the mental state attribution account. While this account will not be a focus of this thesis, it is worth noting given its impact on the field and its conceptual overlap with the dual model (Haslam, 2006) and infrahumanization theory (Leyens et al., 2000, 2001). Like these other models, the mental state attribution account characterises dehumanization as failing to perceive aspects of humanity in another individual or social group that distinguish us from nonhuman animals and

machines. Rooted in the subfields of social cognition and social neuroscience, this account defines dehumanization as misperceiving other individuals or members of outgroups as lacking socio-emotional capacity, agency, and subjective experience. More specifically, they are seen as lacking desires, beliefs, or higher-order cognition (Harris & Fiske, 2006b, 2011). This perspective comprises two key research areas (as cited by Haslam, 2021). The first concerns mind perception and is grounded in research on *mentalizing*, or the theory of mind (see Frith & Frith, 2003). In an investigation into lay understandings of how the human mind can vary from individual to individual, Gray, Gray, and Wegner (2007) proposed that mind perception occurs along the dimensions of *agency* (the mental capacity to plan and decide) and *experience* (the capacity to sense and feel). Gray and colleagues found that adults were seen as having greater agency than children or nonhuman animals and greater experience than inanimate objects (see also Epley et al., 2007).

Elaborating on the process of mind perception, Waytz et al. (2010) note that the perceiver could be motivated to recognise mental states in another human out of a desire for social connection. However, when an individual has many social connections, an indifference towards unacquainted others might lead to a dehumanized perception of them (see Waytz & Epley, 2012). Regarding factors stemming from the perceived minds of others, we are said to consider those more similar to us (i.e., ingroup members) as having more mental states. Consequentially, we are said to be more likely to dehumanize (or *dementalize*) outgroup members with whom we do not share an identity. Waytz and colleagues (2010) further note that either recognising mental states or having a dehumanizing perception of the target is associated with varying behavioural outcomes. For example, should the perceived individual be an acquaintance or a member of an ingroup, cooperation and helping behaviour may be encouraged. However, should the perceived be unfamiliar, different to us or unlikeable, then what might be encouraged instead are indifference, hostility, or harm.

Lastly, I will briefly touch upon another variant of the mental state attribution approach that is the least relevant to my doctoral research. Grounded in the stereotype content model (Fiske et al., 2002), this alternative approach focuses on intergroup social perceptions and

associated neural activity. Dehumanization became the prime focus of this approach after Harris & Fiske's (2006) impactful study that claimed to identify a neural signature for dehumanization. They found that the medial prefrontal cortex was activated using functional MRI methods when participants viewed images of all social groups apart from low-status outgroups, such as people who abuse illegal drugs and unhoused people. Moreover, the researchers interpreted amygdala activation when viewing images of members of these groups as disgust or a fear response. It is worth noting, however, that these findings have been criticised for possible reverse inference issues and that the competence dimension highly correlates with social status (Durante et al., 2013). Nonetheless, the influence of this socio-physiological characterisation of dehumanization has been far-reaching, having been applied to several contexts and social groups alongside the stereotype content model's broader applications (for a review, see Fiske, 2018).

1.3. The Hypothesised Causal Link Between Dehumanization and Harm

Across disciplines, dehumanized individuals and groups are considered at a greater risk of harm (Haslam, 2019, 2021; Haslam & Loughnan, 2014, 2016; Smith, 2011, 2016). Dehumanization is postulated as a critical factor in large-scale acts of genocide (see Stanton, 1998) and interpersonal acts of harm committed by regular people in society (see Zimbardo, 2007). The hypothesised link between dehumanization and harm has been particularly prominent when addressing contexts of inter-ethnic perceptions, conflict, and violence (Bar-Tal, 2000; Maoz & McCauley, 2008; Smith, 2011, 2014). Haslam & Loughnan (2014) noted that the construct of dehumanization is primarily called upon to explain and enable acts of violence (p. 402), also claiming that understanding how dehumanization promotes harm has been the driving force behind recent developments in social psychology concerning the phenomenon (Haslam & Loughnan, 2016).

Within social psychology, theories concerning how dehumanization might encourage harm via moral disengagement with the target were suggested prior to developments concerning subtle forms of dehumanization. Opotow (1990) suggested that a dehumanized

individual is deindividuated and no longer perceived as within the moral community of the given social context, and thus, considerations of fairness do not apply to them. Moral exclusion has been associated with diminished empathy towards the target, leaving the dehumanizer in a position to morally disengage from them, which Bandura (1999) notes can present a temptation to cause them harm as it may seem a morally warranted act. This link between dehumanization and harm was further elaborated upon by Nariman et al. (2020), who reported that the moral exclusion of a minority group could encourage other prejudicial attitudes, such as enhancing the positive relationship between social dominance orientation and right-wing authoritarianism with discriminatory intentions (see also Duckitt, 2001).

In recent decades, a causal link between dehumanization and increased intentions to harm the target has been examined across many intergroup contexts. In the following paragraphs, I will discuss how explicit forms of dehumanization, followed by characterisations of subtle forms of dehumanization, have each been empirically linked to harming the target.

Explicit Dehumanization and Harm

Philosophers have noted that genocide is often preceded by propaganda in which the members of the target outgroup are explicitly described as rats, lice, and parasites (Smith, 2011; Tirrell, 2012). Such blatant forms of dehumanization have also been linked to increased hostility in lab-based work. Kteily and Bruneau (2017) reported that the extent to which US Americans endorse the claim that Arabs are “less than human” predicted hostility towards Arab immigration. The 'Ascent of Human' measure of blatant dehumanization (Kteily et al., 2015) has been utilised in examining many inter-group contexts where links to harm have been suggested. For instance, Kteily and colleagues (2015) showed that the extent to which non-Arab US Americans blatantly dehumanized people who were Arab was associated with reduced empathy and the endorsement of direct harm towards them (see also Kteily et al., 2017). Similarly, Viki et al. (2013) found that British Christians who associated Muslims with animal traits more than human traits showed greater support for torturing Muslim prisoners of war. In examining the predictors of explicit dehumanization, Lindén et al. (2016) suggested

majority ingroup identification and authoritarian and socially dominant personalities encourage the blatant dehumanization of terrorists, which predicted support for torturing them.

Empirical work has also suggested that nonviolent harm can be predicted by explicit dehumanization. For instance, Bruneau et al. (2020) found that blatant dehumanization predicted discrimination against minority Roma children by majority-group teachers in a Hungarian school, whereby explicit dehumanization of Roma students predicted them being disproportionately recommended for lower educational tracks (see also Civitillo et al., 2022). Turning from explicit dehumanization to implicit mental associations between outgroup members and animals, empirical work on this phenomenon has also suggested dehumanization can lead to increased non-violent forms of harm. In a series of implicit association tests, Goff et al. (2008) concluded that non-Black participants in their US sample tended to associate the faces of Black people with apes and that this association then influenced participants to judge the use of violence against a criminal suspect as more justifiable when the suspect was Black, compared to when the suspect was White. Goff and colleagues' findings suggest how dehumanization, at least In the form of semantic associations between outgroup members and nonhuman animals, appears to be associated with the endorsement of harm.

The following paragraphs will focus on the models of subtle dehumanization previously described and how each posits their characterisation of dehumanization predicts harm towards the target.

Subtle Forms of Dehumanization and Harm

According to the dual model of trait-based dehumanization, the more an individual or group is either animalistically or mechanistically dehumanized, the greater their risk of harm (Haslam, 2006; Haslam & Loughnan, 2014). The dual model's framework has been employed in many empirical studies examining a hypothesised link between trait-based dehumanization and increased hostility towards the target (Barber & Davis, 2022; Chen-Xia et al., 2023; Kasper et al., 2022; Morehouse et al., 2023; Rousseau et al., 2023; West & Thomson, 2023). Those seen as lacking human traits such as civility, warmth, or cognitive openness, for example, are

said to be ascribed a lower level of humanness, which can foster feelings of disgust or contempt. Similar to theories on moral exclusion (Bandura, 1999; Opotow, 1990), these feelings may facilitate aggression, as the target is said to fall outside the boundary of moral considerations (Harris & Fiske, 2006b; Maoz & McCauley, 2008). Empirical research has also suggested that trait-based dehumanization facilitates social exclusion (Bastian & Haslam, 2010) and reduces prosocial behaviour (Andrighetto et al., 2014) towards the target. Bastian et al. (2013) suggested that trait-based dehumanization directly predicted the endorsement of harsher punishment for criminals, independent of the nature and the severity of the crime they committed (see also Viki et al., 2012; Zhang et al., 2015).

Regarding infrahumanization theory, Leyens and colleagues (2000, 2001) have suggested that perceiving outgroup members as experiencing secondary emotions to a lesser extent than ingroup members could encourage harm in a manner similar to the moral exclusion hypothesis (Bandura, 1999; Opotow, 1990). Leyens and colleagues (2001) elaborate that the link between their conceptualisation of subtle dehumanization captures a perception of outgroup members as sharing an “essence of humanity” with ingroup members to a lesser degree and, in turn, having dissimilar norms and values (p. 408). Citing Struch & Schwartz (1989), Leyens and colleagues further note that such perceived differences in norms and values have been shown to determine levels of aggression towards outgroup members.

Empirical work has suggested that infrahumanization is associated with reduced prosocial behaviour and increased aggression towards the target. Cuddy et al. (2007) asked participants to estimate the extent to which Hurricane Katrina survivors of their racial ingroup and outgroup experienced uniquely human emotions such as *grief*, *sorrow*, and *mourning*, as well as emotions shared with other animals such as *anger*, *panic*, and *rage*. The less their participants attributed uniquely human emotions to outgroup members, the less willing they were to offer them help. In related experimental work, Vaes et al. (2002, 2003) found that participants were less likely to avoid and more likely to respond prosocially to a stranger who expressed a uniquely human emotion (*disappointment*) than to a stranger who expressed an emotion shared with other animals (*anger*). In research concerning gender-based prejudice, it

has been suggested that men who hold hostile sexist attitudes towards women attribute less complex emotions to them, which in turn might justify their mistreatment (Viki & Abrams, 2003). Empirical work has also suggested that infrahumanization can promote harm in contexts of intergroup conflict, such as Tam et al. (2007), who concluded that infrahumanization is detrimental to intergroup forgiveness of past wrongdoings in Northern Ireland (see also Gaunt, 2009; Tam et al., 2008).

The mental state attribution account also proposes a link between subtle dehumanization and harm. People seen as lacking agentic and experiential mental states might be viewed as lacking moral responsibility or the right to be protected from harm (Gray et al., 2007; as cited by Haslam & Loughnan, 2014). Research examining a link between the denial of mental states to a human target often looks at the intergroup context of gender-based violence. Touching upon the psycho-sociological concept of *objectification* (see Nussbaum, 1995), Loughnan and colleagues (2010) suggest that perpetrators of misogynistic violence deny full personhood to women, specifically mind and moral status. Gender-based violence and its association with the dehumanization of women, in particular, has also been widely examined in the psychological literature (Bernard et al., 2020, 2015; Fredrickson & Roberts, 2016; Seabrook et al., 2019). Correlational studies suggest that men who deny mental states to women are likely to show tendencies towards sexual aggression (Rudman & Mescher, 2012) and intimate partner abuse (Bastian, 2019; see also Pizzirani & Karantzas, 2018). Moreover, survivors of sexual assault may also be dehumanized, as suggested by Loughnan et al. (2013), whereby the dehumanization of survivors was associated with more victim-blaming and less perceived suffering than survivors who were not dehumanized (see also: Bernard, Loughnan, et al., 2015; Taylor, 2020).

Dehumanization Following Harm

In the modern world, minority groups that have been historical victims of colonialism, genocide, and marginalisation continue to struggle in the same systems of oppression that have perhaps become more subtle rather than dismantled. The link between dehumanization and harm is thought to shed light on how such systems can survive generations later. Some

research posits that dehumanization can occur *after* harm has been inflicted as a means of dissonance reduction, whereby the perpetrator dehumanizes their victim to ease their guilt (Baumeister et al., 1994; Castano & Giner-Sorolla, 2006; Waytz et al., 2010). Čehajić and colleagues (2009) supported this claim when they found that reminding group members of violence they have collectively committed against another group (such as Non-Indigenous Chileans and Serbs in Bosnia) raised their sense of ingroup responsibility while simultaneously encouraging victim dehumanization. The researchers concluded that members of perpetrating groups morally disengage with the victims through the denial of secondary emotions, to feel less empathy towards them and, in turn, less collective guilt (see also Bandura, 1999).

Similarly, the apparent dehumanization of historical harm victims is associated with less support for reparation policies (Zebel et al., 2008). Interestingly, this effect has also been reported in response to another person's anguish when the group identity of the perceiver and the perceived are not relevant by Kozak et al. (2006). Kozak and colleagues suggest that perceived suffering seems to discourage acknowledgement of the victim's mind, thoughts, and feelings as a moral disengagement strategy (see Bandura et al., 1996), whereby dehumanizing victims of harm relieves the distress that occurs when an individual's awareness of their suffering becomes salient. This process of moral disengagement has been applied when explaining how apparent dehumanization can lead to less intention to help those in need (Cuddy et al., 2007; Vaes et al., 2003) and increased reluctance to provide intergroup help, such as majority-group opposition to the intake and integration of foreign refugees (Bruneau et al., 2018).

1.4. Limitations of Dehumanization Research

In the past two decades, a large body of psychological literature has been published on the construct of dehumanization using the models detailed in this chapter. However, serious issues remain regarding the theories and metrics used in the psychological study of dehumanization. Firstly, dehumanization research suffers from a lack of consistent operationalization of the construct. The following footnote by Posselt (2017) aptly articulates

how loosely the term can be interpreted: “To dehumanize an individual can mean to depersonalize it, to objectify or to de-subjectify it, to animalize it, to outgroup it, to de-individualize or stereotype it – or, which seems to be the general underlying assumption, to deny him or her equality (sic)” (p. 16). The range of contexts and targets in which apparent dehumanization has been found within empirical literature has grown considerably in recent years. These can sometimes include groups where the sources and consequences of dehumanization can be dubiously variable, such as short people (Kunst et al., 2017), obese people (Kersbergen & Robinson, 2019), asexuals (MacInnis & Hodson, 2012), cyclists (Delbosc et al., 2019), or those who have opposing political views to your own (Pacilli et al., 2015; Roccato et al., 2018). As Bloom (2022) notes, “If everything is dehumanization, nothing is dehumanization”.

There have also been several critiques regarding theoretical shortcomings in psychological approaches to dehumanization and its, at times, insufficient separation from interrelated phenomena such as prejudice, stereotypes, or intergroup preference (Bloom, 2017b; Enock et al., 2021a, 2021b; Lang, 2020; Manne, 2016, 2017; Over, 2020a, 2020b; Smith, 2014, 2016, 2023). Over (2020a, 2020b) summarises the current empirical paradigm of the causal relationship between dehumanization and harm as the *dehumanization hypothesis*, composed of two interrelated claims. The first is that victims of intergroup harm are perceived as similar to nonhuman entities, and the second is that natural inhibitions against inflicting harm on the target are eroded because of this perception. Over then poses challenges to this hypothesis, critically reflecting upon the most prominent theories of dehumanization from philosophy, neuroscience, and social psychology. Emphasising the need to reassess the limits to which dehumanization can be contextually operationalised, Over calls for clarification on how the denial of *negative* mental states, antisocial emotions, and *socially undesirable* traits fit into its existing characterisations.

Trait and Emotion Sociality: A Possible Confound

According to infrahumanization theory, dehumanization occurs when an outgroup is perceived as lacking the capacity to experience secondary emotions, such as remorse or

nostalgia, to the same extent as one's ingroup (Leyens et al., 2000, 2001). Leyens and colleagues have argued that the valence of secondary emotions (how positive or negative they are to experience as an individual) does not factor into how they are attributed to one's ingroup and outgroup members. While some studies have come to similar conclusions (see Castano & Giner-Sorolla, 2006; Cortes et al., 2005; Paladino et al., 2002), Enock et al. (2021b) argue that valence is not equivalent to the sociality of emotions. For instance, guilt is a negative emotion to experience but is not considered antisocial. When one expresses guilt, it is generally considered an appropriate response to wrongdoing. Thus, expressing guilt might convey good character and gain social approval from others (Stearns & Parrott, 2012, as cited by Enock et al., 2021b).

After running a pretest in which emotion valence and sociality were shown to differ, Enock et al. (2021b) applied both dimensions in a series of studies utilising the infrahumanization approach. Specifically, the researchers examined whether outgroup members were denied uniquely human emotions that were antisocial (e.g., arrogance or resentment) to the same extent that they denied prosocial emotions. However, members of the studies' intentionally antagonistic outgroups (e.g., anti-vaxxers, criminals) were perceived as experiencing prosocial secondary emotions to a lesser extent than the ingroup, yet antisocial secondary emotions to a greater extent than the ingroup. Enock and colleagues also found this trend occurring in a non-antagonistic context using a minimal groups design. This pattern diverges from the central hypothesis of the infrahumanization approach (Leyens et al., 2000, 2001), which expects outgroups to be attributed secondary emotions to a lesser extent than the ingroup. Instead, Enock and colleagues' findings supported an approach anchored in intergroup preference, whereby people perceive uniquely human emotions to be experienced to a greater extent by ingroup members when they are prosocial but by outgroup members when antisocial. This trend was also found by Enock & Over (2022), who concluded that reduced intentions to help outgroup members were better predicted by the attribution of antisocial uniquely human emotions rather than infrahumanization.

Haslam's dual model (2006) posits that denying uniquely human traits (*civilised*, *rational*, *moral*, *refined*, *mature*) to an individual or group constitutes animalistic dehumanization, and likewise, denying human nature traits (warm, openminded, depth of character, emotionally responsive, self-determined) constitutes mechanistic dehumanization. However, this model was developed by exclusively relying on socially desirable traits in its theoretical framework and ensuing measures. A series of studies by Enock et al. (2021a) highlighted the need to recognise the denial of undesirable human character traits (such as *arrogant*, *jealous* or *corrupt*) in dehumanization research. Enock and colleagues identified traits that scored high in both dimensions by conducting a pretest where participants indicated how human and how desirable a range of character traits were. Along with human traits that were desirable, socially undesirable uniquely human traits (*corrupt*, *arrogant*, *superficial*) and socially undesirable human nature traits (*jealous*, *selfish*, *stingy*) were a novel addition to the dual model's approach to intergroup attributions. With the inclusion of socially undesirable and desirable traits, Enock and colleagues expected intergroup bias, rather than dehumanization, to explain the variance in trait attributions across numerous intergroup contexts. According to the dual model, outgroup members should be denied uniquely human and human nature traits regardless of trait sociality.

However, this was not the case. Instead, Enock and colleagues (2021a) observed trends indicative of intergroup preference similar to those found by Enock et al. (2021b) and Enock and Over (2022) when examining infrahumanization theory. Desirable human traits were attributed more to the ingroup, but undesirable human traits were attributed more to the outgroup. This bias was evident across three social divisions examined: those who voted to leave or remain in the UK's Brexit referendum, right-wingers' perceptions of immigrants, and perceptions of criminals by individuals with no criminal history. Moreover, non-uniquely human traits were also included when examining the latter two intergroup contexts, for which the same ingroup bias emerged. Enock and Over (2023) observed a similar trend in which outgroup members described using animalistic slurs were ascribed more undesirable human traits rather than dehumanized according to the dual model. These pieces of research challenge

the dual model's framework, as they demonstrate how measuring the attribution of essentially human traits that vary in terms of their social desirability leads to apparent evidence of trait-based dehumanization collapsing into intergroup preference instead.

Considering the above critiques and evidence-based clarifications of some of the leading theories in dehumanization research, I maintain that there are valid grounds to reflect on the current literature surrounding the topic critically. Furthermore, there is reasonable doubt to re-examine one of the most widely accepted qualities of dehumanization: its apparent causal relationship to harm.

Questioning the Causal Link Between Subtle Dehumanization and Harm

One of the core challenges Over (2021a, 2021b) poses to dehumanization research is that victims of harm are often targeted *because* of their perceived humanity. In this view, targets of harm are not misperceived as lacking a mind, morals, or mental states; instead, they are attributed antisocial or socially undesirable aspects of humanity. Noting examples such as football fans making monkey noises at Black players during matches and an account of Jewish people being forced to scrub the streets in Nazi-occupied Austria, Bloom (2017) notes that such intentions to humiliate or reinforce subordination lie precisely in the recognition that the targets are human. Manne (2017) makes a similar case in the context of gender-based violence, noting that it may often be the case that a man's misogyny and recognition of a woman's humanity, rather than dehumanization, will lead them to inflict harm as a means of dominance expression. In the same vein, these acts would not be committed against nonhuman animals or machines, as intentions to humiliate and psychologically subjugate only make sense with human targets. Similarly, Davis (1981) argues that despite their ardent portrayal of Black US Americans as subhuman chattel, laws enforced by the White majority under segregation would not have forbidden Black Americans from accessing education if they did not recognise their humanity and associated potential to learn (p. 101). Furthermore, Lang (2010, 2020) criticises the academic reliance on dehumanization to explain how humans can inflict extreme harm on one another in contexts of intergroup conflict and genocide. Noting how popular construals of dehumanization often minimise the meaning behind one group's

violence, such as shifting power dynamics motivating one group to maintain hierarchical superiority through the subjugation or eradication of another, Lang argues that a social relationship exists between the perpetrator and victim of harm, in which recognition of common humanity is required (see also Sofsky, 1996).

The explanatory power of dehumanization in contexts of harm becomes further questionable when morality is considered. Over (2020a) notes that moral-centric conflict does not exist in the animal kingdom outside of humans, implying that to inflict harm on moral grounds is, by default, a recognition of the target's mental capacity and moral standing. Rai et al. (2017) find evidence of such a paradigm in empirical work, concluding that dehumanization removes the meaning behind morally motivated violence (p. 8514). It has also been argued that perceiving animals, objects, and other people as lacking in humanity does not solely determine harm towards them (Over, 2021a; Smith, 2021b). Some developmental research suggests that differences in the acceptability of harm towards these targets might be socially learned, with younger children tending to be more equalitarian in their views and older children prioritising human safety over nonhuman animals and robots (Reinecke et al., 2021; Wilks et al., 2021). Social learning also plays an important role in shaping one's moral considerations and biases towards different groups of people (for a review, see Over & McCall, 2018). Thus, whether seeing another person or group as less human puts them at risk of harm, independent from the perceiver's pre-existing prejudices, intergroup biases, and moral beliefs surrounding who should be harmed and protected within their social context, is not as straightforward as often assumed in the literature.

An alternative explanation of intergroup harm by Over (2021a, 2021b) is that attributing negative aspects of humanity to outgroup members encourages harm towards them rather than dehumanization (see also: Smith, 2020). Relating to social-psychological models of subtle dehumanization, outgroup harm could be predicted by the perception of negative mental states (e.g., plotting against the ingroup), undesirable human traits (e.g., *controlling*, *jealous*), or antisocial uniquely human emotions (e.g., *resentment*, *pessimism*) in outgroup members, rather than seeing them as lacking in these human qualities overall. Given the

intrinsic real-world value of such research, the models and theories we use to explain intergroup problems such as racism and sexism must represent reality as closely as scientifically possible. Thus, evidence-based support is needed to challenge the assumed causal link between dehumanization and harm.

1.5. The Present Research

This thesis reports 11 studies constituting my doctoral research. This research had three overarching aims. The first was to investigate the extent to which social-psychological findings on subtle forms of dehumanization replicate, specifically those relating to the dual model of trait-based dehumanization (Haslam, 2006) and infrahumanization theory (Leyens et al., 2000, 2001). The second aim was to shed light on whether earlier evidence of trait-based dehumanization and infrahumanization can be better explained by effects associated with the confound of social desirability, such as negative evaluation of the target and intergroup preference. The third aim of my doctoral work was to critically evaluate the hypothesised causal relationship between trait-based dehumanization and harm. These aims converge on the intention to provide a critical yet constructive rethinking of subtle dehumanization models to better our understanding of intergroup perceptions and harm.

In the first empirical chapter of this thesis, I aim to provide an evidence-based re-examination of the dual model of dehumanization (Haslam, 2006) based on a central critique posed by Over (2020a, 2020b). This critique is the theoretical shortcoming whereby the dual model emphasises uniquely human and human nature traits that are socially desirable and overlooks those that are socially undesirable, enabling trait desirability to confound the model, as evidenced by Enock and colleagues (2021a). In Studies 2.1A and 2.1B, I initially aimed to replicate Bastian et al. (2013), whereby an apparent positive relationship was found between the trait-based dehumanization of criminal targets and participants' endorsement of harsher punishments. Studies 2.2A and 2.2B introduced socially undesirable uniquely human and human nature traits to the design, investigating whether trait desirability moderates the previously described relationship. Studies 2.3A and 2.3B then used an experimental design to

manipulate the desirability and humanness of the traits that participants attributed to criminals to determine if one's endorsement of harsh punishment is directly affected by such variability in these two trait dimensions.

In the second empirical chapter of this thesis, I attempt to replicate the phenomenon of infrahumanization following harm initially proposed by Castano and Giner-Sorolla (2006) and further evidenced by Čehajić and colleagues (2009). Across three highly powered and pre-registered studies, the extent of harm experienced by members of an outgroup and the sense of collective responsibility for the harm are manipulated across two novel intergroup contexts. In Studies 3.1 and 3.3, UK residents perceived collective responsibility for the harm experienced by residents of Lagos, Nigeria, due to climate change. In Study 3.2, UK consumers perceived collective responsibility for the harm experienced by textile workers in fast fashion supply chains. Studies 3.1 and 3.2 followed the design of Čehajić and colleagues (2009), whereby the dependent measure was the extent to which participants perceived members of the target outgroup as typically experiencing the same four positive primary emotions, four negative primary emotions, four positive secondary emotions, and four negative secondary emotions these researchers used.

In Study 3.3, we followed Castano and Giner-Sorolla's (2006) design in a more elaborate attempt to replicate the infrahumanization of outgroup members that one's ingroup has harmed. Participants first rated how uniquely human and how positively or negatively valenced 64 emotions were before the experimental manipulation. After reading the same climate change vignettes as in Study 3.1, they attributed these 64 emotions to residents of Lagos. This method gave infrahumanization following harm a fairer chance of replicating with more elaborate measures used. The variables of emotion humanness and emotion valence were also treated as continuous variables, as done by Castano and Giner-Sorolla (2006). An additional hypothesis was tested in Study 3.3, whereby outgroup members were expected to be seen as experiencing more negative emotions when depicted as being harmed compared to when they are not depicted as being harmed. Challenging the hypothesis that we might subtly dehumanize members of outgroups we have harmed, we expected a recognition of

negative affect when harmed to occur regardless of emotion humanness or how collectively responsible participants felt for the harm.

In the third empirical chapter of this thesis, I aimed to re-examine infrahumanization in light of recent critiques and empirical work suggesting infrahumanization theory confounds subtle dehumanization with intergroup preference (Bunce et al., 2024.; Enock et al., 2021b; Enock & Over, 2022; Over, 2021a, 2021b). Should an intergroup preference emerge, outgroup members would perceive a typical ingroup member as feeling prosocial emotions more strongly and a typical outgroup member as feeling antisocial emotions more strongly, regardless of emotion humanness. In two highly powered and pre-registered studies, we included a much larger set of emotions than is typically used in infrahumanization research (68 in total) in our measures, as recently mentioned as a limitation to infrahumanization research (Enock & Over, 2022; Vaes, 2023). The intergroup context examined in Study 4.1 was UK nationals (ingroup) and immigrants to the UK (outgroup), and in Study 4.2, it was British people (ingroup) and Irish Travellers (outgroup).

Chapter 2: Is Outgroup Harm Predicted by Trait-Based Dehumanization or Negative Evaluation?

2.1. Abstract

Previous work has reported that the extent to which participants dehumanized criminals by denying them uniquely human character traits such as refinement, rationality, and morality predicted the severity of the punishment endorsed for them. I revisited this influential finding across six highly powered and pre-registered studies. First, I conceptually replicated the effect reported in previous work, demonstrating that our method was sensitive to detecting relationships between trait-based dehumanization and punishment, should they occur. I then investigated whether the apparent relationship between trait-based dehumanization and punishment was driven by the desirability of the traits incorporated into the stimulus set, their perceived humanness, or both. To do this, I asked participants to rate the extent to which criminals possessed uniquely human traits that were either socially desirable (e.g., cultured, civilised) or socially undesirable (e.g., arrogant, bitter). Correlational and experimental evidence converged on the conclusion that apparent evidence for the relationship between trait-based dehumanization and punishment is better explained by the extent to which participants attribute socially desirable attributes to criminals rather than the extent to which they attribute uniquely human attributes. These studies cast doubt on the hypothesised causal relationship between trait-based dehumanization and harm, at least in this context.

2.2. Introduction

The research reported in this chapter concerns the dual model of dehumanization (Haslam, 2006), an influential trait-based account of subtle dehumanization. According to the dual model, individuals and groups are dehumanized to the extent that they are denied uniquely human character traits. The dual model distinguishes between two forms of dehumanization (Haslam et al., 2004). When outgroup members are *animalistically dehumanized* or perceived as similar to animals, they are thought to possess traits such as civility, refinement, rationality, moral sensibility, and maturity to a lesser extent than the ingroup. When outgroup members are *mechanistically dehumanized* or perceived as similar to robots, they are thought to possess traits such as emotional responsiveness, interpersonal warmth, depth, cognitive openness, and agency to a lesser extent than the ingroup.

According to the dual model, the more an individual or group is either animalistically or mechanistically dehumanized, the greater their risk of being harmed (Haslam, 2006; Haslam & Loughnan, 2014). Haslam and Loughnan (2014) argue that "dehumanization is important as a psychological phenomenon because it can be so common and yet so dire in its consequences" (p. 401). Haslam (2021) further notes, "many studies have examined how dehumanizing perceptions enable harm or provide support for it. Some of this work points to direct links between tendencies to dehumanize others and… aggressive behaviour" (p. 139). Empirical research has suggested that trait-based dehumanization facilitates social exclusion (Bastian & Haslam, 2010) and reduces prosocial behaviour (Andrighetto et al., 2014).

The dual model's framework has been employed in many empirical studies examining a supposed link between trait-based dehumanization and increased hostility towards the target (Barber & Davis, 2022; Chen-Xia et al., 2023; Kasper et al., 2022; Morehouse et al., 2023; Rousseau et al., 2023; West & Thomson, 2023). Bastian and colleagues (2013) conducted an influential set of studies testing the hypothesised association between the denial of human character traits and the endorsement of harsh punishment. The researchers measured how trait-based dehumanization influenced participants' punishment of criminals. Participants were asked to rate their agreement with four items assessing animalistic

dehumanization of criminals: "*I felt like the person in the story was refined and cultured*" [*reversed*], "*I felt like the person in the story was rational and logical, like they were intelligent*" [reversed], "*I felt like the person in the story lacked self-restraint, like an animal*", and "*I felt like the person in the story was unsophisticated*". Participants were also asked to rate their agreement with four items assessing mechanistic dehumanization of criminals: *"I felt like the person in the story was open minded, like they could think clearly about things*" [reversed], "*I felt like the person in the story was emotional, like they were responsive and warm*" [reversed], "*I felt like the person in the story was superficial like they had no depth*", "*I felt like the person in the story was mechanical and cold, like a robot*". Bastian and colleagues (2013) reported that both forms of dehumanization predicted endorsement of harsh punishment for the criminals portrayed in their stimuli, concluding that their participants viewed criminals as "subhuman and beastly*"* (p. 9).

Recently, however, the explanatory value of the dual model has been called into question (Enock et al., 2021a; Over, 2021a, 2021b). According to these critiques, evidence for trait-based dehumanization is often confounded with social desirability. Bastian and colleagues (2013) suggested that criminals were animalistically dehumanized based on the observation that participants judged them to be unsophisticated, lacking self-restraint, unrefined, uncultured, irrational, and unintelligent. Evidence that criminals were mechanistically dehumanized came from the observation that participants viewed them as superficial, cold, and lacking in warmth and responsiveness. These results may reflect dehumanization because the traits criminals were found to lack are those perceived as uniquely or essentially human (Haslam, 2006; Haslam et al., 2004). However, as the traits deemed uniquely human were all socially desirable, evidence for trait-based dehumanization cannot be separated from evidence of negative evaluation. An alternative explanation for the findings of Bastian et al. (2013) is that participants endorse harsh punishment against criminals to the extent they perceive criminals to possess undesirable or antisocial characteristics.

Bastian and colleagues (2013) sought to account for this possibility by statistically controlling for participants' moral outrage at the targets' behaviour in their analysis. They reported that the relationship between trait-based dehumanization and punishment remained even when controlling for moral outrage. While this is interesting and suggests independent effects of dehumanization, it cannot fully address the conceptual weaknesses in how dehumanization was operationalised. A more convincing way to de-confound evidence for trait-based dehumanization from evidence of negative evaluation is to ask participants to rate the target group on traits that are uniquely human but vary from socially desirable to undesirable (Over, 2021a, 2021b). Previous research conducted by Enock et al. (2021a) has established that undesirable character traits such as *jealous*, *spiteful*, and *bitter* are considered unique to humans and socially undesirable. Across three intergroup contexts, the researchers found that participants attributed socially desirable human traits more strongly to the ingroup and socially undesirable traits more strongly to the outgroup (see also: Decker & Lord, 2023; Enock et al., 2021b; Enock & Over, 2022, 2023). Enock et al. (2021a) concluded that intergroup preference may better explain apparent evidence for trait-based dehumanization. However, it is unclear how attributing uniquely human character traits relates to harm. Addressing this question is crucial to understanding the extent to which the dual model of dehumanization can help explain real-world discrimination and intergroup negativity.

We revisit the hypothesised causal relationship between trait-based dehumanization and harm in the context of endorsing harsh punishment for criminals. In Studies 2.1A and 2.1B, we seek to conceptually replicate a key finding of Bastian et al. (2013), suggesting that the extent to which participants animalistically (Study 2.1A) and mechanistically (Study 2.1B) dehumanize criminals predicts the severity of the punishment participants endorse for them. In Studies 2.2A and 2.2B, we adopt a similar design but incorporate socially undesirable traits into our stimulus set. This addition to the design allows us to investigate whether trait-based dehumanization, undesirable trait attribution, or both predict the severity of punishment. Following Bastian et al. (2013), and to understand the generalisability of our findings, we investigate these questions in relation to two different types of crime (violent crime and theft).

In Studies 2.3A and 2.3B, we seek to investigate a similar question using an experimental design and focusing on parole decisions rather than sentencing. We present participants with vignettes in which criminals are described using character traits that differ in how socially desirable they are and whether they are unique to humans. We then measure how these varying descriptions influence participants' parole decisions. This design allows us to directly measure whether there is a causal relationship between trait-based dehumanization and punishment, independent of negative evaluation.

2.2.1. Note on Methods

All studies received ethical approval from the Psychology Departmental Ethics Committee at the University of York (approval number 926). All data collection occurred online, and the studies were created and administered using Qualtrics (https://www.qualtrics.com). Participants were recruited through the online platform Prolific (https://www.prolific.co), with an independent sample recruited for each study. Informed consent was obtained at the start of each session according to approved ethical guidelines. Inclusion criteria for each study included adult participants fluent in English who had never been to prison for committing a crime and had a Prolific approval rating of at least 90% (95% for Studies 2.3A and 2.3B). Increases in Prolific's recommended rate of compensation for participation during data collection meant the reward ranged from approximately £7 per hour in Studies 2.1A and 2.1B to approximately £8 in the other four studies. Assumption testing and analyses were conducted using *SPSS* and *RStudio*. Highly influential cases were identified using Cook’s distance in each study. All studies were pre-registered on AsPredicted.com before commencing data collection. The stimuli used for each study can be seen in Appendix A: Supplementary Materials for Chapter 2.

2.3. Study 2.1A: Does Trait-Based Dehumanization Predict Punishment Decisions for Violent Criminals?

Bastian et al. (2013) presented evidence that the more participants dehumanize violent criminals, the harsher the punishment participants endorse for them. We sought to test whether we could conceptually replicate this relationship between trait-based dehumanization and punishment using terms similar to those used by Bastian and colleagues. In Study 2.1A, participants read a series of scenarios describing fictitious criminals and their violent crimes. Following this, participants rated the extent to which the criminals possessed four character traits that distinguish humans from nonhuman animals (*refined*, *rational and logical*, *has a sense of morality,* and *civilised*), which we refer to as uniquely human traits. Participants also rated the extent to which criminals possessed four character traits that distinguish humans from machines (*openminded*, *emotionally responsive*, *has a depth of character*, and *interpersonally warm*), which we refer to as human nature traits. In investigating traits that distinguish humans from animals and machines, we respect the distinction between animalistic and mechanistic dehumanization, which is fundamental to the dual model (Haslam, 2006, p. 256). Participants also responded to an item measuring how harsh they thought the punishment for violent criminals should be. Following Bastian et al. (2013), we predicted that the less participants attributed these uniquely human traits to criminals, the harsher the punishment they would recommend for criminals.

2.3.1. Method

Participants

A power analysis using G*Power indicated that a sample size of 89 would allow us to detect a medium effect size (f^2 = .15) with an alpha of .05 and power of .95. A medium effect was chosen based on previous findings in the area of trait-based dehumanization. A final sample of 100 participants was collected, with 54 identifying as female, 44 as male, and two as nonbinary. Ages ranged from 18 to 63 (M = 26.5, SD = 8.94). Following our pre-registered exclusion criteria, data submitted by six individuals who failed one or both attention checks

(i.e., gave a response more than 20 points away from the instructed end of the scale) were omitted and replaced. Participation took an average of approximately eight minutes.

Materials

Vignettes. All participants responded to the same five vignettes detailing different scenarios involving violent crimes. An effort was made to ensure that all five vignettes were similar in length, degree of detail, and severity of crimes depicted. In each vignette, the target criminal's age and ethnicity were not indicated, and the scenarios depicted were all set in unspecified locations. Each target's name and pronouns were gender-neutral, though we cannot rule out assumptions made by participants about gender given the prevalence of androcentrism (Bailey et al., 2018; Van Fleet & Atwater, 1997), particularly in contexts of crime (see Ruibyte et al., 2016; Seager & Barry, 2019). All vignettes can be seen in full in Appendix A: Supplementary Materials for Chapter 2. For example: *"Charlie was arrested after a fight broke out in a pub soon after opening time, apparently triggered by a minor disagreement. Charlie smashed a pint glass and used it to stab another customer. Two additional customers received cuts as they tried to hold Charlie back until the police arrived."*

Trait Attribution. After reading each vignette, participants responded to items designed to measure trait-based dehumanization, broadly following the procedure of Bastian et al. (2013). Participants indicated the extent to which they attributed four uniquely human traits (*refined*, *rational and logical*, *has a sense of morality* and *civilised*) and four human nature traits (*openminded*, *emotionally responsive*, *has a depth of character*, and *interpersonally warm*) to the criminals depicted. Participants indicated their agreement with each item (e.g., *I think* [e.g., *Charlie*] *is refined*) using an unmarked sliding scale from 0 (*Strongly Disagree*) to 100 (*Strongly Agree*), with the sliders initially fixed at the midpoint. According to the dual model, lower scores indicated greater dehumanization of violent criminals. An attention check appeared halfway through the dehumanization items for two criminals (*Please move the slider all the way to Strongly Agree/Disagree*).

Harshness of Punishment Endorsed. Using an unmarked sliding scale that ranged from 0 (*Not at all harsh*) to 100 (*Very harsh*), participants were asked to respond to the following question, "*How harsh do you think the punishment for* [e.g., *Charlie*] *should be*?".

Design

Following Bastian et al. (2013), we utilised a within-subjects, correlational design. All participants read the same five vignettes presented in a random order and responded to the same trait attribution and punishment items. Participants' scores for the trait attribution items and the endorsed harshness of punishment items were then averaged across scenarios. The presentation of the items in the trait attribution task was also randomised.

Procedure

Participants were informed that the study would examine how social attributions influence our behavioural intentions towards criminals. After providing informed consent, participants answered a few demographic questions and confirmed that they had never been to prison for committing a crime. The first of five vignettes then followed. After reading the vignette, participants were asked to respond to the trait attribution items, followed by the single item asking them to indicate how harshly they thought the criminal should be punished. Participants repeated the above steps for each of the remaining four vignettes. Each vignette remained on the screen for at least 15 seconds to ensure participants read the stimuli carefully.

2.3.2. Results

Model 1: Animalistic Dehumanization and Punishment

In line with our pre-registered criteria, this analysis omitted two highly influential cases (remaining sample $N = 98$). We first calculated the average attribution score for uniquely human traits and punishment for each participant in the sample. We then conducted a simple linear regression to understand whether the extent to which participants attributed uniquely human traits to criminals predicted the harshness of punishment participants endorsed for them. A significant negative relationship was found, $b = -.56$ [-.75, -.37], $t = -5.93$, $p < .001$, see Figure 3. Thus, the more violent criminals were animalistically dehumanized (by being

denied uniquely human traits), the harsher the punishment participants endorsed. The model explained approximately 27% of the variance in the harshness of punishment scores, R^2 = .27, $F(1, 96) = 35.14$.

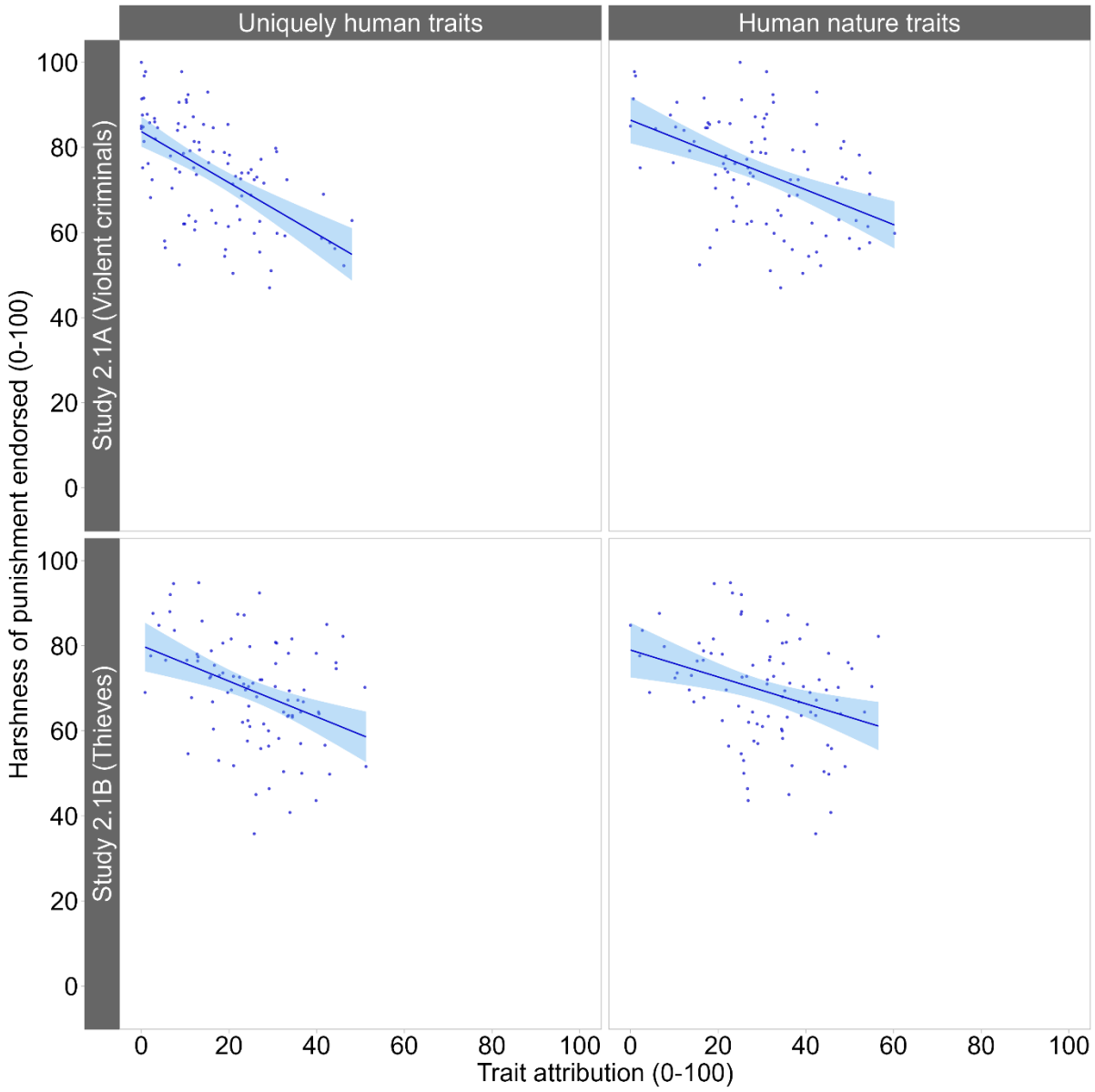

Figure 3. Results of Studies 2.1A and 2.1B

Seemingly in line with Bastian et al. (2013), the greater the extent of animalistic dehumanization (left panels) and mechanistic dehumanization (right panels), the harsher the punishment endorsed for both violent criminals (Study 2.1A, top panels) and thieves (Study 2.1B, bottom panels).

Model 2: Mechanistic Dehumanization and Punishment

In line with our pre-registered criteria, seven highly influential cases were omitted from the analysis (remaining sample $N = 93$). After calculating the average attribution score for human nature traits and punishment for each participant, we conducted a simple linear regression to test whether the attribution of human nature traits predicted the harshness of punishment endorsed for violent criminals. A significant negative relationship was found, $b = -.41$ [-.57, -.24], $t = -4.928$, $p < .001$, see Figure 3.

This relationship shows that greater mechanistic dehumanization (operationalised as the denial of human nature traits) was associated with the endorsement of harsher punishment. The model explains 21% of the variance in the harshness of punishment scores, $R^2 = .21$, $F(1, 91) = 24.29$.

2.4. Study 2.1B: Does Trait-Based Dehumanization Predict Punishment Decisions for Thieves?

Study 2.1B investigates whether the relationship found in Study 2.1A replicates when participants are asked to judge a different type of criminal activity. In Study 2.1B, we examined whether animalistic and mechanistic forms of dehumanization, as operationalised by Bastian et al. (2013), are associated with harsher punishment being endorsed for individuals who commit theft. The design, materials, and analysis plan were similar to that used in Study 2.1A, except that the scenarios involved theft rather than violent crime. In investigating a different type of crime, we follow the example set by Bastian and colleagues and seek to understand the generalisability of our results.

2.4.1. Method

Participants

Based on the same power analysis used in Study 2.1A, a sample of 100 participants was collected, with 55 identifying as male and 45 as female. Ages ranged from 18 to 57

(*M* = 25.5, *SD* = 8.10). The attention checks in Study 2.1A were also used in Study 2.1B. Ten participants failed one or both attention checks, and their data was omitted and replaced as per our pre-registration. Participation took an average of nine minutes.

Materials

The measures of trait-based dehumanization and harshness of punishment endorsed were identical to those used in Study 2.1A.

Vignettes. All participants responded to the same five vignettes, each detailing a crime involving theft (see Appendix A: Supplementary Materials for Chapter 2). As in Study 2.1A, an effort was made to ensure the vignettes were similar in structure and amount of detail. Once again, all the perpetrators had gender-neutral names. An example of one of the theft vignettes is as follows: "*Until their recent arrest, Charlie had worked as a till operator at a local charity shop supporting individuals experiencing homelessness. Charlie had been stealing cash amounts varying from £5 to £50 from the tills almost daily over a five-year period. Police revealed that Charlie had stolen several thousand pounds from the charity shop while working there."*

2.4.1.3. Design and Procedure

The design and procedure of Study 2.1B were identical to those of Study 2.1A.

2.4.2. Results

Model 1: Animalistic Dehumanization and Punishment

Seven highly influential cases were omitted from the analysis (remaining sample *N* = 93). We calculated the average attribution scores for uniquely human trait attribution and punishment for each participant and then conducted a simple linear regression to measure whether trait attribution predicted the harshness of punishment endorsed for thieves. As shown in Figure 3, a significant negative relationship was found, b = -.42 [-.64, -.21], t = -3.97, $p < .001$. Thus, greater animalistic dehumanization of thieves was associated with the endorsement of harsher punishment for them. The model explains approximately 15% of the variance in the harshness of punishment scores, R^2 = .15, $F(1, 91)$ = 15.75.

Model 2: Mechanistic Dehumanization and Punishment.

Eight highly influential cases were omitted from the analysis (remaining sample $N = 92$). After calculating the average score of human trait attribution and punishment, we conducted a simple linear regression to test whether or not human trait attribution predicted the harshness of punishment endorsed for thieves. A significant negative relationship was found, $b = -.27$ [-.47, -.08], $t = -2.77$, $p = .007$. These data show that greater mechanistic dehumanization is associated with the endorsement of harsher punishment for thieves (see Figure 3). The model explains approximately 8% of the variance in the harshness of punishment scores, $R^2 = .08$, $F(1, 90) = 7.65$.

2.5. Study 2.2A: Does Trait-Based Dehumanization or Trait Social Desirability Predict Punishment Decisions for Violent Criminals?

Study 2.2A investigated whether apparent evidence for a relationship between trait-based dehumanization and endorsement of harsh punishment for violent criminals remains when controlling for the desirability of the traits. We tested this by introducing character traits perceived as uniquely human yet socially undesirable into the stimulus set (Enock et al., 2021a; Over, 2021a, 2021b). The dual model predicts that to the extent criminals are denied uniquely human character traits, they will be subjected to harsher punishment. We predict that trait desirability will moderate the relationship between human trait attribution and punishment. More specifically, we predict that the extent to which violent criminals are denied socially desirable character traits and attributed socially undesirable character traits will predict harsh punishment.

2.5.1. Method

Participants

A power analysis using G*Power indicated that a sample size of 119 would allow us to detect a medium effect size ($f^2 = .15$), with three predictors (trait attribution, trait desirability, attribution*desirability), an alpha of .05 and power of .95. To counterbalance the sample

equally and allow for the exclusion of outliers, a sample of 130 was collected. Within the sample, 66 identified as female, 62 as male, and two as nonbinary. Ages ranged from 18 to 55 (M = 28.5, SD = 9.15). Similar to Studies 2.1A and 2.1B, two attention checks were included in this study. Per our pre-registered plan, 16 participants failed one or both attention checks; thus, their data were omitted and replaced. Participation took an average of nearly eight minutes.

Design

This study utilised a mixed design. All participants responded to items designed to measure animalistic dehumanization and mechanistic dehumanization. The desirability of the traits rated by participants was manipulated between subjects: half of the participants rated criminals on the extent to which they possessed socially desirable traits, and half rated criminals on the extent to which they possessed undesirable traits. All participants responded to a single item measuring the harshness of punishment endorsed.

Materials

Vignettes. All participants read the same violent crime vignettes as used in Study 2.1A.

Trait Attribution. After reading each vignette, participants responded to an eight-item scale measuring animalistic dehumanization (four items) and mechanistic dehumanization (four items) of the criminal portrayed. Participants made trait attributions by indicating their agreement with each item using an unmarked sliding scale ranging from 0 (*Strongly Disagree*) to 100 (*Strongly Agree*), all initially positioned at the scale's midpoint. Depending on the condition, the eight trait items were either socially desirable (uniquely human: *cultured, civilised, sophisticated, moral*; human nature: *generous, open-minded, warm, kind*) or socially undesirable (uniquely human: *corrupt, controlling, arrogant, bitter*; human nature: *jealous, selfish, spiteful, cruel*). The lower the score, the more participants dehumanized the criminal target by denying them human traits.

Harshness of Punishment Endorsed. The same single-item scale for measuring the harshness of punishment endorsed in Studies 2.1A and 2.1B was employed in Study 2.2A.

Procedure

The procedure in Study 2.2A mirrored that of Studies 2.1A and 2.1B.

2.5.2. Results

Model 1: Animalistic Dehumanization and Punishment.

Eight highly influential cases were omitted from the analysis (remaining sample N = 122). The regression model tested for a relationship between participants' average scores for uniquely human trait attribution and harshness of punishment endorsed with trait desirability included as a moderator (desirable = 0, undesirable = 1).

The moderated regression showed no significant effect of uniquely human trait attribution on punishment, b = -0.07 [-0.22, 0.09], t = -0.83, p = .408. Thus, when undesirable uniquely human traits were included in the measure of animalistic dehumanization, the previously reported relationship between animalistic dehumanization and the endorsement of harsher punishment by Bastian et al. (2013) was no longer significant.

The interaction between uniquely human trait attribution and trait desirability was significant, b = 1.31 [1.00, 1.63], t = 8.31, p < .001. In line with our prediction, simple slopes showed that the more socially desirable human traits participants attributed to criminals, the less harshly participants thought they should be punished, b = -.58, [0.15, 0.52], t = -5.90, p < .001. The more undesirable traits participants attributed to criminals, the more harshly participants thought they should be punished, b = -.73, [-0.98, -0.49], t = -5.92, p < .001 (see Figure 4). The model explained approximately 38% of the variance in the harshness of punishment endorsed, R^2 = .375, F(3, 118) = 23.61.

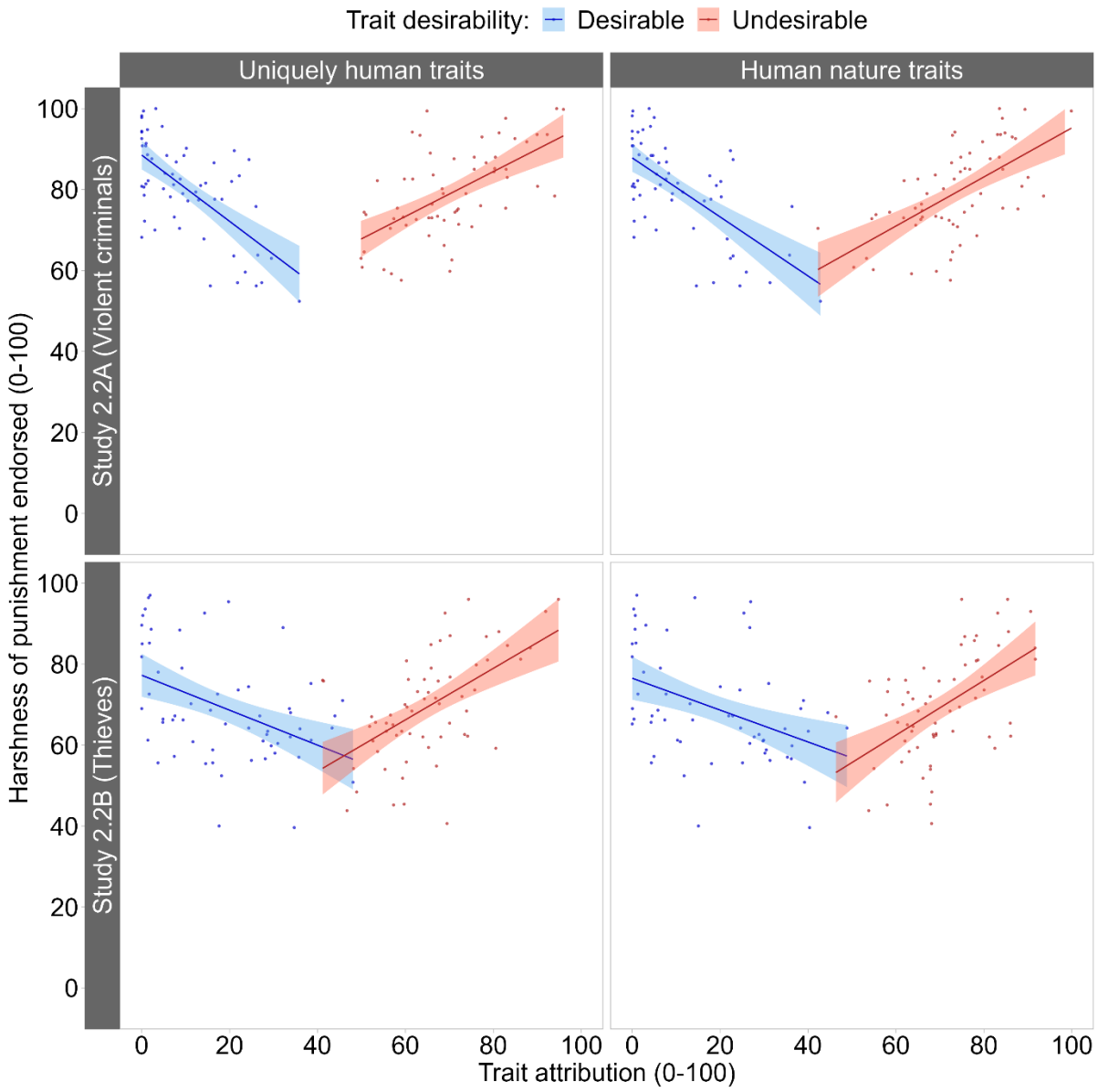

Figure 4. Results of Studies 2.2A and 2.2B

The relationship between trait attribution and punishment for violent criminals (Study 2.2A, top panels) and thieves (Study 2.2B, bottom panels) depends on the social desirability of the traits rather than how uniquely human they are.

Model 2: Mechanistic Dehumanization and Punishment

Eight highly influential cases were omitted from the analysis (remaining sample N = 122). A moderated regression analysis tested for a relationship between the average

scores of human trait attribution and harshness of punishment endorsed to violent criminals and whether this interacted with trait desirability.

The moderated regression showed no significant effects of human nature trait attribution on punishment, b = -0.04 [-0.19, 0.12], t = -0.44, p = .658. The effect reported by Bastian et al. (2013), whereby mechanistic dehumanization predicted harsher punishment endorsement, which we replicated in Studies 2.1A and 2.1B, did not appear when undesirable human nature traits were included in our measures.

The interaction between uniquely human trait attribution and trait desirability was significant, b = 1.37 [1.06, 1.68], t = 8.65, p < .001. Simple slopes indicated that the more participants attributed socially desirable traits to criminals, the less harshly they thought those criminals should be punished b = -0.73, [-0.96, -0.51], t = -6.46, p < .001. As shown in Figure 4, the more participants attributed socially undesirable traits to criminals, the more harshly they thought those criminals should be punished, b = .64 [0.42, 0.86], t = 5.77, p < .001. The model explained 39% of the variance in the harshness of punishment scores, R^2 = .389, F(3, 118) = 25.05.

2.6. Study 2.2B: Does Trait-Based Dehumanization or Trait Social Desirability Predict Punishment Decisions for Thieves?

Study 2.2B sought to replicate the results of Study 2.2A but with thieves as the target group rather than violent criminals. We examined whether the apparent relationship between trait-based dehumanization and the endorsement of harsh punishment for thieves is better explained by the desirability of the traits incorporated into the stimulus set. We investigate this question using a similar design and procedure to Study 2.2A, with the exception that the vignettes are those used in Study 2.1B detailing crimes involving theft. As in Study 2.2A, we hypothesise that trait desirability will moderate the relationship between human trait attribution and punishment. More specifically, we predict the extent to which criminals are denied socially desirable character traits and attributed socially undesirable character traits will predict endorsement of harsher punishment.

2.6.1. Method

Participants

The power analysis described in Study 2.2A informed the sample size for Study 2.2B. A separate sample of 130 participants was collected, of whom 74 identified as male, 53 as female, and three as nonbinary. Ages ranged from 18 to 59 (M = 26.6, SD = 7.48). Data submitted by 20 participants who did not pass one or both checks were omitted and replaced. Participation took an average of nine and a half minutes.

Design

This study utilised a mixed-methods design, matching that of Study 2.2A. The same attention checks used in Studies 2.1A, 2.1B, and 2A were used in Study 2.2B.

Materials

Vignettes. All participants responded to the same five vignettes used in Study 2.1B detailing scenarios involving criminals committing theft.

Trait Attribution and Harshness of Punishment Endorsed. The same scales used in Study 2.2A when measuring animalistic dehumanization, mechanistic dehumanization, and harshness of punishment endorsed were used in Study 2.2B.

Procedure

The procedure in Study 2.2B was identical to that of Study 2.2A, except for the vignettes describing crimes involving theft rather than violence.

2.6.2. Results

Model 1: Animalistic Dehumanization and Punishment

Eight highly influential cases were omitted from the analysis (remaining sample N = 122). A moderated regression tested for a relationship between average scores of uniquely human traits and harshness of punishment endorsed and whether this interacted with trait desirability. The moderated regression showed no significant effects of uniquely human trait attribution on punishment b = .13 [-0.05, 0.30], t = 1.44, p = .152. Replicating the results

of Study 2.2A, when socially undesirable traits were incorporated into the stimulus set, there was no longer any relationship between trait-based dehumanization and punishment.

The interaction between uniquely human trait attribution and trait desirability was significant, $b = 1.08$ [0.74, 1.42], $t = 6.22$, $p < .001$. As illustrated in Figure 4, the more participants attributed socially desirable traits to criminals, the less harshly participants felt they should be punished, $b = -.43$, [-0.65, -0.21], $t = -3.88$, $p < .001$. The more participants attributed undesirable traits to criminals, the more harshly participants felt they should be punished, $b = .65$, [0.38, 0.91], $t = 4.85$, $p < .001$. The model explained about 25% of the variance in endorsed harshness of punishment scores, $R^2 = .247$, $F(3, 118) = 12.89$.

Model 2: Mechanistic Dehumanization and Punishment

The analysis omitted six highly influential cases (remaining sample $N = 124$). A moderated regression tested for a relationship between human trait attribution and punishment and whether this interacted with trait desirability. As in Study 2.2A, and contradicting the findings of Bastian et al. (2013), the moderated regression showed no significant relationship between human trait attribution and punishment, $b = 0.15$ [-0.04, 0.34], $t = 1.57$, $p = .119$.

However, the interaction between human nature trait attribution and trait desirability was significant, $b = .95$ [0.57, 1.33], $t = 5.00$, $p < .001$. As can be seen in Figure 4, simple slopes showed that the more participants attributed socially desirable human traits to criminals, the less harshly participants thought they should be punished, $b = -.33$, [-0.55, -0.11], $t = -2.99$, $p = .003$. The more participants attributed socially undesirable human traits to criminals, the more harshly participants thought they should be punished, $b = .62$, [0.31, 0.92], $t = 4.01$, $p < .001$. The model explained approximately 17% of the variance in the harshness of punishment scores, $R^2 = .174$, $F(3, 120) = 8.44$. These data suggest that the apparent relationships between animalistic and mechanistic dehumanization and punishment reported in previous research (Bastian et al., 2013) are better explained by the social desirability of the traits.

2.7. Study 2.3A: Does Animalistic Dehumanization or Trait Desirability Predict Parole Decisions?

In Study 2.3A, we used an experimental design to examine further the hypothesised causal relationship between trait-based dehumanization and punishment when controlling for the social desirability of human traits incorporated into the stimuli. We described criminals with traits that varied in desirability and perceived humanness, creating a 2*2 design. We then measured participants' willingness to endorse parole for each criminal described. To allow for character traits to be integrated into the vignettes, the context in which the criminals are described differed from those of Studies 2.1A through 2.2B. Rather than describing a criminal act, the vignettes of Study 2.3A each give an account of how other people describe a criminal as they approach a parole hearing. Thus, the dependent measure also differed from Studies 2.1A through 2.2B to fit the different contexts of the vignettes. The dual model predicts that criminals who are described in uniquely human terms will be more likely to be granted parole. We predicted that criminals described in socially desirable terms would be more likely to be granted parole. In principle, this design allows us to detect independent effects of dehumanization and trait sociability or an interaction between the two. In Study 2.3A, we specifically measure the extent to which animalistic dehumanization is causally related to parole decisions. Thus, our measures included uniquely human traits and those shared with other animals. We predict that participants will be more likely to endorse parole for criminals described with socially desirable traits, regardless of whether or not those traits are uniquely human or shared with other animals.

2.7.1. Method

Participants

A power analysis using G*Power, with effect size specification as in SPSS, indicated that a sample size of 135 would allow us to detect a medium effect size ($\eta_p^2 =$ .09) with a 2*2 factorial, repeated measures design, an alpha of .05, and a power of .95. To counterbalance the sample equally, a sample of 136 participants was collected, of whom 78 identified as

female, 55 as male, two as nonbinary, and one who preferred not to indicate their gender identity. Ages ranged from 18 to 63 (M = 24.9, SD = 6.97). All participants were adults fluent in English who had never been to prison for committing a crime. Due to a noticeable increase in failed attention checks during pilot data collection, the minimal approval rating on Prolific was raised from 90% to 95%. Despite this, data from 46 participants were omitted and replaced due to failed attention checks. The high rate of failed attention checks was likely due to a broader issue with Prolific at the time of data collection: a surge in new users after a social media post promoting the platform went viral (see Charalambides, 2021). Three participants were mistakenly recruited after the intended sample size had been met, and thus, their data were excluded from analyses. Including data submitted by excess participants in analyses yielded the same results as those reported. Participation took an average of seven minutes.

Materials

Vignettes. All participants responded to the same four vignettes, each detailing a different scenario in which a criminal's eligibility for parole was assessed. Efforts were made to ensure that all four vignettes were similar in length, degree of detail, and contextual aspects, such as how long the criminal had spent in prison and who was described as attributing the traits to the criminal. In each vignette, the criminal's name and pronouns were gender-neutral, their age and ethnicity were not indicated, and their crime and sentence were not specified. The four vignettes for Study 2.3A can be seen in Appendix A: Supplementary Materials for Chapter 2. The criminal was described as *cultured*, *civilised*, *sophisticated*, and *moral* in the uniquely human socially desirable condition. In contrast, the criminal was described as *corrupt*, *controlling*, *arrogant,* and *bitter* in the uniquely human socially undesirable condition. In the animalistic desirable condition, the criminal was described as *energetic*, *trusting*, *genuine,* and *having curiosity*, while in the animalistic undesirable condition, the criminal was described as *uncultured*, *unrefined*, *unsophisticated*, and *stupid*.

The following is an example of a vignette describing a criminal with uniquely human socially desirable traits: "*Alex, known by locals in their hometown as having always been sophisticated, has recently begun their first parole hearing at the local courthouse. Having*

been tried and convicted 36 months ago, a report by one of the prison's counsellors notes that other prisoners often refer to Alex as being civilised and moral in character. Alex was also described by the counsellor as exhibiting a cultured demeanour since their arrival."

Agreement with Parole. The dependent variable, agreement with granting parole, was measured using the following single-item measure: "*I think* (*Alex/Sam/Robin/Jamie) should be granted parole"*'. This measure appeared after each vignette, and participants indicated their agreement using an unmarked sliding scale ranging from *Strongly Disagree* (0) to *Strongly Agree* (100). The slider's starting point was always centred at 50.

Attention Check. An additional paragraph describing a criminal named Charlie was included, largely similar to the other four paragraphs. However, the following sentence was included in the middle of the paragraph: "*This paragraph is an attention check: please move the slider all the way to Strongly Disagree on the left-hand side"*. Data submitted by any participants who did not respond within 20 points of the instructed end of the 100-point scale were omitted and replaced.

Design

This study adopted a 2(trait humanness: uniquely human, shared) X 2(trait desirability: desirable, undesirable) within-subjects factorial design. Counterbalancing ensured that each vignette was associated with each trait category an equal number of times across the participant sample, resulting in four trait-type orders. The trait words were randomly allocated to the position in which they appeared in each vignette using a random order function in Excel. Mirror versions of the trait orders were then created. These two trait-order conditions were also counterbalanced between participants, which was done to control for possible primacy and recency effects of the order in which traits appeared.

Procedure

After participants provided informed consent, they responded to the same demographic questions and inclusion checks as in the other studies. Participants were then shown the first of the four vignettes. After reading the vignette, participants were asked to respond to a single item measuring their agreement with granting parole to the criminal

depicted. Participants then repeated the above steps for the remaining three vignettes. The order in which the vignettes were presented to participants was randomised. Each vignette appeared on the screen for at least 15 seconds to maximise the chance that participants read all the relevant information. Participants were debriefed and redirected to Prolific to collect their reward after completing the questionnaire.

2.7.2. Results

A 2*2 within-subjects ANOVA was conducted to examine how variations in the desirability (desirable or undesirable) and humanness (uniquely human or shared with other animals) of the traits used to describe criminals influenced participants' agreement with granting them parole. In line with our prediction, a significant main effect of trait desirability was found, $F(1, 135) = 369.43$, $p < .001$, $\eta_p^2 = .73$. Criminals described with socially undesirable traits ($M = 38.8$, $SE = 1.72$) were less likely to be granted parole than those described with desirable traits ($M = 77.8$, $SE = 1.44$), 95% CI [34.98, 43.01]; see Figure 5. A main effect of trait humanness was also found, $F(1, 135) = 51.62$, $p < .001$, $\eta_p^2 = .28$. Contrary to the predictions of the dual model, however, criminals who were described with uniquely human traits ($M = 53.0$, $SE = 1.36$) were less likely to be granted parole than those described with traits shared with other animals ($M = 63.7$, $SE = 1.50$), 95% CI [-13.66, -7.76].

A significant interaction between trait humanness and trait desirability was also found, $F(1, 135) = 54.67$, $p < .001$, $\eta_p^2 = .29$. We conducted paired samples *t*-tests with a Bonferroni-corrected alpha level of .025 to examine interaction effects. Criminals described using undesirable uniquely human traits ($M = 27.9$, $SE = 1.93$) were less likely to be granted parole than were criminals described using desirable uniquely human traits ($M = 78.0$, $SE = 1.70$), $t(135) = 20.75$, $p < .001$, 95% CI [-54.94, -45.37], *Cohen's d* = 1.78. Contrary to the dual model, participants tended to agree more with granting parole to criminals described with uniquely human traits that were desirable than those that were undesirable by a difference of about 50 points on the 100-point scale.

Similarly, criminals described using undesirable traits shared with other species (*M* = 49.8, *SE* = 2.32) were less likely to be granted parole than were those described using desirable traits shared with other species (*M* = 77.6, *SE* = 1.6), *t*(135) = 10.57, *p* < .001, 95% CI [-33.05, -22.63], *Cohen's d* = .91. Demonstrating the impact of differences in trait desirability, participants were more likely to agree with granting parole to criminals described with desirable shared traits than those described with undesirable shared traits by a difference of about 28 points on the 100-point scale.

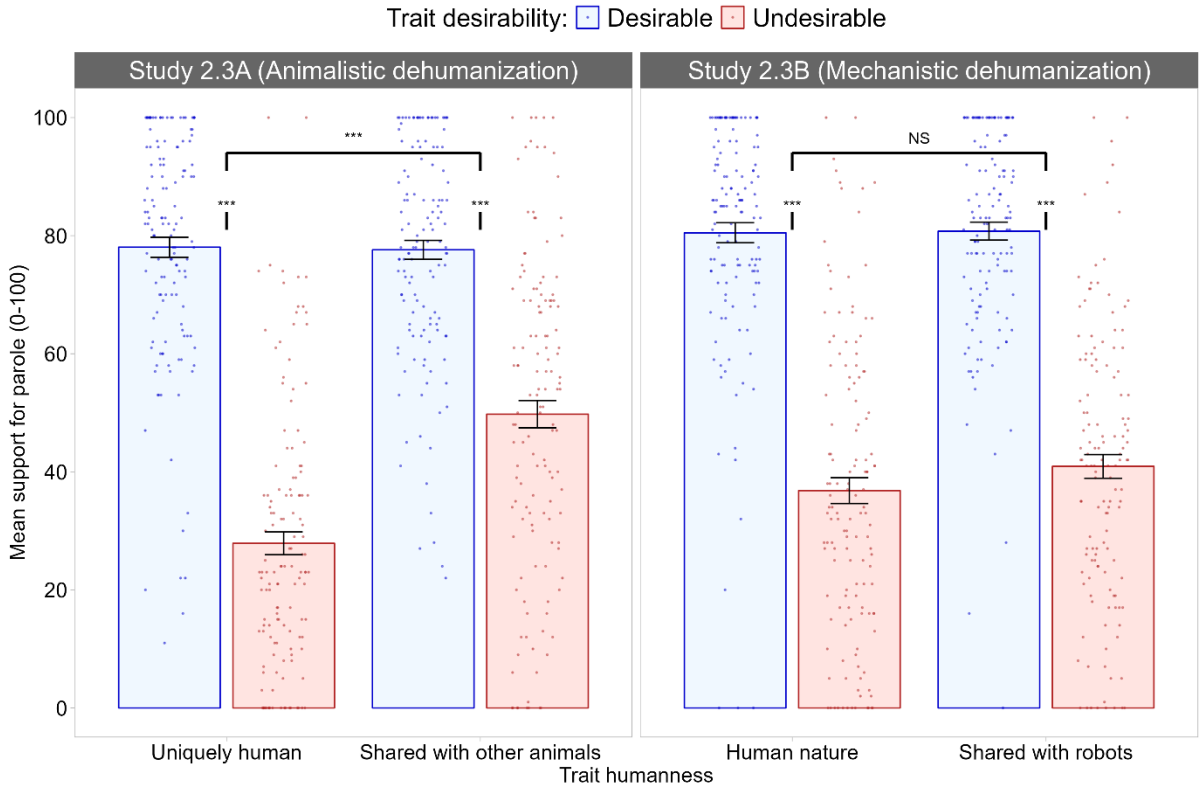

Figure 5. Results of Studies 2.3A and 2.3B

In both Study 2.3A (left) and Study 2.3B (right), criminals described with undesirable traits were less likely to be granted parole than criminals described with desirable traits, regardless of whether or not those traits were essentially human.

Note: Error bars represent ±1 *SE*. *** denotes *p* < .001, *NS* denotes non-significant.

2.8. Study 2.3B: Does Mechanistic Dehumanization or Trait Desirability Predict Parole Decisions?

Study 2.3B had a similar design and method to Study 2.3A. Again, we employed an experimental manipulation in which we manipulated the perceived humanness and sociality of the traits with which criminals were described and measured how these descriptions influenced participants' parole decisions. In Study 2.3B, we specifically tested for an influence of mechanistic dehumanization by including human nature traits and traits shared with robots in our measures.

As in Study 2.3A, we predicted that criminals described with undesirable traits would be less likely to be granted parole than those described using desirable traits.

2.8.1. Method

Participants

The power analysis described in Study 2.3A also informed the sample size for Study 2.3B. A new sample of 136 participants was collected, of whom 76 identified as female, 56 as male, two as nonbinary, and two did not indicate their gender identity. Ages ranged from 18 to 57 (M = 26.4, SD = 8.29). The inclusion criteria were identical to those used in Study 2.3A, including a minimum Prolific approval rating of 95%. Data from 35 participants were omitted and replaced due to failed attention checks. As in Study 3.3A, the high rate of failed attention checks was likely due to a surge in new Prolific users after a social media post promoting the platform went viral at the time of data collection (see Charalambides, 2021). Five participants were mistakenly recruited after the intended sample size had been met, and thus, their data were excluded from analyses. Including data submitted by excess participants in analyses yielded the same results as those reported. Participation took an average of just under seven minutes.

Materials

The agreement with granting parole scale and attention check were the same as those used in Study 2.3A.

Vignettes. All participants responded to the same four vignettes used in Study 2.3A but with different trait words. The desirable human words were *generous*, *openminded*, *warm,* and *kind*. The undesirable human words were *jealous*, *selfish*, *spiteful,* and *stingy*. The desirable traits shared with robots were *helpful*, *disciplined*, *calm,* and *efficient*. The undesirable traits shared with robots were *cold*, *inflexible*, *superficial,* and *passive*. The following vignette is an example of the undesirable shared condition:

"*Sam is currently applying for parole after being convicted of a crime just over three years ago. In assessing Sam's suitability, the parole committee gathered reports from prison staff and other inmates. Guards patrolling the prison grounds noted Sam as being passive. Other prisoners mention Sam as exhibiting superficial behaviour with them for the most part. The prisoner who shares a cell with Sam has referred to them as the most inflexible cell-mate they have ever had. In last week's parole hearing, Sam's responses indicated a cold character."*

Design and Procedure

The design and procedure were identical to that of Study 2.3A.

2.8.2. Results

A 2*2 within-subjects ANOVA was conducted to examine how variations in trait humanness (human nature or shared with robots) and trait desirability (desirable or undesirable) influenced participants' parole decisions. As illustrated in Figure 5, a significant main effect of trait desirability was found, $F(1, 135) = 409.60$, $p < .001$, $\eta_p^2 = .75$. Criminals described using undesirable traits ($M = 38.9$, $SE = 1.82$) were less likely to be granted parole than were criminals described using desirable traits ($M = 80.6$, $SE = 1.44$), 95% CI [-45.85, -37.68]. Participants tended to see criminals less deserving of parole by about 42 points on the 100-point scale when described with undesirable rather than desirable traits.

No significant main effect of trait humanness was found, $F(1, 135) = 2.782$, $p = .098$, $\eta_p^2 = .02$. Participants were no more likely to grant parole to criminals who were described using human nature traits ($M = 58.7$, $SE = 1.53$) than those described using traits shared with

robots (M = 60.8, SE = 1.33), 95% CI [-4.79, 0.41]. Unlike in Study 2.3A, no interaction effect between trait humanness and trait desirability was found, $F(1, 135) = 2.49$, $p = .117$, $\eta_p^2 = .018$.

2.9. Chapter 2 General Discussion

Across six highly powered and pre-registered studies, we examined the hypothesised causal relationship between trait-based dehumanization and harm. The dual model of dehumanization (Haslam, 2006) posits that individuals and groups are sometimes subtly dehumanized by being denied human character traits. To the extent that groups are dehumanized in this way, they are considered vulnerable to harm (Haslam, 2019; Haslam & Loughnan, 2016). The work of Bastian and colleagues (2013) has been cited in support of this claim. Bastian et al. (2013) reported that the less human traits participants attributed to criminals, the harsher the punishment participants endorsed for them.

We initially sought to replicate the dehumanization effect reported by Bastian et al. (2013) in a conceptually similar design. In Study 2.1A, we examined the relationship between animalistic and mechanistic dehumanization, as operationalised by Bastian and colleagues, and the harshness of punishment endorsed by participants. In both studies, we successfully replicated previous findings, demonstrating that our paradigm was sensitive to finding predictive relationships between trait-based dehumanization and harm, should they occur.

In Studies 2.2A and 2.2B, we investigated the extent to which the previously reported relationship between trait-based dehumanization and harm can be explained by the social desirability of the traits incorporated into the stimulus set. The dual model (Haslam, 2006) has previously been critiqued for failing to clearly distinguish evidence for trait-based dehumanization from evidence of negative evaluation (Bloom, 2022; Enock et al., 2021a; Over, 2021a, 2021b). In line with the dual model (Haslam, 2006), Bastian and colleagues (2013) operationalised animalistic dehumanization as a reduction in the extent to which participants viewed criminals as possessing traits such as sophistication and refinement. They operationalised mechanistic dehumanization as a reduction in the extent to which participants

viewed criminals as possessing traits like warmth and depth. As these human traits are all socially desirable, it is unclear whether harm was predicted by dehumanization or negative evaluation.

We incorporated undesirable human traits into our stimulus set (e.g., *bitter* and *spiteful*) to tease apart the influence of dehumanization and negative evaluation in predicting harm. If trait-based dehumanization explains harm, the previously reported relationship between dehumanization and punishment should remain even when undesirable human traits are incorporated into the stimulus set. If the previously reported relationship is better explained by negative evaluation, then trait desirability should moderate the relationship with punishment. In support of the latter claim, Studies 2.2A and 2.2B showed that the more desirable human traits participants attributed to criminals, the less harshly participants thought they should be punished. The more undesirable human traits participants attributed to criminals, the more harshly participants thought they should be punished. The amount of variance explained by the models in Studies 2.2A and 2.2B increased by an average of about 75% compared to their respective models in Studies 2.1A and 2.1B. Given that the inclusion of trait desirability was the main difference between the studies, this increase suggests a strong independent influence of this variable on how traits are attributed, further highlighting the importance of controlling for trait desirability when examining trait-based dehumanization.

In Studies 2.3A and 2.3B, we sought to further distinguish between these two competing hypotheses using an experimental manipulation. We described criminals in traits that varied in perceived humanness and sociality and measured the influence of these varying descriptions on participants' parole decisions. This experimental design allowed us to directly test the hypothesised causal relationship between trait-based dehumanization and punishment. Converging with the findings of Study 2.2, we found that criminals described with undesirable traits were less likely to be granted parole than criminals described with desirable traits, regardless of whether or not those traits were uniquely human. There was no evidence for the hypothesis that criminals described with uniquely human terms would be more likely to be granted parole.

These findings fit with broader critiques of social-psychological models of dehumanization. Enock and colleagues (2021a) showed that what appears to be evidence for trait-based dehumanization of immigrants and political groups is better explained by negative evaluation. Similarly, Enock et al. (2021b) presented evidence that what appears to be emotion-based dehumanization of seven different outgroups is better explained by intergroup preference, another effect driven by the confound of social desirability. In these studies, participants were more likely to attribute prosocial emotions to the ingroup regardless of whether or not they were uniquely human. Participants were also more likely to attribute antisocial emotions to the outgroup, regardless of whether or not they were uniquely human. In further work, Enock and Over (2022) presented evidence that the apparent relationship between emotion-based dehumanization and reductions in prosocial behaviour is better explained by negative evaluation.

An interesting trend that emerged in Study 2.3A was that criminals were less likely to be granted parole when they were described with undesirable uniquely human traits compared to undesirable traits shared with other animals. In direct contrast to the predictions of infrahumanization theory, this suggests that shared undesirable traits (e.g., uncultured, stupid) might have less of an impact or be seen as less incriminating than undesirable uniquely human traits (e.g., corrupt, arrogant). This difference may explain the significant main effect of trait humanness found in Study 2.3A. Note that we did not see an analogous interaction when investigating mechanistic dehumanization in Study 2.3B. This difference between studies was not predicted and consequently needs to be treated with caution. Further research could examine in more detail whether seeing someone as possessing essentially undesirable human traits might garner more hostility than those seen as possessing undesirable traits shared with other animals. One possibility is that it is more vilifying to be seen as machine-like, possessing traits like ‘cold’ and ‘superficial’ than being seen as animal-like.

It is important to acknowledge that we considered only one target group in this study – criminals. We based this decision on the influence the findings of Bastian et al. (2013) have had on the literature. However, there may be more evidence for the hypothesised causal

relationship between trait-based dehumanization and harm in other intergroup contexts. In addition to examining additional intergroup contexts, future research should also incorporate more trait terms into stimulus sets. Research on dehumanization has been critiqued for using relatively small stimulus sets (Vaes, 2023). Indeed, some studies have used a single trait term to assess dehumanization. For example, Leidner et al. (Leidner et al., 2013) measured dehumanization by asking participants to rate the extent to which they agreed that members of the target outgroup experienced compassion. It will always remain possible that evidence for the causal relationship between dehumanization and harm could be found with a more sensitive paradigm.

We are not trying to argue that trait-based dehumanization never occurs. Instead, our argument is considerably more modest. Taken in conjunction with other recent results, it is apparent that evidence for trait-based dehumanization has often been confounded with evidence for negative evaluation (Bloom, 2022; Enock et al., 2021a, 2021b; Over, 2021a, 2021b). The results of the current study add to this growing body of critiques by showing that the findings of Bastian et al. (2013), often cited as evidence for the claim that trait-based dehumanization leads to an increased risk of harm, are considerably less convincing than they first appear. It is imperative that future research tests whether there is evidence for trait-based dehumanization when trait desirability is controlled for, given the grave importance of understanding predictors of intergroup harm in the real world.

Chapter 3: Infrahumanization Following Harm or Recognising Negative Affect when Harmed?

3.1. Abstract

Previous research has suggested that subtle dehumanization can occur as a consequence of harming others. According to this research, participants who feel a sense of collective responsibility for harming an outgroup may perceive members of that group as lacking in uniquely human emotions (Castano & Giner-Sorolla, 2006; Čehajić et al., 2009). I sought to understand whether these results replicate in two novel intergroup contexts: understanding UK residents' perceptions of those who experience the harmful effects of climate change in Nigeria and employees working in exploitative conditions in fast fashion supply chains. Across three well-powered and pre-registered studies, I successfully manipulated perceived harm against the target groups and perceived ingroup responsibility for that harm. However, I found no evidence for the previously reported relationship between ingroup responsibility for harm and the attribution of uniquely human emotions towards target outgroup members. Instead, a more straightforward pattern was found in all three studies, whereby participants perceived outgroup members to experience negative emotions to a greater extent when harm towards them has been made salient. The often-cited relationship between the need to justify harm and the tendency to subtly dehumanize harmed outgroup members may be less common than is often assumed.

3.2. Introduction

According to infrahumanization theory, a prominent emotion-based model of subtle dehumanization, outgroups are dehumanized to the extent that they are seen as lacking in uniquely human emotions compared to one's ingroup (Leyens et al., 2000, 2001). Also known as secondary emotions, uniquely human emotions are considered complex and long-lasting, including nostalgia, pride, shame, and euphoria. In contrast, emotions shared with other animals (also known as primary emotions) are considered more primal and fleeting, representing experiences humans perceive as sharing with other animals, such as happiness, fear, sadness, and surprise. The distinction between uniquely human emotions and those we share with nonhuman animals is based on lay perceptions of human and animal emotion experiences, with some empirical evidence suggesting that people spontaneously categorise emotions along this dimension (Rodríguez-Torres et al., 2005). Empirical work has suggested that we infrahumanize many kinds of outgroups, including regional and national ones (Cortes et al., 2005; Paladino et al., 2002; Rodríguez-Pérez et al., 2011), immigrants (Banton et al., 2020; Costello & Hodson, 2009, 2011), refugees (Azevedo et al., 2021; Gómez-Martínez & de la Villa Moral-Jiménez, 2018), and members of minority ethnic groups (Bruneau et al., 2020; De Dreu et al., 2011; DeLuca-McLean & Castano, 2009).

One reason researchers across many disciplines are interested in the construct of dehumanization is that it is thought to be causally related to harm. Philosophers have noted that genocide is often preceded by propaganda in which the members of the target outgroup are explicitly described as rats, lice, and parasites (Smith, 2011; Tirrell, 2012). Such blatant forms of dehumanization have also been linked to increased hostility in lab-based work (Kteily & Bruneau, 2017). Kteily and Bruneau (2017) reported that the extent to which US Americans endorse the claim that Arabs are "less than human" predicted hostility towards Arab immigration.

Subtle forms of dehumanization are also hypothesised to be causally related to harm (Bruneau et al., 2020; Civitillo et al., 2022; Goff et al., 2008; Maoz & McCauley, 2008; Obermann, 2011; Rudman & Mescher, 2012; Viki et al., 2012, 2013; Zebel et al., 2008). For

example, Cuddy et al. (2007) asked participants to estimate the extent to which Hurricane Katrina survivors of their racial ingroup and outgroup experienced uniquely human emotions such as grief, sorrow, and mourning, as well as emotions shared with other animals such as anger, panic, and rage. The less their participants attributed uniquely human emotions to outgroup members, the less willing they were to offer them help. In related experimental work, Vaes et al. (2003) found that participants responded more prosocially to a stranger who expressed a uniquely human emotion (*disappointment*) than to a stranger who expressed an emotion shared with other animals (*anger*).

Recently, the hypothesised causal relation between dehumanization and harm has been questioned (Bloom, 2017, 2022; Manne, 2016, 2018; Over, 2021a, 2021b). Individuals and groups may often be harmed because of uniquely human qualities they are thought to possess. Perceiving a group to be enemies, criminals, or traitors, terms that make the most sense when applied to humans, might motivate harm or oppression towards members of that group (Lang, 2010; Manne, 2016; Over, 2021a, 2021b; Rai & Fiske, 2011). Lab-based research has shown that the apparent causal relationship between infrahumanization and social behaviour can often be explained by a confound in the stimuli used. It has been suggested that the uniquely human qualities incorporated into stimulus sets are often more prosocial than those shared with other animals (Bunce et al., 2024; Over, 2021b). When appropriate controls are put in place, lowered attributions of prosocial and socially desirable qualities predict a reduction in prosocial behaviour and increases in antisocial behaviour towards outgroup members, not lowered attributions of uniquely human qualities (Enock et al., 2021a, 2021b; Enock & Over, 2022, 2023).

However, researchers have pointed out that dehumanization may be related to harm in other ways (Vaes et al., 2021). Perpetrators may justify the harm they have caused by coming to perceive their victims as less than human. The sociologist Luft (2019) suggests that dehumanization might sometimes manifest as a result of violence rather than a predecessor to it. Based on research into cases of extreme intergroup violence, including the Rwandan genocide and the Holocaust, Luft argues that dehumanizing language may serve to reinforce

social norms surrounding ongoing violence and justify the violence already undertaken by perpetrators. In particular, dehumanizing discourse may emerge to help alleviate the trauma and guilt that perpetrators feel in response to killing fellow humans (Luft, 2015). In support of this analysis, perpetrators of extreme intergroup harm sometimes refer to their harmed targets as less than human (Hatzfeld, 2003; Smith, 2011, 2021b).

The hypothesis that dehumanization may occur as a consequence of, or response to, harm has also been investigated with more subtle forms of dehumanization in lab-based settings. In a seminal paper in this field, Castano and Giner-Sorolla (2006) investigated whether participants would be more likely to deny outgroup members uniquely human emotions when they believed their ingroup was responsible for causing the outgroup harm. In one study, Castano and Giner-Sorolla presented participants with a scenario in which humans were either responsible for killing a large number of extra-terrestrial aliens or not. Participants were then asked to rate the extent to which they thought the aliens experienced 59 emotions on a 5-point Likert scale. Participants' emotion attributions were correlated with previously reported data on the extent to which these emotions are perceived as unique to humans (Demoulin et al., 2004). Results suggested that when participants felt responsible for their ingroup harming members of an outgroup, they attributed uniquely human emotions to them to a lesser extent than when they did not feel responsible. Castano and Giner-Sorolla (2006) broadly replicated this effect of infrahumanization following harm in two further studies using real-world contexts. These post-harm contexts involved British participants rating the emotion typicality of Indigenous Australians, and European US Americans rating the emotion typicality of Native Americans.

Further evidence for a relationship between collective responsibility for harm and infrahumanization was provided by Čehajić et al. (2009). Serbian participants were asked to estimate the extent to which Bosnian Muslims experienced 16 emotions. In one condition, a sense of collective responsibility for the genocide of Bosniaks in the early 1990s was made salient, while in the other two conditions, it was not. For the emotion ratings, four emotions were unique to humans and positive to experience (*tenderness*, *hope*, *admiration*, and *love*),

four were unique to humans and negative to experience (*remorse*, *guilt*, *shame*, and *resentment*), four were shared with other animals and positive to experience (*happiness*, *pleasure*, *euphoria*, and *joy*) and four were shared with other animals and negative to experience (*sadness*, *disgust*, *anger*, and *fear*). Participants rated Bosnian Muslims as typically feeling uniquely human emotions to a lesser extent when ingroup responsibility for harm was made salient compared to when it was not. These results were replicated in another context by Čehajić and colleagues, whereby European Chileans rated Indigenous Chileans as experiencing uniquely human emotions to a lesser extent when encouraged to feel responsible for the harm European colonialism brought upon the Native peoples of Chile. Overall, the studies described by Castano and Giner-Sorolla (2006) and Čehajić et al. (2009) reflect an apparent motivation to alleviate negative affect in response to harming fellow human beings by attributing fewer uniquely human emotions to them.

However, these pieces of research claiming to provide evidence of infrahumanization following harm are not without limitations. For instance, the crucial interaction between group condition and emotion humanness did not always reach significance, such as in Study 1 in Castano and Giner-Sorolla (2006) and Studies 1 and 2 in Čehajić et al. (2009). However, the researchers broke down these nonsignificant interactions and treated the trends found within as evidence for infrahumanization following harm. Each piece of research also had sampling limitations. Castano and Giner-Sorolla (2006) recruited relatively small samples of psychology undergraduates in their first two studies, as did Čehajić and colleagues (2009) in their first study. The sample collected by Čehajić and colleagues (2009) for their Study 2 comprised teenage students at a specific secondary school. The frequency of nonsignificant trends and lack of diverse sampling in the two most influential papers that claim to show evidence of infrahumanization following harm warrant an examination of how replicable this finding is in other contexts.

Considering the above concerns, as well as recent theoretical and empirical critiques of infrahumanization theory and its role in intergroup harm more broadly, we sought to measure whether we could replicate the occurrence of infrahumanization in response to harm

in two contemporary contexts – the harm caused to people by climate change and the harm caused to people by the fast fashion industry. Across three well-powered and pre-registered studies, we sought to test whether the previously reported relationship between feeling collective responsibility for harming an outgroup and the infrahumanization of its members replicates in two contemporary intergroup contexts.

Following previous research, we compare participants' performance across three conditions. In the *harm responsible condition*, we inform participants that the outgroup has experienced harm and that the ingroup is responsible for causing that harm. We compare emotion attribution in this condition to two other conditions. In the *harm not responsible condition*, we informed participants that the outgroup had experienced the same harm but that responsibility for the harm fell on a third party unrelated to the participants' ingroup. In the control *no harm condition*, we describe members of the same outgroup without any mention of harm. Regarding the predictions of previous research employing infrahumanization theory (Castano & Giner-Sorolla, 2006; Čehajić et al., 2009), the crucial comparison is between the two harm conditions. If infrahumanization occurs as a response to collective harm, then participants should attribute uniquely human emotions to the outgroup more strongly in the harm not responsible condition than in the harm responsible condition.

3.2.1. Note on Methods

All studies received ethical approval from the Psychology Departmental Ethics Committee at the University of York (Ethics ID no. 127). All data collection occurred online, and the studies were created and administered using Qualtrics (https://www.qualtrics.com). Participants were recruited through the online platform Prolific (https://www.prolific.co), with an independent sample recruited for each study. In each study, we incorporated manipulation checks to ensure that participants believed the group had been harmed and to ensure that participants felt collective responsibility for this harm. Informed consent was obtained at the start of each session according to approved ethical guidelines. Participants were rewarded at an approximate rate of £9 per hour in all studies. Assumption testing and subsequent analyses

were conducted using *SPSS* and *RStudio*. All post hoc tests were Bonferroni corrected for multiple comparisons. Highly influential cases were identified using Cook's distance in each study. All Studies were pre-registered on AsPredicted.com before commencing data collection. The vignettes used in each study, main effects and detailed breakdowns of interactions that are not theoretically relevant for Studies 3.1 and 3.2, and results of alternative analyses when testing the two hypotheses of Study 3.3 can be seen in Appendix B: Supplementary Materials for Chapter 3.

3.3. Study 3.1: Testing for Infrahumanization of Outgroup Members Harmed by Climate Change

In Study 3.1, we examined whether UK residents perceived residents of Lagos, Nigeria, as typically experiencing uniquely human emotions to a lesser extent when they felt collective responsibility for the harm they experience due to climate change, following the same design as Čehajić and colleagues (2009). This intergroup context was chosen because Lagos is a region considered particularly vulnerable to the consequences of climate change, with locals already experiencing some harmful effects (Mixed Migration Centre, 2023; Opeyemi, 2020). Thus, from the participant's perspective, the connection between their individual actions as members of an advantaged group and the current harm experienced by members of a disadvantaged outgroup is arguably more relatable and relevant than historical cases of intergroup harm, as examined in previous work.

We asked participants to read paragraphs about the residents of Lagos and then rate the extent to which residents of Lagos typically experienced the same selection of 16 different emotions used by Čehajić and colleagues (2009). The content of the paragraphs differed between the three conditions. In the harm responsible condition, the responsibility of UK residents for the harm experienced by the residents of Lagos due to climate change was emphasised. In the harm not responsible condition, the responsibility of multinational oil companies for this harm was emphasised. In the no harm condition, participants read a

paragraph about residents of Lagos in which no harm or effects of climate change were mentioned.

We sought to measure whether participants would rate residents of Lagos as typically experiencing uniquely human emotions to a lesser extent when participants felt a sense of collective responsibility for the harm caused to the target group by climate change, compared to when participants did not feel responsible for that harm or when no harm was emphasised.

3.3.1. Method

Participants

A power analysis using G*Power indicated that a sample size of 249 would allow us to detect a medium effect size (η_p^2 = .06) with an alpha of .05 and power of .95 using a mixed ANOVA. A final sample of 252 participants was included to allow for an equal number of 84 participants between condition groups. All participants were adult UK nationals currently residing in the UK who were fluent in English and had an approval rating of at least 95% on Prolific. Data from two participants were omitted and replaced as they failed attention checks. Data from two further participants were excluded as they were mistakenly recruited after the pre-registered sample size of 252 participants had been met. Including data submitted by excess participants in analyses did not change the results. Participants' ages ranged from 18 to 79 (M = 41.3, SD = 15.9), and 158 identified as female, 92 as male and two as nonbinary. An effort was made to ensure the sample was balanced regarding participants' political orientation, with 127 indicating they were left-leaning and 125 right-leaning.

Materials

Vignettes. There were three between-subject conditions: *Harm responsible*, *Harm not responsible* and *No harm*. In the two harm conditions, participants first read the same four paragraphs detailing the extent to which climate change has harmed the residents of Lagos, Nigeria. Those in the harm responsible condition then read further paragraphs detailing the contribution of UK residents to climate change. In contrast, those in the harm not responsible condition read further paragraphs detailing the contribution of multinational oil companies.

Participants in the *no harm* condition read paragraphs describing life for the residents of Lagos without mentioning climate change or harm. The vignettes can be seen in Appendix B: Supplementary Materials for Chapter 3.

Manipulation checks. All participants completed the same two manipulation checks. Participants were first asked, *"How harmful do you think climate change is to the residents of Lagos?*". This manipulation check appeared after the paragraphs describing Lagos, with or without mentioning the harmful effects of climate change in the region. It was included to ensure that participants in the two harm conditions perceived the residents of Lagos as experiencing more harm than participants in the no harm condition.

The second manipulation check asked, "*How responsible do you think UK citizens are for the effects of climate change on residents of Lagos?"* which appeared after the second set of paragraphs for participants in the two harm conditions. This check was included to ensure that a greater sense of collective responsibility for the harm experienced by the residents of Lagos due to climate change was felt by participants in the harm responsible condition than those in the harm not responsible condition. Participants in the no harm condition also responded to this manipulation check, which was presented alongside the *perceived harm* manipulation check once they read the single set of paragraphs in this condition.

Both manipulation checks were responded to using unmarked sliding scales ranging from 0 (*Not at all*) to 100 (*Very much so*), with the sliders initially fixed at the midpoint of the scale (*Somewhat*).

Outgroup emotion typicality ratings. Following the between-subjects experimental manipulation, all participants rated the extent to which residents of Lagos typically experience 16 different emotions. The selection of emotions was the same as used by Čehajić et al. (2009), consisting of four uniquely human emotions that are positive to experience (*tenderness, hope, admiration, love*), four uniquely human emotions that are negative to experience (*remorse, guilt, shame, resentment*), four emotions shared with other animals that are positive to experience (*happiness, euphoria, pleasure, joy*), and four emotions shared with other animals that are negative to experience (*sadness, disgust, anger, fear*). We chose the

emotions used by Čehajić et al. (2009) because their design closely mirrors that used in many other studies of infrahumanization (e.g., Leyens et al., 2000, 2001; Paladino et al., 2002; Vaes et al., 2003; Viki & Abrams, 2003) and because personal communication with Castano and Giner-Sorolla revealed these researchers no longer had access to the specific list of stimuli used in their previous work (2006). Participants responded to the extent to which they believed the outgroup to experience each item using an unmarked sliding scale ranging from 0 (*Not at all*) to 100 (*Very much so*), with the sliders initially fixed at the midpoint of the scale ('*Somewhat'*). The 16 emotion items were presented in a randomised order. An attention check (*Residents typically: please indicate 'not at all'*) appeared approximately midway through this scale.

Design

This study had a 3 (Condition: harm responsible, harm not responsible, and no harm; between) X 2 (Emotion humanness: uniquely human and shared with other animals; within) X 2 (Emotion valance: positive and negative; within) mixed design. Data were analysed using a mixed ANOVA for the main analysis.

Procedure

Three separate versions were created on Prolific/Qualtrics, one for each experimental condition. Data collection for each condition took place separately, allowing us to ensure participants only participated in one condition.

Interested individuals were informed on Prolific that the study aimed to examine how people ascribe emotions to different groups of individuals. After providing informed consent, participants answered brief demographic questions, including gender identity, age, political orientation, and checks for the inclusion criteria. Participants then proceeded to the experimental stimuli.

Participants in the two harm conditions read four paragraphs describing harm to residents of Lagos due to climate change and then responded to the first manipulation check measuring their perceived extent of harm. Participants in these two harm conditions then read four more paragraphs intended to manipulate their sense of collective responsibility for the

harm. Following this, participants responded to the second manipulation check, measuring participants' sense of collective responsibility for the harm.

Participants in the no harm condition read four paragraphs describing aspects of life for residents of Lagos without mentioning climate change or responsibility for harm. These participants then responded to the same two manipulation checks one after another because there was only one set of paragraphs in this condition. Each paragraph was presented for at least eight to ten seconds (without an upper limit) to ensure participants read each section carefully.

After reading the respective paragraphs for the experimental manipulation, participants in all three conditions responded to the outgroup emotion typicality items. Following this, participants could provide optional feedback on the study and were debriefed before returning to Prolific to complete the submission and receive payment. The study took an average of five and a half minutes to complete.

3.3.2. Results

Manipulation Checks

The Perceived Extent of Harm. A one-way between-subjects ANOVA revealed a significant difference in the perceived extent of harm experienced by residents of Lagos as a result of climate change, $F(2, 249) = 25.09$, $p < .001$, $\eta^2 = .17$. Post hoc comparisons indicated no significant difference between the harm responsible ($M = 89.4$, $SE = 1.93$) and the harm not responsible conditions ($M = 86.3$, $SE = 1.93$), $p = .77$, 95% CI [-3.46, 9.68]. However, the perceived extent of harm was significantly lower in the no harm condition ($M = 71.4$, $SE = 1.93$) than in the harm responsible condition, $p < .001$, 95% CI [-24.63, -11.49] or in the harm not responsible condition, $p < .001$, 95% CI [-21.52, -8.38]. Participants in the two harm conditions tended to see residents of Lagos as experiencing greater harm, by about 17 points on the 100-point scale, than those in the no harm condition. Therefore, the perceived extent of harm was manipulated as intended between conditions in Study 3.1, as illustrated in the left panel of Figure 6.

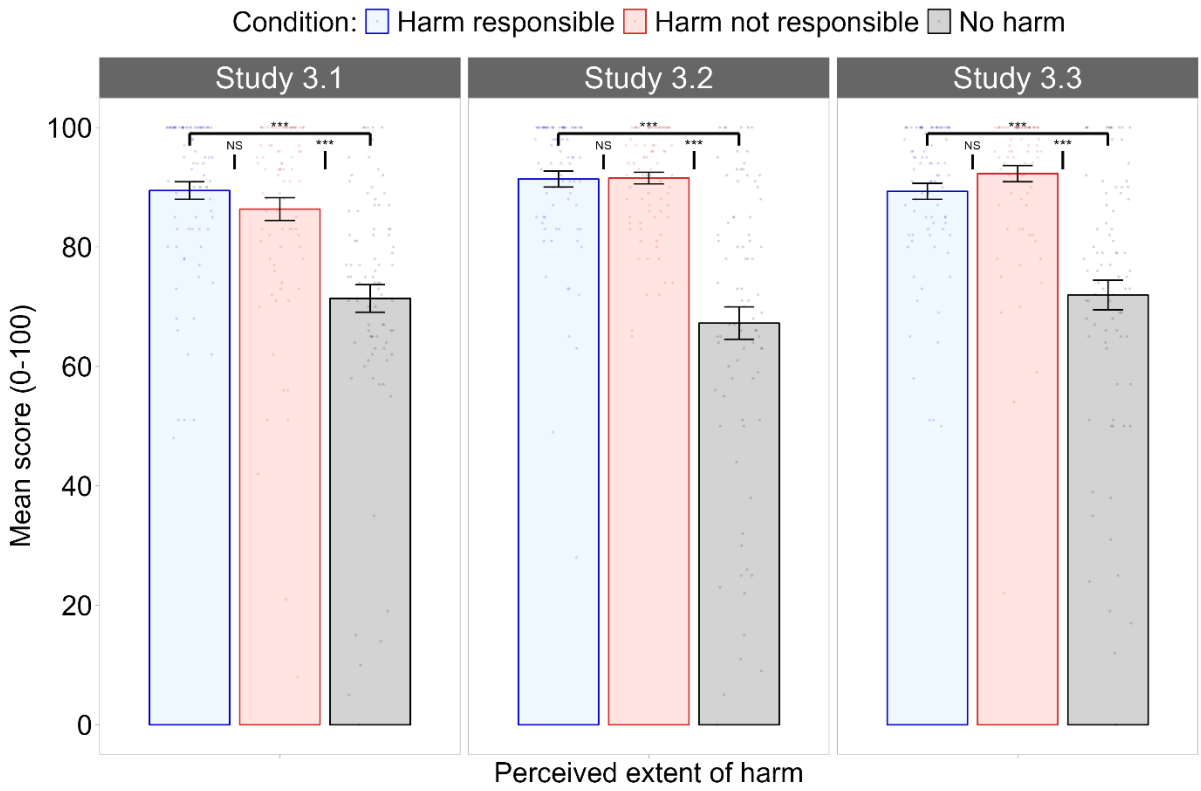

Figure 6. Mean scores for perceived extent of harm in Studies 3.1, 3.2, and 3.3

Perceived extent of harm was manipulated as intended in Study 3.1 (left panel), Study 3.2 (middle panel), and Study 3.3 (right panel). Participants in the harm responsible and harm not responsible conditions perceived outgroup members as experiencing significantly more harm than participants in the no harm condition.

Note: Error bars represent ±1 *SE*. *** denotes $p < .001$; *NS* denotes non-significant.

Sense of Collective Responsibility. A one-way between-subjects ANOVA revealed a significant difference between conditions in the extent to which participants felt a sense of collective responsibility for the harm caused by climate change, $F(2, 249) = 17.84$, $p < .001$, $\eta^2 = .13$. Post hoc comparisons revealed that participants in the harm responsible condition felt significantly more collective responsibility for the harm ($M = 63.2$, $SE = 2.81$) than did those in the harm not responsible condition ($M = 44.2$, $SE = 2.81$), $p < .001$, 95% CI [9.44, 28.59], or those in the no harm condition ($M = 41.4$, $SE = 2.81$), $p < .001$, 95% CI [12.22, 31.37]. There was no significant difference in collective responsibility between the harm not

responsible and no harm conditions, $p = 1$, 95% CI [-6.79, 12.36]. Participants' sense of collective responsibility for harm tended to be about 19 points higher on the 100-point scale in the harm responsible condition than in the harm not responsible condition. Therefore, this variable was manipulated as intended between conditions, as seen in the left panel of Figure 7. Taken together, results from the manipulation checks suggest that this is an intergroup context in which we should see evidence of infrahumanization similar to that reported in previous research (Castano & Giner-Sorolla, 2006; Čehajić et al., 2009).

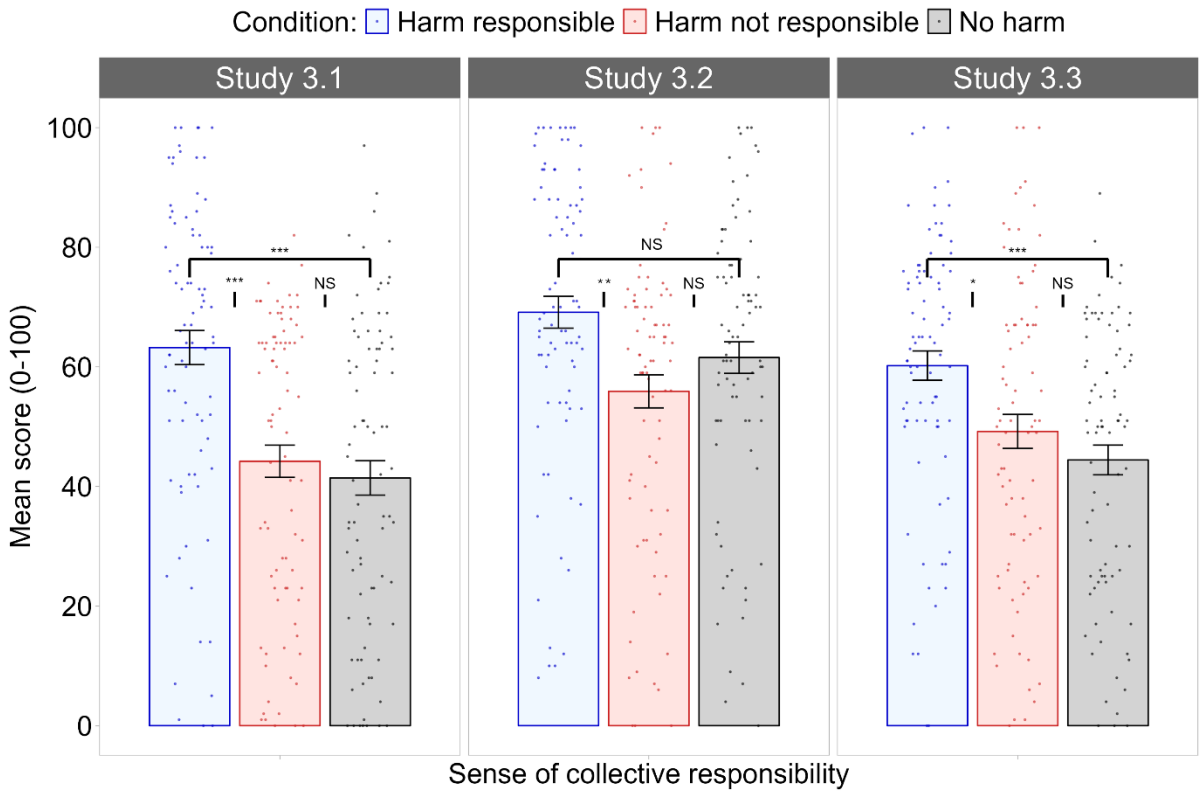

Figure 7. Mean scores for sense of collective responsibility in Studies 3.1, 3.2, and 3.3

Sense of collective responsibility was manipulated as intended in Study 3.1 (left panel), Study 3.2 (middle panel), and Study 3.3 (right panel). Participants in the harm responsible condition felt that their ingroup was significantly more responsible for the harm experienced by outgroup members than participants in the harm not responsible condition.

Note: Error bars represent ±1 *SE*. *** denotes $p < .001$; ** denotes $p < .005$; * denotes $p < .05$; *NS* denotes non-significant.

Emotion Ratings

We conducted a 3 (Condition: harm responsible, harm not responsible, no harm; between) X 2 (emotion humanness: uniquely human, shared with other animals; within) X 2 (emotion valence: positive, negative; within) mixed ANOVA to examine if the target group were perceived as experiencing uniquely human emotions to a lesser extent when ingroup collective responsibility for harm was made salient compared to when ingroup responsibility was not made salient or no harm against the target group was described. The results of Study 3.1 are illustrated in Figure 8. The main effects and a breakdown of the two-way interactions are reported in Appendix B: Supplementary Materials for Chapter 3.

We found significant two-way interactions between harm condition and emotion humanness $F(2, 249) = 20.270$, $p < .001$, $\eta_p^2 = .14$, and between harm condition and emotion valence, $F(2, 246) = 57.714$, $p < .001$, $\eta_p^2 = .32$. The two-way interactions were each qualified by a significant three-way interaction between harm condition, emotion humanness, and emotion valence, $F(2, 249) = 19.209$, $p < .001$, $\eta_p^2 = .13$. This three-way interaction was broken down using Bonferroni corrected pairwise comparisons between the three conditions at each level of emotion humanness and emotion valence.

Positive Uniquely Human Emotions. The extent to which residents of Lagos were thought to experience positive uniquely human emotions did not differ between the harm responsible condition ($M = 53.9$, $SE = 1.61$) and the harm not responsible condition ($M = 50.7$, $SE = 1.61$), $p = .475$, 95% CI [-2.27, 8.71]. Residents of Lagos were seen as typically experiencing positive uniquely human emotions to a greater extent in the no harm condition ($M = 64.8$, $SE = 1.61$) than in either the harm responsible condition, $p < .001$, 95% CI [5.36, 16.34], or the harm not responsible condition, $p < .001$, 95% CI [8.58, 19.57]. These results suggest that residents of Lagos were seen as typically experiencing positive uniquely human emotions to a lesser degree of about 12 points on the 100-point emotion typicality scale when described as being harmed compared to when they were not, irrespective of whether participants felt collectively responsible for the harm.

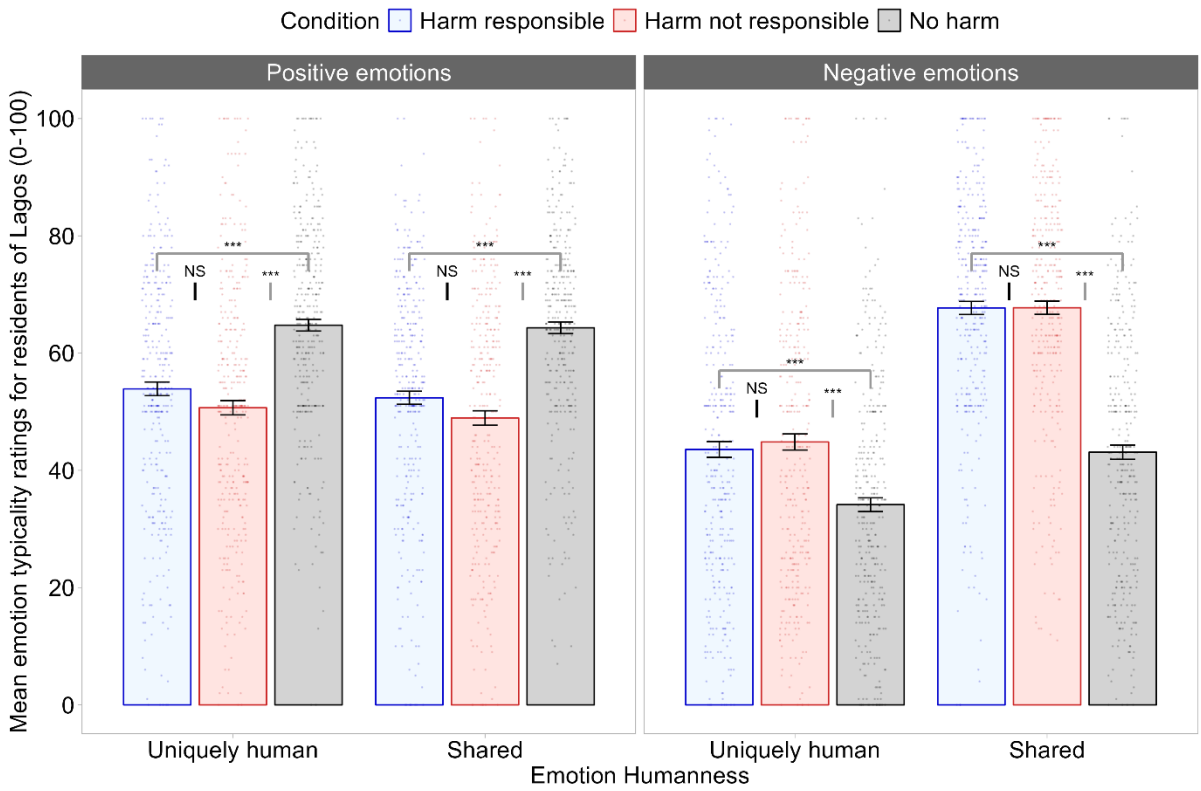

Figure 8. Results of Study 3.1

Mean emotion attributions by harm condition, emotion humanness, and emotion valence in Study 3.1 are shown. We observed no evidence that collective responsibility for harm induced infrahumanization of the outgroup. Instead, residents of Lagos were thought to experience negative emotions to a greater extent and positive emotions to a lesser extent when harm against them was made salient.

Note: Error bars represent ±1 *SE*. Shared = emotions shared with other animals. *** denotes $p < .001$; *NS* denotes non-significant.

Negative Uniquely Human Emotions. The extent to which residents of Lagos were thought to experience negative uniquely human emotions did not differ between the harm responsible ($M = 43.6$, $SE = 1.66$) and harm not responsible ($M = 44.8$, $SE = 1.66$) conditions, $p = 1$, 95% CI [-6.95, 4.39]. Residents of Lagos were seen as typically experiencing negative uniquely human emotions to a lesser extent in the no harm condition ($M = 34.1$, $SE = 1.66$) than in the harm responsible condition, $p < .001$, 95% CI [-15.09, -3.75], or in the harm not

responsible condition, $p < .001$, 95% CI [-16.37, -5.03]. When residents of Lagos were described as harmed, participants saw them as typically experiencing negative uniquely human emotions to a greater degree of about 10 points on the 100-point emotion typicality scale, compared to when they were not described as being harmed. As with positive uniquely human emotions, participants' sense of collective responsibility for the harm did not seem to affect how negative uniquely human emotions were perceived.

Positive Shared Emotions. A similar pattern emerged for positive emotions shared with other animals as with positive uniquely human emotions. There was no significant difference in the perceptions of the experience of positive shared emotions between the harm responsible condition ($M = 52.4$, $SE = 1.82$) and the harm not responsible condition ($M = 48.9$, $SE = 1.82$), $p = .545$, 95% CI [-2.75, 9.63]. Participants rated residents of Lagos as typically experiencing positive emotions shared with other animals to a greater extent in the no harm condition ($M = 64.3$, $SE = 1.82$) than in the harm responsible condition, $p < .001$, 95% CI [5.75, 18.13], or in the harm not responsible condition, $p < .001$, 95% CI [9.12, 21.57]. Similar to positive uniquely human emotions, participants' sense of collective responsibility for the harm did not affect how residents of Lagos were seen to typically experience positive shared emotions. When participants perceived residents of Lagos as experiencing harm, they saw them as typically experiencing positive shared emotions to a lesser degree of about 14 points on the 100-point emotion typicality scale compared to when the outgroup members were not seen as experiencing harm.

Negative Shared Emotions. A similar pattern emerged for negative emotions shared with other animals and negative uniquely human emotions. No significant difference in ratings of negative shared emotions was found between the harm responsible condition ($M = 67.7$, $SE = 1.85$) and the harm not responsible condition ($M = 67.7$, $SE = 1.85$), $p = 1$, 95% CI [-6.32, 6.26]. Participants rated residents of Lagos as typically experiencing negative emotions shared with other animals to a lesser extent in the no harm condition ($M = 43.1$, $SE = 1.85$) than in the harm responsible condition, $p < .001$, 95% CI [-30.89, -18.31], or in the harm not responsible condition, $p < .001$, 95% CI [-30.92, -18.34]. Participants' sense of collective

responsibility for harm did not influence the extent to which they thought residents of Lagos typically experienced negative shared emotions, a trend found for all four emotion categories. Like negative uniquely human emotions, residents of Lagos were seen as typically experiencing negative shared emotions to a greater degree of about 25 points on the 100-point outgroup emotion typicality scale when they were described as harmed compared to when they were not.

3.3.3. Study 3.1 Discussion

Our pattern of results does not align with previous research suggesting that a sense of collective responsibility for harm leads to the infrahumanization of the victim group, whereby individuals subtly dehumanize members of an outgroup they believe their ingroup has harmed (Castano & Giner-Sorolla, 2006; Čehajić et al., 2009). Rather, our data show a somewhat more straightforward pattern. Participants estimated that the residents of Lagos were likely to feel negative emotions to a greater extent and positive emotions to a lesser extent when harm to this group was made salient, regardless of the humanness of the emotions and regardless of ingroup responsibility for this harm. This pattern is in striking contrast to previous findings.

While we did not replicate the central finding of previous research, our manipulation checks showed that we successfully manipulated perceived harm and perceived collective responsibility for that harm. These successful manipulations suggest that we should have been able to detect infrahumanization effects if they occur in this context.

3.4. Study 3.2: Testing for Infrahumanization of Outgroup Members Harmed by the Fast Fashion Industry

In Study 3.2, we sought to test whether collective responsibility for harm induced a tendency to infrahumanize the outgroup in another contemporary intergroup context. In this study, we examined whether UK residents rated textile workers in fast fashion supply chains as experiencing uniquely human emotions to a lesser extent when ingroup responsibility for harm was made salient compared to when it was not.

We chose this intergroup context for several reasons. The exploitation of textile workers is often referred to as dehumanizing (see: Ghani, 2020; Yousefi, 2020). Importantly, and like the climate change context used in Study 3.1, the fast fashion context provides the opportunity to frame the participant ingroup (UK residents) as either responsible or not responsible for the harm experienced by outgroup members. As in Study 3.1, this contemporary context of intergroup harm, where a participant's actions as a member of a privileged group are directly contributing to the harm experienced by members of a disadvantaged outgroup, is arguably more relatable for participants than historical cases of harm, or entirely hypothetical ones, as previous research has examined (Castano & Giner-Sorolla, 2006; Čehajić et al., 2009).

As in Study 3.1, UK residents participated in one of three between-subject conditions. In the two harm conditions, participants read a series of paragraphs about the extent to which textile workers experience exploitation in fast fashion supply chains. In the harm responsible condition, participants then read about the high rates of fast fashion consumption in the UK. In the harm not responsible condition, participants read about the negligence of CEOs and boards of directors in fast fashion companies who prioritise profit over safe and fair working conditions for textile workers. In the no harm condition, participants read a few paragraphs detailing the history and global presence of the fast fashion industry without mentioning the harm textile workers experience. If people tend to infrahumanize others when they feel responsible for harm against them (as is predicted by past work in infrahumanization theory), we would expect to see greater attribution of uniquely human emotions in the harm not responsible and no harm conditions than in the harm responsible condition. We included the same emotion items as in Study 3.1, the same selection used by Čehajić and colleagues (2009).

3.4.1. Method

Participants

The sample size estimation was the same as in Study 3.1. A sample of 252 participants was collected. All participants were adult UK nationals currently residing in the UK who were fluent in English and had an approval rating of at least 95% on Prolific. In the sample, 154 participants identified as female, 95 as male, two as nonbinary and one as unspecified. Ages ranged from 18 to 77 (M = 41.1, SD = 15.6). We sought to recruit participants across the political spectrum, with 127 identifying as left-leaning and 125 as right-leaning.

Materials

The manipulation checks measuring the perceived extent of harm and sense of collective responsibility, the 16-item outgroup emotion typicality scale and attention checks were the same as in Study 3.1, with the target outgroup and contextual details adapted to the fast fashion context.

Vignettes. As in Study 3.1, the stimuli varied between the three between-subject conditions, with data from 84 participants collected for each (*harm responsible*, *harm not responsible* and *no harm*). In the two harm conditions, participants read four paragraphs detailing the extent of harm that poor working conditions and exploitation have on textile workers along fast fashion supply chains. After these four paragraphs, those in the *harm responsible* condition read four paragraphs describing the growing demand for fast fashion products UK consumers have and how this is complicit with the harm experienced by textile workers, allowing it to continue. Participants in the *harm not responsible* condition instead read four paragraphs framing fast fashion companies' CEOs and boards of directors as responsible for the harm due to their prioritising profits over safe and fair working conditions for textile workers in the supply chains. Participants in the *no harm* condition read four paragraphs detailing the history of fast fashion and the number of textile workers the fast fashion industry employs. In this vignette, there was no mention of harm.

Procedure and Design

The procedure and design in Study 3.2 were identical to those used in Study 3.1.

3.4.2. Results

Manipulation Checks

The Perceived Extent of Harm. A one-way between-subjects ANOVA revealed a significant difference between the three conditions in the perceived extent of harm experienced by textile workers in fast fashion supply chains, $F(2, 249) = 57.99$, $p < .001$, $\eta^2 = .32$. Bonferroni-corrected post hoc comparisons revealed no significant difference in perceived harm between the harm responsible ($M = 91.4$, $SE = 1.84$) and harm not responsible conditions ($M = 91.5$, $SE = 1.84$), $p = .1$, 95% CI [-6.42, 6.09]. The perceived extent of harm was significantly lower among participants in the no harm condition ($M = 67.3$, $SE = 1.84$) than those in the harm responsible condition, $p < .001$, 95% CI [-30.37, -17.86], or the harm not responsible condition, $p < .001$, 95% CI [-30.54, -18.03]. Participants in the two harm conditions tended to see textile workers as experiencing greater harm, by about 24 points on the 100-point perceived harm scale, than participants in the no harm condition. Therefore, the extent of perceived harm was manipulated as intended between the three harm conditions. Mean scores for participants' perceived extent of harm in Study 3.2 are illustrated in the middle panel of Figure 6.

Sense of Collective Responsibility. A one-way between-subjects ANOVA revealed that the collective responsibility felt by participants significantly differed between conditions, $F(2, 249) = 6.089$, $p = .003$, $\eta^2 = .05$. Bonferroni-corrected post hoc comparisons revealed that participants in the harm responsible condition felt significantly more collective responsibility ($M = 69.1$, $SE = 2.69$) than those in the harm not responsible condition ($M = 55.9$, $SE = 2.69$), $p = .002$, 95% CI [4.06, 22.39]. Thus, participants' sense of collective responsibility was manipulated as intended between conditions as this tended to be higher among those in the harm responsible condition, by about 13 points more on the 100-point scale, than those in the harm not responsible condition. No significant differences in collective responsibility were found between the no harm condition ($M = 61.6$, $SE = 2.69$) and either the harm responsible condition, $p = .142$, 95% CI [-16.75, 1.58], or the harm not responsible condition, $p = .418$, 95% CI [-3.53, 14.81]. It is worth noting that without Bonferroni correction, participants' sense of collective responsibility was significantly higher in the harm responsible

condition than in the no harm condition (p = .047). Mean scores for participants' sense of collective responsibility in Study 3.2 are illustrated in the middle panel of Figure 7.

Emotion Ratings

We conducted a 3 (condition: harm responsible, harm not responsible, no harm; between) X 2 (emotion humanness: uniquely human, shared with other animals, within) X 2 (emotion valence: positive, negative, within) mixed ANOVA to examine if collective responsibility for harm influenced the extent to which participants perceived textile workers in fast fashion supply chains as typically experiencing uniquely human emotions and positive emotions. The main effects and a breakdown of the two-way interactions are reported in Appendix B: Supplementary Materials for Chapter 3. The results of Study 3.2 are illustrated in Figure 9. We found significant two-way interactions between harm condition and emotion humanness, $F(2, 249) = 5.647$, $p = .004$, $\eta_p^2 = .04$, and between harm condition and emotion valence, $F(2, 246) = 29.730$, $p < .001$, $\eta_p^2 = .19$. As in Study 3.1, these were qualified by a significant three-way interaction between harm condition, emotion humanness, and emotion valence, $F(2, 249) = 10.170$, $p < .001$, $\eta_p^2 = .08$. This three-way interaction was broken down using Bonferroni-corrected pairwise comparisons between the three conditions at each level of emotion humanness and emotion valence.

Positive Uniquely Human Emotions. The extent to which textile workers were thought to experience positive uniquely human emotions did not differ between the harm responsible condition (M = 32.6, SE = 2.00) and the harm not responsible condition (M = 28.9, SE = 2.00), p = .601, 95% CI [-3.18, 10.44]. Textile workers were seen as typically experiencing positive uniquely human emotions to a greater extent in the no harm condition (M = 42.30, SE = 2.00) than in the harm responsible condition p = .002, 95% CI [2.92, 16.54] or in the harm not responsible condition, $p < .001$, 95% CI [6.55, 16.54]. Textile workers were seen as typically experiencing positive uniquely human emotions to a lesser degree, about 11 points less on the 100-point emotion typicality scale, when described as being harmed compared to when they were not, irrespective of whether participants felt collectively responsible for the harm.

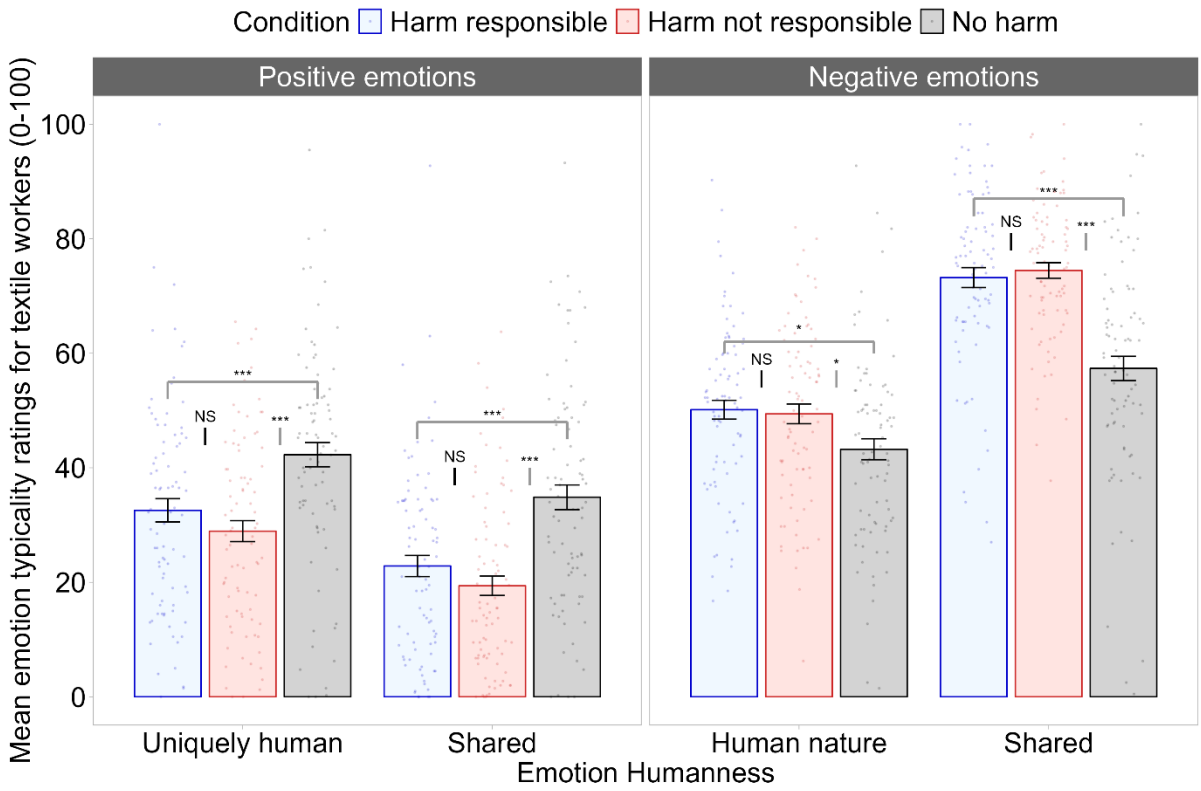

Figure 9. Results of Study 3.2

Mean emotion attributions by condition, emotion humanness, and emotion valence are displayed. We found no evidence that inducing a sense of collective responsibility for the harm experienced by textile workers leads participants to infrahumanize them. Converging with the results of Study 3.1, participants seemed to rate outgroup members as experiencing negative emotions more strongly when harm against them was made salient.

Note: Error bars represent ±1 *SE*. *Shared* = emotions shared with other animals. *** denotes $p < .001$; * denotes $p < .05$; *NS* denotes non-significant.

Negative Uniquely Human Emotions. The extent to which textile workers were thought to experience negative uniquely human emotions did not differ between the harm responsible ($M = 50.2$, $SE = 1.74$) and harm not responsible ($M = 49.4$, $SE = 1.74$) conditions, $p = 1$, 95% CI [-5.19, 6.64]. Textile workers were seen as typically experiencing negative uniquely human emotions to a lesser extent in the no harm condition ($M = 43.2$, $SE = 1.74$) than in either the harm responsible condition $p = .015$, 95% CI [-12.82, -0.99], or the harm not

responsible condition, p = .037, 95% CI [-12.10, -0.27]. When textile workers were described as harmed, participants saw them as typically experiencing negative uniquely human emotions to a greater degree, about 7 points on the 100-point emotion typicality scale, compared to when they were not described as being harmed. As with positive uniquely human emotions, participants' sense of collective responsibility for the harm did not seem to affect how negative uniquely human emotions were perceived.

Positive Shared Emotions. A similar pattern emerged for positive emotions shared with other animals as with positive uniquely human emotions. There was no significant difference in the perceptions of textile workers' experiences of positive shared emotions between the harm responsible condition (M = 22.9, SE = 1.91) and the harm not responsible condition (M = 19.4, SE = 1.91), p = .606, 95% CI [-3.05, 9.95]. Participants rated textile workers as typically experiencing positive emotions shared with other animals to a greater extent in the no harm condition (M = 34.9, SE = 1.91) than in the harm responsible condition, p < .001, 95% CI [5.51, 18.51], or in the harm not responsible condition, p < .001, 95% CI [8.96, 21.96]. When participants perceived textile workers as experiencing harm, they saw them as typically experiencing positive shared emotions to a lesser degree, about 14 points on the 100-point emotion typicality scale, compared to they were not seen as experiencing harm. This was irrespective of participants' sense of collective responsibility for the harm.

Negative Shared Emotions. A similar pattern emerged for negative emotions shared with other animals and negative uniquely human emotions. No significant difference in ratings of negative shared emotions was found between harm responsible condition (M = 73.2, SE = 1.77) and the harm not responsible condition (M = 74.5, SE = 1.77), p = 1, 95% CI [-7.26, 4.77]. Participants rated textile workers as typically experiencing negative emotions shared with other animals to a lesser extent in the no harm condition (M = 57.4, SE = 1.77) than in the harm responsible condition, p < .001, 95% CI [-21.89, -9.86], or in the harm not responsible condition, p < .001, 95% CI [-23.13 -11.10]. Participants' sense of collective responsibility for harm did not influence the extent to which they thought textile workers typically experienced negative shared emotions, a trend found for all four emotion categories.

Like negative uniquely human emotions, textile workers were seen as typically experiencing negative shared emotions to a greater degree of about 17 points on the 100-point outgroup emotion typicality scale when they were described as harmed compared to when they were not.

3.4.3. Study 3.2 Discussion

Using the context of harm inflicted on fast fashion textile workers, we did not replicate the previously reported relationship between feeling collective responsibility for harm and infrahumanization. However, we observed a pattern of results that was very similar to that in Study 3.1. When harm was made salient, workers in fast fashion supply chains were thought to experience negative emotions to a greater extent and positive emotions to a lesser extent, regardless of emotion humanness or ingroup responsibility.

There are several possible reasons for our failure to replicate previous findings. One possibility is that we happened to choose intergroup contexts in which infrahumanization does not occur. Previous research has examined the relationship between collective responsibility for harm and infrahumanization in response to hypothetical scenarios with aliens as well as in real-world contexts such as responses to atrocities committed during colonialism and genocide (Castano & Giner-Sorolla, 2006; Čehajić et al., 2009). We chose more contemporary and relatable intergroup contexts, focusing on the responsibility of UK residents for harm towards people affected by climate change and fast fashion textile workers. However, we successfully manipulated perceived harm and, when comparing the harm responsible and harm not responsible conditions, a sense of collective responsibility for harm, suggesting these are contexts where we would likely observe infrahumanization following harm if it occurs. Nevertheless, it remains possible that collective responsibility induces infrahumanization in some contexts but not others.

Another possibility is that the design we chose was insensitive to the effects of infrahumanization. Our design broadly mirrored that of Čehajić et al. (2009); however, infrahumanization research has been criticised more broadly for using a small number of

emotion terms in each condition (Enock & Over, 2022; Vaes, 2023). Castano and Giner-Sorolla (2006) asked participants to rate a larger number of emotion terms that vary continuously along the dimensions of humanness and valence. In Study 3.3, we returned to the intergroup context of climate change used in Study 3.1. However, we sought to test the relationship between collective responsibility and infrahumanization with a more sensitive design similar to that used by Castano and Giner Sorolla (2006).

3.5. Study 3.3: Testing for Infrahumanization of Outgroup Members Harmed by Climate Change Using Continuous Measures

In Study 3.3, we again manipulated perceptions of harm inflicted on residents of Lagos caused by climate change and collective responsibility for this harm. This time, rather than measuring infrahumanization using a small number of emotions clustered into discrete categories, we asked participants to rate how typically the outgroup members experience a much larger range of emotions (64 in total). Before the manipulation, we asked participants to rate these emotions on the dimensions of humanness and valence. Our initial intention was to use the same set of emotions used by Castano and Giner-Sorolla (2006). However, personal communication with the authors revealed no record of the exact stimuli they used. Instead, we selected 64 emotion terms based on previous research in the area, including Demoulin et al. (2004) and Enock et al. (2021b).

In this design, infrahumanization is said to be reflected in a negative relationship between ratings of emotion humanness and typicality of outgroup members such that the more uniquely human an emotion is perceived to be, the less it is ascribed to outgroup members (Castano & Giner-Sorolla, 2006; Vaes, 2023). This effect is thought to be independent of emotion valence. If collective responsibility for harm induces infrahumanization of the outgroup, then there should be a stronger negative relationship between emotion humanness and outgroup typicality in the harm responsible condition than in the harm not responsible condition.

However, we suggest an alternative hypothesis based on our findings of Studies 3.1 and 3.2. We suggest that victims of harm are perceived as experiencing negative emotions to a greater extent and positive emotions to a lesser extent regardless of emotion humanness or whether or not one feels collective responsibility for the harm. Thus, we expected a stronger negative relationship to emerge between emotion valance and outgroup emotion typicality in both the harm responsible and harm not responsible conditions than in the no harm condition.

3.5.1. Method

Participants

Our primary analysis was done using emotion items rather than participants. A sensitivity analysis using G*Power indicated that including 64 emotions in the regression models would allow us to detect a medium effect size ($f^2 = .15$), with an alpha of .05 and a power of .85. We recruited 252 participants (84 per condition) to ensure the average scores for emotion valence, humanness, and outgroup typicality were reliable. All participants were adult UK nationals currently residing in the UK who were fluent in English and had an approval rating of at least 95% on Prolific. Among participants, 132 identified as female, 117 as male, two as nonbinary, and one preferred not to indicate their gender identity. Ages ranged from 18 to 81 ($M = 43.9$, $SD = 15.2$). As in the previous studies, an effort was made to ensure the sample was balanced in political orientation, with 127 participants being left-leaning and 125 being right-leaning. Data submitted by five participants were omitted and replaced due to missing responses.

Design and Analysis Strategy

All participants responded to the same emotion humanness, emotion valence, and outgroup emotion typicality measures. Participants were randomly allocated to one of three between-subject conditions: harm responsible, harm not responsible, and no harm. This study, therefore, took a 3 (condition: harm responsible, harm not responsible, and no harm; between) X emotion humanness (continuous; within) X emotion valance (continuous; within) mixed design, with perceived emotional experiences of residents of Lagos as the dependent variable.

Previous work on infrahumanization following harm suggests a stronger negative relationship between emotion humanness and outgroup emotion typicality ratings should emerge in the harm responsible condition compared to the harm not responsible and no harm conditions, irrespective of emotion valence. However, we suggest that a stronger negative relationship between emotion valence and outgroup emotion typicality ratings should emerge in the two harm conditions compared to the no harm condition, irrespective of emotion humanness.

We followed the analysis plan Castano and Giner-Sorolla (2006) used to test these two hypotheses. We first created average scores for emotion humanness, emotion valence and outgroup emotion typicality for each condition separately. These average scores for individual emotions were then treated as observations in the main analyses, where we regressed average emotion humanness and valence scores onto outgroup emotion typicality ratings. In testing for infrahumanization following harm, our three models were:

1. **Outgroup typicality** (**DV) ~ emotion humanness** (**IV) in the harm responsible condition**
2. **Outgroup typicality** (**DV) ~ emotion humanness** (**IV) in the harm not responsible condition**
3. **Outgroup typicality** (**DV) ~ emotion humanness** (**IV) in the no harm condition**

The coefficients of each model were then compared with Fisher's Z transformation using the *cocor* package in *RStudio* (Diedenhofen & Musch, 2015; see also: Fisher, 1992; Zou, 2007). This process determined whether a significantly stronger negative relationship existed between emotion humanness and outgroup emotion typicality ratings for the harm responsible condition than the harm not responsible condition.

The above steps were repeated to test our alternative prediction, with emotion valence entered as the predictor variable rather than emotion humanness. Thus, the three models compared when testing our alternative prediction of negative affect driving emotion perception in harmed outgroup members were as follows:

4. **Outgroup typicality (DV) ~ emotion valence (IV) in the harm responsible condition**
5. **Outgroup typicality (DV) ~ emotion valence (IV) in the harm not responsible condition**
6. **Outgroup typicality (DV) ~ emotion valence (IV) in the no harm condition**

We expected a stronger negative relationship between emotion valence and outgroup typicality ratings in the two harm conditions compared to the no harm condition, regardless of emotion humanness.

To test for the effects of humanness independently of valence (models 1-3) and valence independently of humanness (models 4-6), we needed to ensure the two constructs were orthogonal in the data. Otherwise, we would need to control for the effects of one as a covariate when measuring the other and vice versa. Before conducting the regression analyses, we checked for correlations between ratings of emotion humanness and emotion valence. No significant correlations were found between emotion humanness and emotion valence among data submitted by participants in the harm responsible condition (r = -.03, p = .814), the harm not responsible condition (r = -.06, p = .628), or in the no harm condition (r = -.03, p = .817). These nonsignificant correlations imply that emotion humanness and emotion valence were orthogonal in our dataset. Thus, a nonlinear relationship between emotion humanness and emotion valence meant partial regressions were unnecessary (Freund et al., 2010). Therefore, a single predictor and outcome variable were entered into each regression model.

As in Studies 3.1 and 3.2, we also conducted one-way ANOVAs to check the data from the manipulation checks to be sure that the perceived extent of harm and sense of collective responsibility for the harm were manipulated as intended between the conditions.

Materials

Emotion Humanness. After briefly describing how some emotions might be considered unique to humans and some we share with other animals, participants were given the following prompt: *"To what extent do you think each emotion is experienced by humans*

and other animals equally, or by humans only?". Participants then rated each of the 64 emotions along the dimension of humanness using an unmarked sliding scale ranging from 0 (*Humans and other animals equally*) to 100 (*Humans only*), with the sliders initially fixed at the scale's midpoint. An attention check (*Please indicate 'Humans only')* was included roughly halfway through the scale.

Emotion Valence. After briefly describing how some emotions might be considered positive to experience and others negative to experience, participants were given the following instruction: *"How does each emotion make people feel?".* Participants then rated each of the 64 emotions on how positively or negatively valenced they were using an unmarked sliding scale ranging from 0 (*Extremely negative*) to 100 (*Extremely positive*). The sliders were initially fixed at the scale's midpoint (*Neither positive nor negative*) for each item. An attention check (*Please indicate 'Extremely negative'*) was included approximately halfway through the scale.

Filler Task. A filler task was included before the experimental stimuli to minimize carryover effects from the initial emotion-scoring tasks. An example of a *divergent uses task* (see Gilhooly et al., 2007; Guilford, 1967), the filler task involved asking participants to think of as many uses for a brick as possible in one minute, entering their responses into an open text box.

Vignettes. We used the same climate change vignettes for each condition described in Study 3.1, which can be seen entirely in Appendix B: Supplementary Materials for Chapter 3.

Outgroup Emotion Typicality Ratings. After reading the vignettes, all participants were asked to respond to 64 outgroup emotion typicality items. Each item read as "*Residents of Lagos typically feel X*", with X being one of the 64 emotions. Participants responded to all 64 emotion typicality items using an unmarked sliding scale ranging from 0 (*Not at all*) to 100 (*Very much so*), with the sliders initially fixed at the midpoint of the scale (*Somewhat*). The 64 emotion items were presented randomly, apart from an attention check (*Residents typically: please indicate 'not at all'*), which appeared approximately halfway through the task.

An effort was made to include a selection of emotions balanced across perceived humanness and valence based on pretest data collected by Enock et al. (2021b). A balanced sample meant that, on average, emotions considered unique to humans were significantly higher in humanness than those rated as shared with other animals, but these categories did not differ in how positively or negatively valenced they were. Likewise, the average valence of emotions rated as positive to experience was significantly more positive than those rated as negative to experience, but positive and negative emotions did not differ in how uniquely human they were perceived to be. The emotions used in Study 3.3 were *admiration*, *agitation*, *amusement*, *anger*, *astonishment*, *attachment*, *attraction*, *awe*, *bitterness*, *compassion*, *confusion*, *contentment*, *depression*, *desire*, *disappointment*, *disenchantment*, *disgust*, *disillusion*, *dread*, *empathy*, *enjoyment*, *envy*, *euphoria*, *excitement*, *fascination*, *fear*, *fury*, *gratitude*, *grief*, *guilt*, *happiness*, *hate*, *hope*, *horror*, *humility*, *irritation*, *jealousy*, *joy*, *loneliness*, *love*, *lust*, *optimism*, *pain*, *panic*, *passion*, *pleasure*, *pride*, *relief*, *remorse*, *repulsion*, *resentment*, *resignation*, *sadness*, *self-satisfaction*, *serenity*, *shame*, *shyness*, *sorrow*, *spite*, *surprise*, *sympathy*, *tenderness*, *terror,* and *triumph*.

Manipulation Checks. The same manipulation checks as those in Study 3.1 were used in the present study. These were included to ensure that the perceived extent of harm experienced by residents of Lagos and the sense of collective responsibility felt by participants were manipulated as intended.

Procedure

Participants were informed that the study was designed to help understand how people ascribe emotions to different groups of individuals. After reading the information sheet, participants were asked to indicate their consent to participation. Brief demographic questions followed, including gender identity, age, political orientation, and checks for inclusion criteria.

All participants then completed the emotion humanness and emotion valence ratings for each of the 64 emotions. Half of the sample completed the humanness ratings first, and half completed the valence ratings first. In both the humanness and the valence rating tasks, participants responded to a total of 65 items. These were the 64 randomised emotion items

and one attention check item, which always appeared approximately halfway through. The 65 items were displayed in blocks of 17 items at a time, with the final block containing 14 items for both the humanness and valence rating tasks. Participants then completed the filler task before moving on to the experimental stimuli for the study, the vignettes. The procedure for the section of this study involving vignettes was identical to that of Study 3.1.

After the vignettes, all participants were asked to respond to the outgroup emotion typicality block. After completing the outgroup emotion typicality ratings, participants could provide optional feedback and were debriefed before finishing the study. Participation took an average of 19 minutes.

3.5.2. Results

Manipulation Checks

The Perceived Extent of Harm. A one-way between-subjects ANOVA revealed a significant difference in the perceived extent of harm experienced by residents of Lagos as a result of climate change, $F(2, 249) = 35.996$, $p < .001$, $\eta^2 = .22$. Post hoc comparisons indicated no significant difference between the harm responsible ($M = 89.5$, $SE = 1.81$) and harm not responsible conditions ($M = 91.9$, $SE = 1.81$), $p = 1$, 95% CI [-8.56, 3.80]. However, the perceived extent of harm was significantly lower in the no harm condition ($M = 72.0$, $SE = 1.81$) than in either the harm responsible condition, $p < .001$, 95% CI [-23.72, -11.36], or the harm not responsible condition, $p < .001$, 95% CI [-26.10, -13.74]. Participants in the two harm conditions tended to rate the perceived extent of harm about 19 points higher on the 100-point scale than those in the no harm condition. Therefore, the perceived extent of harm was manipulated as intended between conditions in Study 3.3, as illustrated in the right panel of Figure 6.

Sense of Collective Responsibility. A one-way between-subjects ANOVA revealed a significant difference between conditions in the extent to which participants felt a sense of collective responsibility for the harm caused by climate change, $F(2, 249) = 9.854$, $p = < .001$, $\eta^2 = .07$. Post hoc comparisons revealed that participants in the harm responsible condition

felt significantly more collective responsibility for the harm (M = 60.2, SE = 2.60) than did those in the harm not responsible condition (M = 48.7, SE = 2.60), p = .006, 95% CI [2.66, 20.37] and those in the no harm condition (M = 44.4, SE = 2.60), p < .001, 95% CI [6.91, 24.62]. There was no significant difference in collective responsibility between the harm not responsible and no harm conditions, p = .745, 95% CI [-4.61, 13.11]. Participants in the harm responsible condition tended to rate their sense of collective responsibility for the harm experienced by residents of Lagos about 12 points higher on the 100-point scale than those in the harm not responsible condition. Thus, participants' sense of collective responsibility was manipulated as intended between conditions in Study 3.3, as seen in the right panel of Figure 7.

Emotion Ratings

Before conducting the main analyses, we checked whether emotion humanness and valence correlated. There were no significant correlations between ratings of emotion humanness and emotion valence in the harm responsible condition, r = -.030, p = .814; the harm not responsible condition, r = -.062, p = .628; or the no harm condition, r = -.030, p = .817. Thus, we did not need to control for potential effects of valence when examining the relationships between outgroup typicality and emotion humanness, and it was not necessary to control for potential effects of humanness when examining the relationships between outgroup typicality and emotion valence.

Testing for Infrahumanization Following Harm. A series of three linear regressions were conducted in which the relationship between emotion humanness and outgroup emotion typicality ratings was examined. The first model measured the relationship in the harm responsible condition, for which *Bitterness* was identified as highly influential using Cook's distance and, thus, omitted (leaving a final N = 63). A significant negative relationship for the harm responsible condition was found, $F(1, 61) = 5.667$, b = -.128, p = .02, R^2 = .09. The second model measured the relationship in the harm not responsible condition, with *Bitterness*, *Disenchantment* and *Disillusion* identified as being highly influential and, therefore,

omitted (leaving a final $N = 61$). A significant negative relationship was also found for the harm not responsible condition, $F(1, 59) = 7.84$, $b = -.202$, $p = .007$, $R^2 = .12$.

The third model measured the relationship in the no harm condition, for which *Optimism* and *Terror* were excluded because they were identified as highly influential (leaving a final N = 62). A significant negative relationship between emotion humanness and outgroup emotion typicality ratings was also found in the no harm model, $F(1, 60) = 5.168$, $b = -.150$, $p = .027$, $R^2 = .08$. In all three of the conditions, the more human an emotion was, the less residents of Lagos were seen as typically experiencing it.

Our main question of interest was whether the regression coefficients differed significantly between conditions. Fisher's *Z* transformations were performed using the *cocor* independent groups function in *RStudio*. These tests revealed no significant difference in the slopes between the harm responsible model ($b = -.128$, $N = 63$) and harm not responsible model ($b = -.202$, $N = 61$), $z = 0.413$, $p = .68$, *Cohen's q* = .08. Additionally, no difference in the relationship was found between the no harm model ($b = -.150$, $N = 62$) and the harm responsible model ($z = 0.122$, $p = .90$, *Cohen's q* = .02), or between the no harm and harm not responsible models ($z = 0.291$, $p = .77$, *Cohen's q* = .05). The extent to which perceptions of emotion humanness predicted outgroup emotion typicality ratings did not significantly differ between the three harm conditions. We did not replicate prior work, which suggests that outgroup members are seen as experiencing uniquely human emotions to a lesser extent when a sense of collective responsibility for harming them is elicited, compared to when it is not. The three models with emotion humanness as the predictor and outgroup emotion typicality as the outcome are plotted in Figure 10.

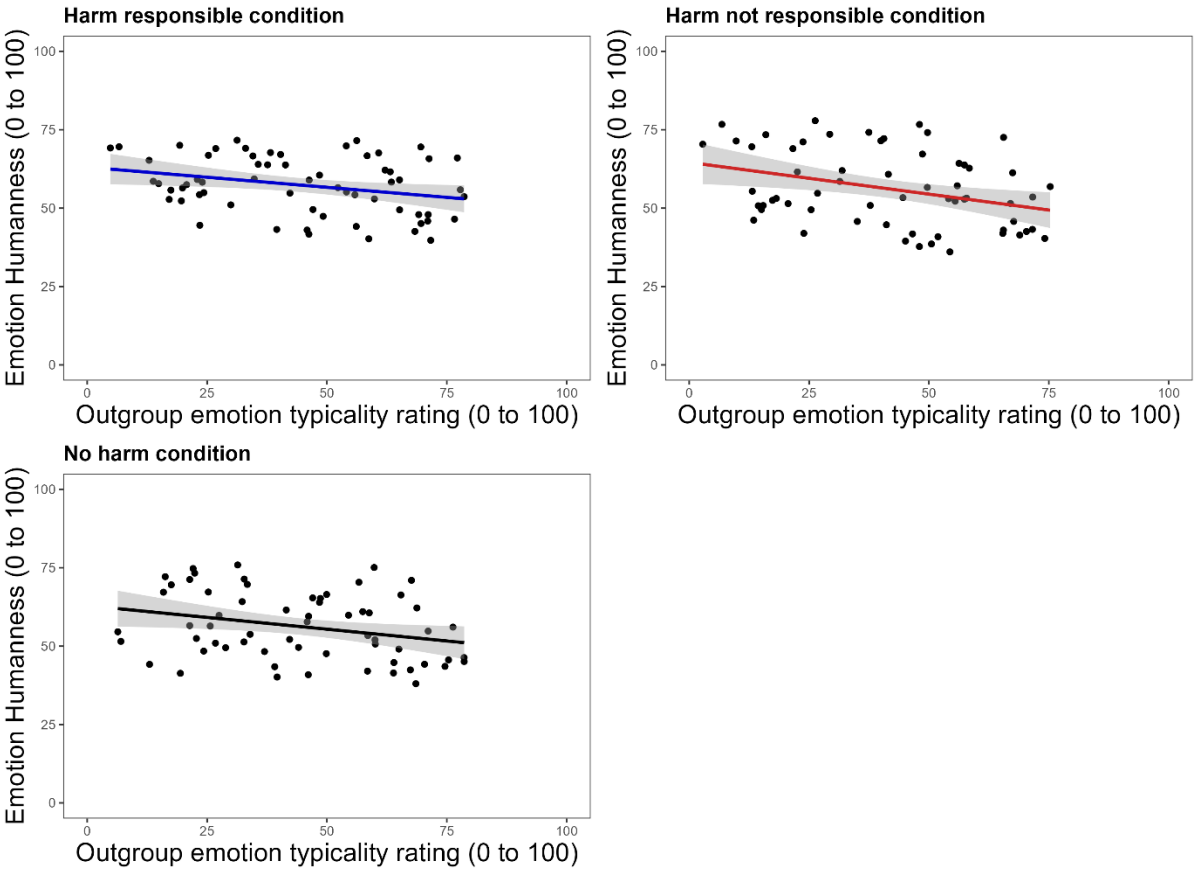

Figure 10. Results of Study 3.3: Testing for infrahumanization following harm

The three regression models with emotion humanness as the predictor and emotion typicality for outgroup members (residents of Lagos) as the outcome are displayed. The small negative relationship between these variables did not differ between the three conditions.

Testing for Negative Affect when Harmed. We ran another series of three linear regressions to test our alternative explanation whereby outgroup members are seen as experiencing negative emotions to a greater extent when they are perceived as being harmed than when they are not. In each model, emotion valence was entered as the predictor and outgroup emotion typicality ratings as the outcome. The first regression model measured the relationship in the harm responsible condition, for which the emotions *Guilt*, *Shame*, and *Love* were identified as highly influential to the model and omitted (leaving a final $N = 61$).

A significant negative relationship was found in the harm responsible model, $F(1, 59) = 52.51$, $b = -.198$, $p < .001$, $R^2 = .47$. The second model measured the relationship

in the harm not responsible condition, with *Guilt*, *Love,* and *Shame* again identified as a highly influential and omitted (leaving a final $N = 61$). A significant negative relationship was also found for the harm not responsible model, $F(1, 59) = 116.3$, $b = -.314$, $p = < .001$, $R^2 = .66$. In the third model, the no harm model, the emotions *Euphoria,* and *Grief* were identified as highly influential and omitted (leaving a final $N = 62$). Unlike the two harm models, a significant positive relationship between emotion valence and outgroup emotion typicality ratings was found in the no harm model, $F(1, 60) = 283.9$, $b = .329$, $p < .001$, $R^2 = .83$.

A Fisher's Z transformation was performed using the *cocor* independent groups function in *RStudio* to test whether the relationship between emotion valence and the perceived typicality of emotions in residents of Lagos significantly differed between the two harm conditions and the no harm condition. The difference between the negative relationship in the harm responsible model ($b = -.198$, $N = 61$) and the positive relationship in the no harm model ($b = .329$, $N = 61$) was significant, $z = 2.934$, $p = .003$, *Cohen's q* = .54. A significant difference was also found between the negative relationship in the harm not responsible model ($r = -.314$, $N = 61$) and the positive relationship in the no harm model, $z = 3.609$, $p < .001$, *Cohen's q* = .67. As predicted, participants rated outgroup members as typically experiencing negative emotions to a greater extent, and positive emotions to a lesser extent, when seen as experiencing harm compared to when they are not. No significant difference was found between the harm responsible and harm not responsible models, $z = .672$, $p = .502$, *Cohen's q* = .13. The three regression models with emotion valence as the predictor and outgroup emotion typicality as the outcome are plotted in Figure 11.

When testing for both infrahumanization following harm and negative affect when harmed, the same results as those reported were found when all 64 emotions were included in the regression models, when any emotion identified as highly influential in one model was removed from all three models ($N = 59$ in all models), and when performing Fisher's Z transformations on respective Pearson correlation coefficients rather than regression coefficients. Results for these alternative analyses are reported in Appendix B: Supplementary Materials for Chapter 3.

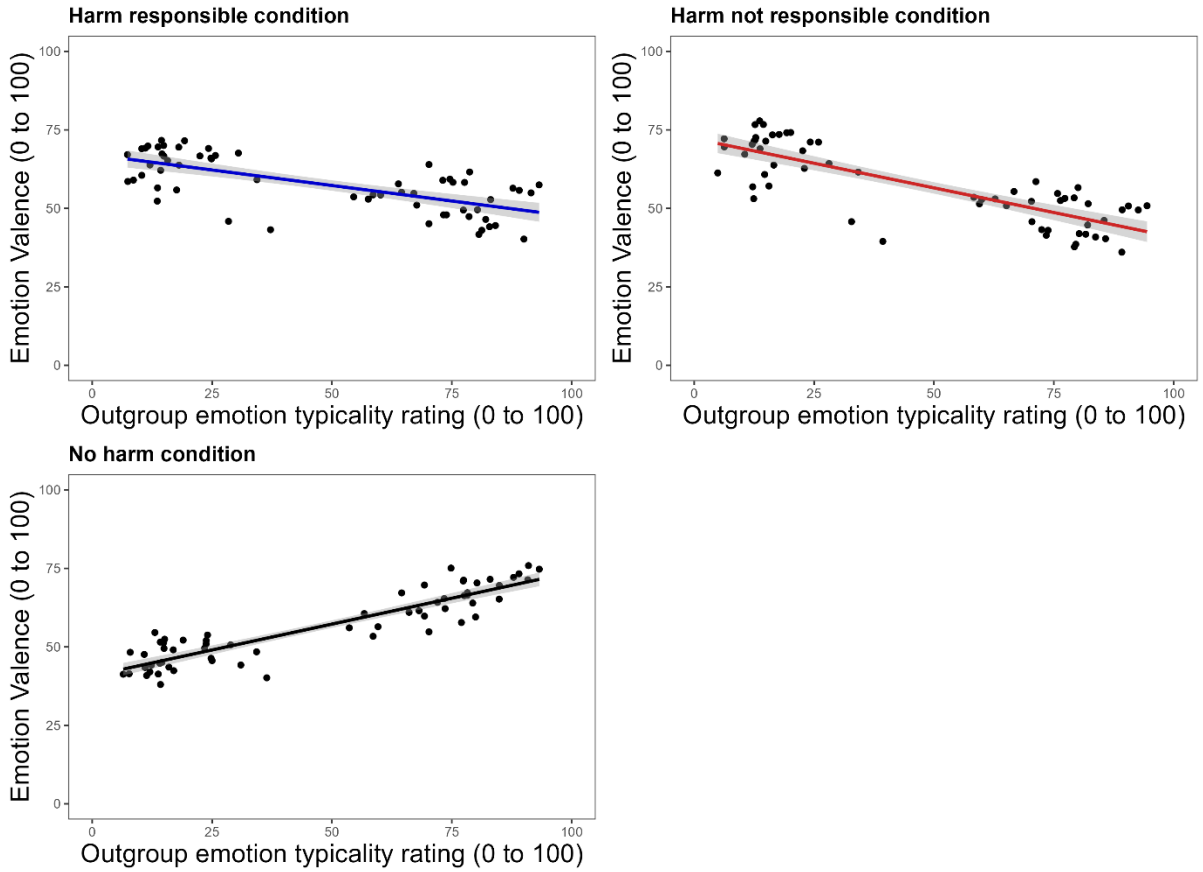

Figure 11. Results of Study 3.3: Testing for negative affect when harmed

The three regression models with emotion valence as the predictor and emotion typicality for outgroup members (residents of Lagos) as the outcome are displayed. As predicted, a negative relationship emerged in both the harm responsible and harm not responsible conditions, with each found to differ significantly from the positive relationship that emerged in the no harm condition.

3.5.3. Study 3.3 Discussion

Despite having a more sensitive design than in Studies 3.1 and 3.2, we again found no evidence that feeling collective responsibility for harm induces infrahumanization of the outgroup. Again, we found evidence for an alternative pattern of emotion attribution that aligned with the results of Studies 3.1 and 3.2. As predicted, we found that outgroup members

were seen as experiencing negative emotions to a greater extent and positive emotions to a lesser extent when harm against them was made salient.

3.6. Chapter 3 General Discussion

Previous research has suggested that subtle dehumanization can occur as an outcome of, or response to, intergroup harm. Castano and Giner-Sorolla (2006) and Čehajić et al. (2009) each reported that participants are more likely to deny members of a group uniquely human emotions when they believe their ingroup to be responsible for harming them. Across three well-powered and pre-registered studies, we sought to measure whether these influential findings replicate in two new intergroup contexts: the harm caused by climate change and the harm caused by fast fashion. In all three studies, we manipulated the extent to which UK residents felt responsible for the harm caused to another group. In Studies 3.1 and 3.2, we presented participants with 16 emotions that varied in how human they are perceived to be and how positive they are to experience. We found no evidence for the hypothesis that feeling responsible for harm reduces the extent to which participants attribute uniquely human emotions to outgroup members. Instead, we found evidence for a much simpler pattern of emotion attribution. When the harm experienced by a group was made salient, participants believed members of that group typically experience more negative and less positive emotions, regardless of emotion humanness and their sense of collective responsibility.

One plausible reason for failing to replicate previous findings is that we used an insufficiently sensitive design. In Studies 3.1 and 3.2, we broadly followed the methods of Čehajić et al. (2009). Like almost all studies on infrahumanization, Čehajić and colleagues presented participants with a relatively small number of emotions (16) that varied categorically in perceived humanness and valence. While this approach is prevalent in the empirical literature (Azevedo et al., 2021; Cuddy et al., 2007; Delgado et al., 2009; Delgado Rodríguez et al., 2012; Enock et al., 2021b; Kteily et al., 2015; Rodríguez-Pérez et al., 2011; Vaes et al., 2002, 2003; Wohl et al., 2012), it has recently been criticised as potentially obscuring evidence

for subtle dehumanization such as infrahumanization effects (Enock & Over, 2022; Vaes, 2023).

In Study 3.3, we utilised a design incorporating many emotions in a manner more similar to that used by Castano and Giner Sorolla (2006). In this study, we asked participants to rate 64 emotions along the dimensions of humanness, valence, and typicality of the outgroup. We then compared the strength of the relationship between these variables between conditions. The same participants who responded to the dependent variable provided the humanness and valence scores, with mean scores reflecting their subjective categorisation of each emotion. This design feature helps to avoid inconsistency in how human or valenced an emotion is considered should pre-test data from a different sample group be used. Such inconsistency is likely, considering the subjectivity and inconsistency in how emotions are categorised even between published pieces of infrahumanization research. For example, *disgust* is treated as uniquely human by DeLuca-McLean and Castano (2009) but as a shared emotion by Prati and Giner-Sorolla (2018). As Castano and Giner-Sorolla (2006) noted, using continuous variables also allows for more emotions to be included in the measures (p.806), providing more reliable mean scores for each emotion variable. Thus, this approach allowed us to test our hypotheses using more representative variable data compared to a factorial design with pre-collected variable data, which is standard in infrahumanization research.

However, the design of Study 3.3 treats the data as correlational within each condition. Thus, continuous variables do not necessarily improve the ability to infer causal associations between emotion humanness or valence and emotion attribution. To avoid an overly complex design, we did not include a between-subjects manipulation regarding the influence different levels of emotion humanness, or valence might have. Future research could incorporate such a between-subject manipulation into a similar design to help infer possible causal effects in the perception of harmed outgroup members' emotional experiences. Nevertheless, we found no evidence that participants were more likely to infrahumanize the outgroup in the harm responsible condition, allowing us to conclude that the causal effect of collective responsibility for harm on the perceived emotional experiences of harmed outgroup members is limited,

challenging the claims of Castano and Giner-Sorolla (2006) and Čehajić et al. (2009). Instead, converging with the results of Studies 3.1 and 3.2, we found that participants perceive outgroup members as experiencing negative emotions more strongly when harm they have experienced is made salient.

It is interesting to consider why we did not observe the same emotion attribution trends reported by Castano and Giner-Sorolla (2006) and Čehajić et al. (2009). One possibility is that infrahumanization following harm may occur in some intergroup contexts but not others. Castano and Giner Sorolla (2006) tested their hypothesis with the harm committed by humans against hypothetical aliens and the real-world harm inflicted on Indigenous Americans and Aboriginal Australians by European colonisers. Čehajić et al. (2009) tested their hypothesis in the contexts of the Serbian genocide against Bosnian Muslims in the 1990s and the colonial harm inflicted by Europeans on Indigenous Chileans in colonial times. On the other hand, we tested our hypotheses with the more contemporary contexts of harm caused to people by climate change and exploitation in fast fashion supply chains. These are contexts often mentioned in the media and may feel more current, tangible, and relatable to people. Indeed, these are cases of ongoing suffering and exploitation that many individuals actively and frequently contribute to today.

Despite our manipulation checks having broadly demonstrated that the perceived harm and sense of collective responsibility for harm were manipulated between participant conditions as intended, further contextual differences might explain why the previously reported effects failed to replicate. It would be interesting for future research to investigate these possible differences (e.g., the impact of past harm compared to ongoing harm on intergroup perceptions) and in which contexts infrahumanization following harm is more likely to occur. As we only tested for infrahumanization, examining other forms of dehumanization in post-harm intergroup contexts would be an interesting avenue for future research. For example, it remains possible that while infrahumanization might not occur, explicit or linguistic dehumanization or a denial of human rights and equal status, as described by Luft (2015, 2019) and Smith (2011, 2020, 2021b), could emerge in post-harm contexts.

It is worth pointing out that while the results of Castano and Giner-Sorolla (2006) and Čehajić et al. (2009) are regularly cited as evidence for the claim that infrahumanization follows harm, some of the results they report are not especially strong. For example, the crucial interaction between the humanness of the emotion terms and conditions reported by Castano and Giner-Sorolla (2006) is statistically significant in their Study 2 but only marginally significant in Study 1, reported as "$p < .06$" (p. 808). In Čehajić et al. (2009), the results are considerably weaker. The crucial interaction between group membership and emotion attribution is marginally significant in their Study 1 ($p = .08$) and nonsignificant in Study 2 ($p = .83$). Against what is widely considered statistically appropriate, the authors break down these nonsignificant interactions into simple comparisons and report the pattern of results that they predicted. Thus, there may be many intergroup contexts in which these small effects are difficult, or even impossible, to detect.

A trend observed In Study 2 of Castano and Giner-Sorolla (2006) that we also found in our Study 3.2 is that in all conditions, the more uniquely human an emotion is considered to be, the less outgroup members were seen as typically experiencing it. The reason for this trend remains unclear, and any suggestion as to why it occurred beyond the specific constructs examined remains speculative. One possibility is that this result suggests the overall infrahumanization of the outgroup. However, as neither of these studies included a measure of ingroup emotion attribution, we cannot rule out the possibility that the pattern of emotion attribution would be similar for ingroups and outgroups. Similarly, trends associated with the impact of emotion valence on emotion perception might also have benefitted from ingroup measures alongside the outgroup measures. For instance, we found in Study 3.3 that the more negative an emotion was, the less typical of outgroup members participants in the no harm condition considered it to be. The vignette in this condition focused on socio-economic aspects of life for residents of Lagos without mentioning harm, instead detailing the strong local industries and recreational activities. Focusing on such details might have portrayed a somewhat unbalanced or idealistic image of their experience, leading participants to see residents of Lagos as generally experiencing more positive and less negative emotions.

Perhaps a vignette that encourages no relationship between emotion valence and outgroup emotion attribution would have been a better comparison point, which future research should consider in such designs. While infrahumanization theory (Leyens et al., 2000, 2001) does not make predictions specifically concerning the perceived valence of outgroup members' emotions, comparisons to ingroup emotion attribution measures might have also shed light on such trends. Perhaps the interaction between emotion humanness and valence might have revealed different results for ingroup members, with the comparison to outgroup perceptions providing possible evidence for infrahumanization or more evidence for our alternative hypothesis. It would be valuable to explore these questions further in future research.

In Studies 3.1 and 3.2, outgroup members were seen as experiencing negative uniquely human emotions to a lesser extent than negative shared emotions. This pattern could be due to several factors. It is possible that within the emotion selection of Čehajić and colleagues (2009) used in both studies, the negative shared emotions and negative uniquely human emotions might not be equivalent in terms of how negative they are perceived. Research suggests people tend to minimise or underestimate outgroup members' pain and negative emotional experiences (Riva & Andrighetto, 2012) due to limited motivation to empathise with outgroup members relative to ingroup members (Cikara et al., 2017). Batson & Ahmad (2009) also suggest that people are less likely to detect and attend to another's suffering when the victim is distant in space, time, or kinship or belongs to a different racial, political, or social group (as cited by Cikara et al., 2017). This research highlights the need for equivalence between the emotion categories in terms of their valence, as the observed trend could be due to the negative uniquely human emotions being more negative than the shared negative emotions.

The negative shared emotions might also have seemed more applicable to the context when judging the emotional experiences of harmed outgroup members than the negative uniquely human emotions. Gray & Wegner (2009) argue for the importance of moral agency in emotion perception, whereby moral agents cause a moral event, and moral patients experience its effects. This consideration is highly relevant to the intergroup contexts we

examined, where the target outgroups are disproportionate victims of harm and not the perpetrators. Thus, it could be argued that experiencing the negative uniquely human emotions in this selection, particularly *shame*, *guilt*, and *remorse*, might imply moral agency in the situation and, thus, are not as applicable to victims of climate change or fast fashion exploitation as the less agentic shared negative emotions like *sadness*, *anger*, or *fear* (see Russell & Fehr, 1994). However, we can only speculate why such a trend emerged, as our designs did not include measures of ingroup emotion attribution, empathy, or the perceived moral agency of the emotions included in our studies. Further research could examine the role of these variables in such designs, which could help to explain the trend of harmed outgroup members being seen as experiencing negative emotions to a lesser extent when they are uniquely human than when they are shared with other species.

Further research is also needed to address the replicability issues we encountered when examining apparent infrahumanization following harm and to help clarify the explanatory value of social-psychological models of dehumanization in post-harm intergroup contexts. Importantly, we are not questioning whether dehumanization can sometimes occur more broadly as a response to harm. Many theorists, including Smith (2011, 2021) and Luft (2015, 2019), have suggested that dehumanization may occur as a means by which to justify harm. Broadly in line with this view, perpetrators of harm sometimes refer to their victims as less than human in qualitative interviews (Hatzfeld, 2003). Instead, we suggest that the extent to which dehumanization following harm reveals itself as a denial of uniquely human emotions may be less widespread than often assumed. Given the real-world importance of understanding the motivations and psychological processes involved in intergroup hostility and harm, social psychology must rely on valid and replicable findings going forward.

Chapter 4: Infrahumanization or Intergroup Preference?

4.1. Abstract

According to infrahumanization theory, people perceive members of outgroups to experience uniquely human emotions (e.g., nostalgia, empathy) to a lesser extent than members of their ingroup. Recently, however, this view has been critiqued. Enock et al. (2021b) presented empirical work to show that intergroup preference might better explain what appeared to be evidence for infrahumanization. Participants attribute prosocial emotions more strongly to their ingroup but antisocial emotions more strongly to their outgroup. Importantly, they do so regardless of whether these emotions are perceived to be uniquely human. These results, however, have proved controversial. One source of ambiguity is that most studies testing infrahumanization use a small number of emotion terms. It is thus unclear whether, where observed, effects apply to emotions in general or just to the select number of emotions chosen as stimuli. I pitted infrahumanization theory against the intergroup preference account in an improved design that incorporated 68 emotions in total (17 per condition). In both intergroup contexts investigated, I found that apparent evidence for infrahumanization is better explained by intergroup bias. Implications for broader psychological research on dehumanization and intergroup prejudice are discussed.

4.2. Introduction

In this chapter, I present work concerning infrahumanization theory (Leyens et al., 2000, 2001). According to this influential theory, individuals are thought to attribute uniquely human emotions more strongly to their ingroup than to outgroups. Uniquely human emotions are complex and long-lasting emotions that we assume other animals do not experience, such as *nostalgia*, *optimism*, *sympathy*, and *empathy*. On the other hand, emotions that we assume are shared by humans and other animals are fleeting and grounded in contextual instinct, such as *joy*, *fear*, *sadness*, and *anger*. Apparent evidence for infrahumanization has been found for a wide range of target outgroups, including women (Gaunt, 2013; Viki & Abrams, 2003), refugees (Azevedo et al., 2021; Bruneau et al., 2018), other nationalities (Leyens et al., 2003; Paladino et al., 2004; Viki & Calitri, 2008), people with disabilities (Capozza et al., 2016; Rodríguez et al., 2016), in minimal group contexts (Demoulin et al., 2009; Simon & Gutsell, 2020), and contexts of intergroup conflict (Andrighetto et al., 2012; Tam et al., 2008).

In recent years, several criticisms have been made regarding the theoretical models and methodologies used in dehumanization research (Bloom, 2017, 2022; Lang, 2010, 2020; Manne, 2016; Over, 2021a, 2021b; Rai et al., 2017; Smith, 2016, 2023). One source of criticism is the methodological shortcomings of the stimuli used to measure a dehumanizing perception of outgroup members. Over (2021a, 2021b) notes that when measuring subtle forms of dehumanization, such as infrahumanization, previous research has relied on measuring prosocial or desirable aspects of humanity. For example, uniquely human emotions frequently incorporated into previous research include *hope*, *optimism*, *admiration*, *remorse*, *compassion*, and *love*. Moreover, there is a noticeable omission of antisocial or undesirable aspects of humanity in the empirical literature on infrahumanization. As a result of this confound, we cannot disentangle evidence of subtle dehumanization from evidence of an intergroup bias, whereby one might attribute more desirable aspects of humanity to their ingroup and less desirable aspects of humanity to the outgroup.

Enock and colleagues (2021b) pitted infrahumanization theory against an alternative intergroup preference account by incorporating stimuli that were perceived to be uniquely

human but antisocial (e.g., contempt, scorn) into the stimulus set alongside uniquely human emotions that were prosocial (e.g., humility, hope). Across seven intergroup contexts, the researchers' findings demonstrated that initial evidence of infrahumanization was better explained by intergroup preference once emotional sociality was considered. Their participants saw members of their ingroup as experiencing prosocial uniquely human emotions more than outgroup members. Crucially, however, they saw outgroup members as experiencing antisocial uniquely human emotions more strongly than ingroup members. Enock and Over (2022) provided further evidence of the confounding influence of emotion sociality on apparent infrahumanization effects. They showed that a lesser attribution of prosocial emotions (and a greater attribution of antisocial emotions), rather than a lesser attribution of uniquely human emotions, predicted reduced prosocial behaviour towards the target.

These results, however, have proved controversial. A methodological limitation of much previous research on subtle forms of dehumanization is that a small number of emotion terms tend to be included in the stimulus set (Enock & Over, 2022; Vaes, 2023). Vaes (2023) notes that the presence or absence of one particular emotion in a category with only two or three others might be an unreliable method of differentiating between uniquely human or shared aspects of humanity, and likewise, whether they are prosocial or antisocial. It is worth noting here that while only three or four items were used in each emotion category in previous work (Enock, Tipper, et al., 2021; Enock & Over, 2022, 2023a), all of the emotions included showed a broadly consistent pattern of attribution within categories across many intergroup contexts, suggesting the effects are not overly limited in generalisability. However, the emotion items were chosen as the most exemplary of each category (e.g., prosocial uniquely human emotions, antisocial emotions shared with other animals). Each emotion category contained emotions closer to the extreme ends regarding how uniquely human or shared with animals and how prosocial or antisocial they are. Therefore, the effects found in this work that challenge previous infrahumanization findings warrant further investigation across a wider spread of emotion stimuli.

In order to address this critique, we sought to pit the predictions of infrahumanization theory (Leyens et al., 2000, 2001) against an intergroup preference account (Enock et al., 2021b; Over, 2021a) in an improved design incorporating considerably more emotions into the stimulus set. Whereas previous research typically used between 6 and 18 emotions (two to four emotions per condition, e.g., Demoulin et al., 2004; Prati & Giner-Sorolla, 2018; Vaes et al., 2003), we chose 68 emotions (17 per condition). These emotions vary in terms of whether they are considered uniquely human or shared with other species and whether or not they are generally considered prosocial or antisocial. Infrahumanization theory predicts that, in this improved design, participants will attribute uniquely human emotions more strongly to their ingroup than their outgroup, regardless of whether those emotions are prosocial or antisocial. According to this view, there should be no difference in the attribution of emotions shared with other species to the ingroup and the outgroup. The intergroup preference account predicts that participants will attribute prosocial emotions more strongly to their ingroup and antisocial emotions more strongly to their outgroup regardless of whether or not they are uniquely human. In principle, both effects could occur in parallel; we may observe evidence of infrahumanization and intergroup preference. We tested these different hypotheses in two intergroup contexts.

The intergroup context we examined in Study 4.1 was that of UK nationals and immigrants to the UK. It is widely claimed that immigrants are a social group that is often the target of dehumanizing language and treatment, which is seen as a global social issue (Amnesty International, 2021; United Nations, 2018, 2020). At the individual level, social-psychological literature has suggested that immigrants may be seen as less human by the majority group members of the host society (Esses et al., 2008; Hodson & Costello, 2007; Kteily & Bruneau, 2017; Markowitz & Slovic, 2020, 2021; Utych, 2018). Research has also suggested that immigrants are vulnerable to infrahumanization (Banton et al., 2020; Prati et al., 2016). Dehumanizing narratives when discussing immigrants have grown to permeate mainstream media and political discussions in the UK, particularly about asylum seekers (Lazović, 2021; Webber, 2023).

In Study 4.2, we examined British people's perceptions of the emotional experiences of Irish Travellers, an ethnic minority group native to Ireland with an estimated population of 300,000 in the UK. Also known as Gypsy Travellers, Pavees, or Mincéirs, members of this group face a distinct form of racism known as *antigypsyism* that is also experienced by groups across Europe, such as the Romani, Sinti, Sami, and Scottish Travellers. Antigypsyism maintains a homogenising and vilifying perception of these groups in society, in turn leading to systemic discrimination, stigma, and violent practices towards them (Alliance Against Antigypsyism, 2017; Ljujic et al., 2012). Antigypsyism has been referred to as "a form of dehumanization and institutional racism nurtured by historical discrimination" by the European Commission Against Racism and Intolerance (2020, p. 3). Media portrayals of Irish Travellers in the UK have also been colloquially referred to as dehumanizing (Cojocaru, 2022; Murray, 2020). Empirical work has demonstrated how other groups collectively labelled as “Gypsies” tend to be blatantly dehumanized (Kteily et al., 2015; Tileagă, 2007) as well as subtly dehumanized (Dalsklev & Kunst, 2015; Kteily et al., 2015; Marcu & Chryssochoou, 2005; Martínez et al., 2012; Miranda et al., 2014) by majority group members in UK and other European societies. Thus, examining whether infrahumanization accurately captures the forms of bias members of the Irish Traveller community face is an urgent matter for social-psychological research.

We restricted our sample to people with right-wing political views in both studies. UK nationals who vote on the right of the political spectrum tend to have less favourable views of immigration in general (Banton et al., 2020; Richards et al., 2023). Research has also demonstrated that right-leaning individuals tend to score higher in both right-wing authoritarianism (Altemeyer, 1981) and social dominance orientation (Sidanius & Pratto, 1999), both of which have been linked to higher rates of antigypsyism and general prejudice towards marginalised groups (Duckitt & Sibley, 2007; Todosijević & Enyedi, 2002; Zick et al., 2011). Other research suggests that right-leaning individuals might be more likely to infrahumanize members of a threatening outgroup (Sánchez & García, 2016). Thus, including right-leaning political orientation as an inclusion criterion for participants should provide the

best opportunity to provide evidence for infrahumanization if it occurs (see also Crawford, 2012).

4.2.1. Note on Methods

All studies received ethical approval from the Psychology Departmental Ethics Committee at the University of York (Ethics ID no. 2264). All data was collected online. We created and administered the studies using Qualtrics (https://www.qualtrics.com) and recruited participants using Prolific (https://www.prolific.co). Independent samples were recruited for each study. Informed consent was obtained at the start of each session according to approved ethical guidelines. All studies were pre-registered on AsPredicted.com before commencing data collection. For each study, removing outlier cases (whose scores were >3 SD away from the mean) yielded the same results as when all data were included in analyses. Thus, the latter is reported for both Study 4.1 and Study 4.2. Participants were rewarded at an approximate rate of £9 per hour for all studies. Assumption testing and subsequent analyses were conducted using *SPSS* and *RStudio*. All post hoc tests were Bonferroni-corrected for multiple comparisons.

4.3. Study 4.1: Testing for Infrahumanization or Derogation of Immigrants to the UK

4.3.1. Method

Participants

A power analysis using MorePower 6.0.4 found a minimum sample size of 126 participants would allow us to detect an interaction with a medium effect size (partial eta squared .06) with an alpha of .05 and power of .80. The lower power of .85, compared to the .95 power in Studies 2.1A to 3.2, was chosen as a cost-effective way to collect enough data that would allow us to detect a medium effect. To allow for counterbalancing, we recruited 130 participants for this study. All participants were adult UK nationals born in the UK, currently residing in the UK, fluent in English, with a right-leaning political orientation and a minimal

approval rating of 95% on Prolific. The sample comprised 75 people who identified as male and 55 as female, with ages ranging from 19 to 77 (*M* = 48.1, *SD* = 15.1).

Materials

Emotion Attribution. A selection of 68 different emotion terms was made for the measures used in both studies. We made this selection based on pre-test data collected by Bunce and colleagues (2024), who compiled a list of 250 emotion terms used in previous research measuring infrahumanization. These researchers asked 200 participants to indicate the extent to which they think humans experience each emotion compared to other animals. The sliding scale participants used to rate each emotion's humanness ranged from 0 (*Shared with other species)* to 100 (*Uniquely human*). Bunce and colleagues also asked a separate sample of 200 participants to rate the same 250 emotions regarding how kind or unkind a person who typically experiences each emotion is likely to be. The sliding scale participants used to rate the sociality of each emotion ranged from 0 (*Extremely unkind*) to 100 (*Extremely kind*). Thus, an emotion was considered prosocial if rated on the higher end of this scale and antisocial if rated on the lower end.

In deciding which emotions to include in our measures out of the 250 taken from Bunce and colleagues' pre-test, we first omitted 16 emotions that less than 80% of the participants in the pre-test knew the definition of (such as *quandary* and *rancour*). These lesser-known emotion terms would likely have the same degree of unfamiliarity in our participant sample, which we wanted to avoid. We also omitted synonyms within the selection (e.g., *hope* was retained while *hopeful* was removed) and emotional terms we considered unclear or vague in their meaning (e.g., *sentiment*, *quiet*) or those that may have multiple interpretations (e.g., *humility*, as it could be interpreted as feeling humiliated or feeling humbled). Minor re-phrasings were then made to ensure that emotion terms were in a noun format (e.g., *vengeful* was changed to *vengefulness*). These steps were taken to minimise the risk of emotion terms carrying grammatical inconsistency, repetition, or ambiguity within our selection.

We categorised the remaining emotion terms based on their respective mean scores along the dimensions of humanness and sociality. The resulting four emotion categories were

uniquely human prosocial emotions (e.g., *modesty*, *empathy*), uniquely human antisocial emotions (e.g., *bitterness*, *pessimism*), prosocial emotions shared with other animals (e.g., *attachment*, *comfort*), and antisocial emotions shared with other animals (e.g., *aggression*, *possessiveness*). Emotions with mean scores near the midpoint (50) for the humanness and sociality scales (e.g., *lust*, *worry*) were omitted from the selection at this point. Omitting these emotions that were considered neutral in terms of their humanness and sociality was necessary to allow our measures to reflect the factorial design widely used in infrahumanization research, whereby each emotion lay within one of the four emotion categories.

Omitting neutral emotions was also necessary to ensure that our final selection of emotions was balanced along the dimensions of humanness and sociality. Specifically, this meant that the categories of uniquely human prosocial emotions and uniquely human antisocial emotions did not differ in terms of humanness, nor did prosocial emotions shared with other animals and antisocial emotions shared with other animals. However, each uniquely human category was rated as significantly higher in humanness than the two shared categories. Similarly, the categories of uniquely human prosocial emotions and prosocial emotions shared with other animals did not differ in their average sociality ratings, nor did uniquely human antisocial emotions and antisocial emotions shared with other animals. However, each prosocial category was rated significantly higher in sociality than the two antisocial categories. We conducted two repeated measures ANOVAs on our initial selection of emotions, analysing trends in emotion humanness and emotion sociality mean scores between the four emotion categories. We then substituted emotions until the desired relationships between the four categories were confirmed with as large a selection of emotions as possible. Our final selection of 68 emotions (17 per emotion category) was balanced in terms of emotion humanness and emotion sociality as intended, with the full selection listed in Table 1.

Table 1. The selection of 68 emotions used in Studies 4.1 and 4.2

Uniquely Human Prosocial	**Uniquely Human Antisocial**	**Shared Prosocial**	**Shared Antisocial**
Admiration	Animosity	Adoration	Abandonment
Elation	Bitterness	Affection	Aggression
Empathy	Contempt	Attachment	Agitation
Enchantment	Deceit	Attraction	Anger
Gratitude	Disdain	Comfort	Annoyance
Hope	Disenchantment	Contentment	Frustration
Humility	Disillusionment	Delight	Fury
Inspiration	Envy	Desire	Greed
Joviality	Hatred	Enjoyment	Hostility
Modesty	Horror	Excitement	Impatience
Nostalgia	Pessimism	Fascination	Irritation
Optimism	Resentment	Happiness	Possessiveness
Passion	Scorn	Joy	Rage
Remorse	Shamelessness	Love	Ruthlessness
Repentance	Spite	Peacefulness	Tension
Sympathy	Vengefulness	Pleasure	Terror
Wonder	Wrath	Trust	Unhappiness

Notes: Shared Prosocial = Prosocial emotions shared with other animals; *Shared Antisocial* = Antisocial emotions shared with other animals.

Ingroup Emotion Ratings. Each participant was asked to indicate how strongly they thought a typical UK national feels each of the 68 emotions. For example, *"How strongly do you think a typical UK national feels excitement?"*. Participants responded using an unmarked sliding scale for each emotion item that ranged from 0 (*Not at all*) to 100 (*Very Strongly*), with the sliders initially fixed at the midpoint ('*Somewhat'*).

Outgroup Emotion Ratings. Participants were asked to indicate the extent to which they thought a typical immigrant to the UK felt each of the 68 emotions. For example, "*How strongly do you think a typical immigrant to the UK feels contempt?*". Participants responded to each item using the same unmarked sliding scale as with ingroup emotion ratings.

Attention Check. An attention check appeared roughly halfway through the items in both the ingroup emotion rating and outgroup emotion rating scales. Each attention check instructed participants to move the slider all the way to one end of the scale (e.g., *This is an attention check, please indicate 'Not at all'*). Responses within 20 points of the instructed end of the scale were accepted to accommodate the visual layout of the labels on the sliding scale.

Explicit Dehumanization. Participants were asked to respond to the 'Ascent of Human' scale (Kteily et al., 2015). This involved rating both ingroup members and outgroup members separately in terms of how evolved they are. Participants were presented with an image of five silhouettes, each depicting a stage of human evolutionary development, progressing from a primitive ape-like humanoid at one side of the image to a fully developed modern human at the opposite side. Participants were given the following instruction for the explicit dehumanization measure of ingroup members: *"People can vary in how human-like they seem. Some people seem highly evolved, whereas others seem no different than lower animals. Using the image below, indicate with the slider how evolved you consider UK nationals* (*who were born and currently reside in the UK) to be*". The instruction for the measure of explicit dehumanization of outgroup members was the same, apart from the final sentence, which read "*Using the image below, indicate with the slider how evolved you consider immigrants to the UK to be*". In line with Kteily and colleagues (2015), members of an outgroup (immigrants) are considered to be explicitly dehumanized if they are rated significantly lower on the ‘Ascent of Human’ scale (i.e., less evolved) than members of the participant ingroup (UK nationals).

Design

Study 4.1 had a 2 (group: ingroup and outgroup) X 2 (emotion humanness: uniquely human and shared with other animals) X 2 (emotion sociality: prosocial and antisocial) within-subjects design. Data were analysed using a repeated measures ANOVA. Counterbalancing ensured that half of the participants responded to ingroup emotion ratings then outgroup emotion ratings, and half responded to outgroup emotion ratings followed by ingroup emotion ratings. The order in which the ingroup and outgroup explicit dehumanization scales were

presented corresponded with the counterbalanced order of the ingroup and outgroup emotion ratings. Therefore, half of the participants rated ingroup members followed by outgroup members on the 'Ascent of Human' scale, and the other half responded in the opposite order.

Procedure

The study was published on Prolific with the title "Social aspects of emotion perception" and a link to the Qualtrics survey. The study description informed interested individuals that it was designed to help us understand how people ascribe emotions to different groups of individuals and gave a brief outline of the procedure, payment, and inclusion criteria to take part. Once participants opened the survey, a detailed information page was provided before they were asked to provide their informed consent to participate in the study. Participants were asked to respond to a few demographic questions and checks for inclusion criteria before progressing to the ingroup and outgroup emotion ratings. After these emotion attribution scales, participants were asked to respond to the measures of explicit dehumanization of ingroup and outgroup members. At the end of the study, participants were debriefed and allowed to provide feedback. They were then redirected to Prolific to submit their data and receive financial payment for their participation. Participation took an average of 11.25 minutes to complete.

4.3.2. Results

Emotion Attribution

We conducted a 2 (group: ingroup and outgroup) X 2 (emotion humanness: uniquely human and shared with other animals) X 2 (emotion sociality: prosocial and antisocial) repeated measures ANOVA to examine the competing hypotheses from infrahumanization theory (Leyens et al., 2000, 2001) and the intergroup preference account (Enock et al., 2021a; Over, 2021a).

There was a significant main effect of group $F(1, 129) = 5.655$, $p = .019$, $\eta_p^2 = .04$, whereby a typical UK national was seen as feeling emotions more strongly overall ($M = 53.8$, $SE = 0.69$) than a typical immigrant to the UK ($M = 52.4$, $SE = 0.63$). A significant main effect

of emotion humanness was also observed $F(1, 129) = 255.180$, $p < .001$, $\eta_p^2 = .66$. Emotions shared with other animals ($M = 55.7$, $SE = 0.60$) were seen as being felt more strongly than uniquely human emotions ($M = 50.4$, $SE = 0.61$) overall. No significant main effect of emotion sociality was found, $F(1, 129) = 1.528$, $p = .219$, $\eta_p^2 = .01$.

A significant two-way interaction was found between group and emotion humanness, $F(1, 129) = 23.43$, $p < .001$, $\eta_p^2 = .15$. However, simple comparisons did not follow the pattern predicted by infrahumanization theory. There was no significant difference in participant ratings of how strongly participants thought a typical UK national felt uniquely human emotions ($M = 50.4$, $SE = 0.74$) compared to a typical immigrant ($M = 50.4$, $SE = 0.67$), $p = .965$, 95% CI [-1.36, 1.42], $\eta_p^2 < .001$. Participants tended to rate a typical UK national as feeling emotions shared with other animals more strongly ($M = 57.1$, $SE = 0.70$) than a typical immigrant ($M = 54.3$, $SE = 0.66$), $p < .001$, 95% CI [1.58, 4.09], $\eta_p^2 = .14$.

As predicted by our intergroup preference hypothesis, there was a significant two-way interaction between group and emotion sociality, $F(1, 129) = 20.926$, $p < .001$, $\eta_p^2 = .14$. Participants tended to rate a typical member of their ingroup, UK nationals, as feeling prosocial emotions more strongly ($M = 57.7$, $SE = 1.21$) than a typical member of the outgroup, immigrants to the UK ($M = 50.6$, $SE = 1.17$), $p < .001$, 95% CI [4.48, 9.60], $\eta_p^2 = .19$. Antisocial emotions were rated as being felt by a typical immigrant ($M = 54.1$, $SE = 1.27$) more strongly than a typical UK national ($M = 49.9$, $SE = 1.34$), $p = .004$, 95% CI [-7.01, -1.34], $\eta_p^2 = .06$.

Importantly, each of the two-way interactions was qualified by a significant three-way interaction between group, emotion humanness, and emotion sociality, $F(1, 129) = 4.549$, $p = .035$, $\eta_p^2 = .04$. Pairwise comparisons followed the pattern predicted by the intergroup preference account. Participants tended to rate a typical UK national as feeling prosocial uniquely human emotions more strongly ($M = 53.6$, $SE = 1.16$) than a typical immigrant to the UK ($M = 48.7$, $SE = 1.23$), $p < .001$, 95% CI [2.43, 7.28], $\eta_p^2 = .11$. Crucially, they tended to rate a typical immigrant to the UK as feeling antisocial uniquely human emotions more strongly ($M = 52.1$, $SE = 1.34$) than a typical UK national ($M = 47.3$, $SE = 1.35$), $p < .001$, 95% CI [-7.71, -1.88], $\eta_p^2 = .08$. Participants tended to rate a typical ingroup member as feeling

uniquely human emotions more strongly than a typical outgroup member by roughly 5 points on the 100-point scale when the emotions were prosocial, and about 5 points less strongly when they were antisocial. As in the two-way interaction between group and emotion humanness, this result shows an intergroup preference towards seeing members of one's ingroup as more prosocial and outgroup members as more antisocial rather than infrahumanization.

Pairwise comparisons for the ratings of emotions shared with other animals revealed similar trends to those for uniquely human emotions. Participants tended to rate a typical UK national as feeling prosocial emotions shared with other animals more strongly (M = 61.8, SE = 1.35) than a typical immigrant to the UK (M = 52.5, SE = 1.26), $p < .001$, $\eta_p^2 = .21$. Participants also rated a typical immigrant as feeling antisocial emotions shared with other animals more strongly (M = 56.1, SE = 1.26) than a typical UK national (M = 52.5 SE = 1.37), $p = .017$, $\eta_p^2 = .04$.Similar to the trend of intergroup preference that emerged for how uniquely human emotions were perceived between groups, participants tended to rate a typical ingroup member as feeling shared emotions more strongly than a typical outgroup member, by about 6 points on the 100-point scale, when the shared emotions were prosocial. When they were antisocial, participants tended to rate a typical outgroup member as feeling them more strongly than a typical ingroup member by about 10 points. The results of Study 4.1 are illustrated in Figure 12.

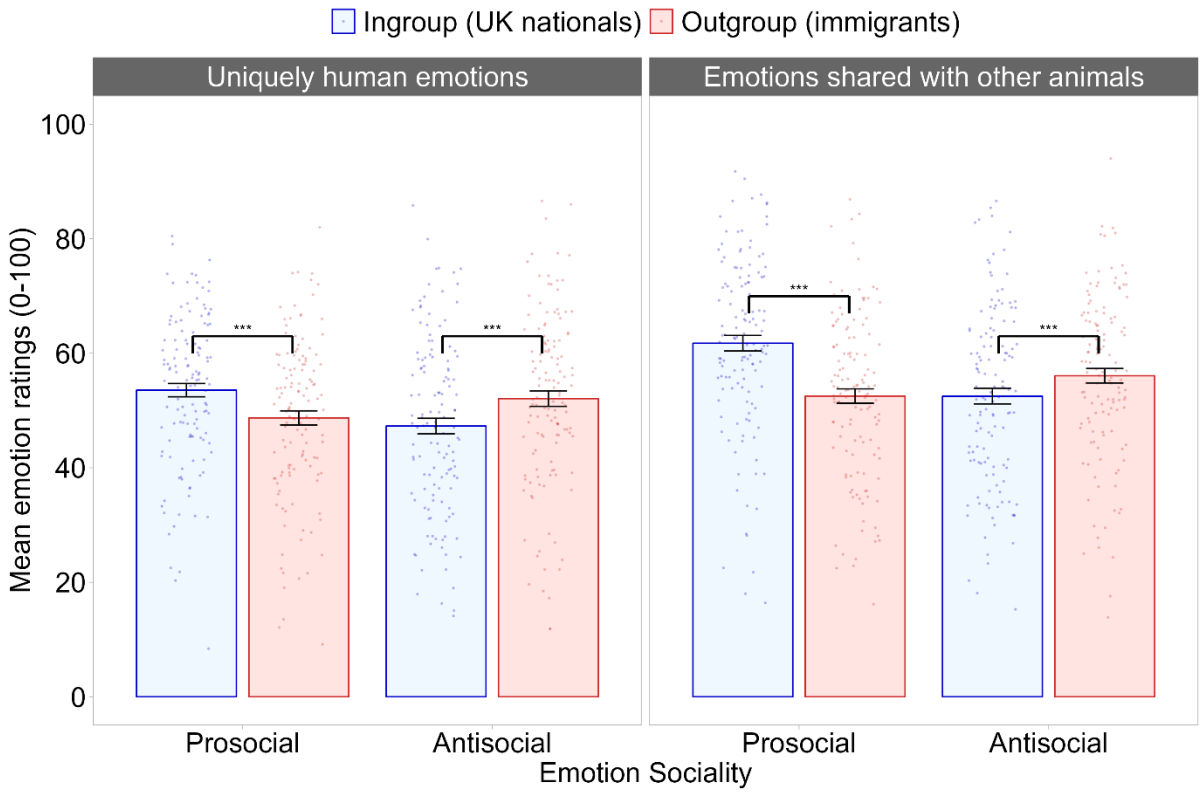

Figure 12. Results of Study 4.1

Rather than infrahumanization, participants attributed prosocial emotions more strongly to their ingroup (UK nationals) and antisocial emotions more strongly to their outgroup (immigrants), regardless of emotion humanness.

Note: Error bars represent ±1 *SE*. *** denotes $p < .001$.

Explicit Dehumanization

In order to check whether this is the type of context in which we would expect to see evidence of infrahumanization if it occurs, we compared explicit dehumanization ratings of UK nationals (the ingroup) and immigrants to the UK (the outgroup). A paired samples *t*-test revealed a significant difference in how evolved participants rated members of each group, $t(129) = 4.61$, $p < .001$, *Cohen's d* = .40. Participants tended to rate immigrants as significantly less evolved ($M = 83.3$, $SD = 23.1$) than UK nationals ($M = 91.4$, $SD = 12.1$). Average 'Ascent of Human' ratings for each group are plotted in Figure 13, confirming that this is an intergroup context where we might expect to observe evidence of infrahumanization if it occurs.

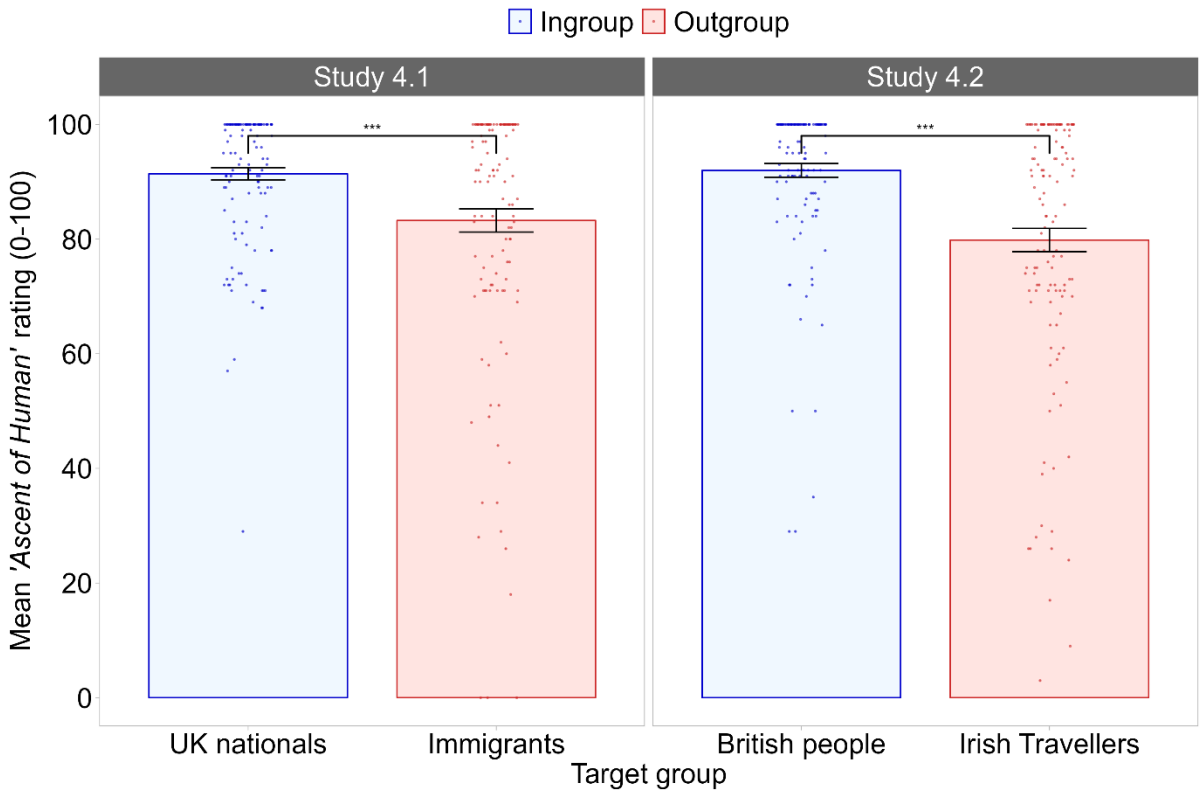

Figure 13. Explicit dehumanization ratings in Studies 4.1 and 4.2

Mean ratings for ingroup and outgroup members on the 'Ascent of Human' scale (Kteily et al., 2015) in Study 4.1 (left panel) and Study 4.2 (right panel).

Note: Error bars represent ±1 *SE*. *** denotes $p < .001$.

4.4. Study 4.2: Testing for Infrahumanization or Derogation of Irish Travellers

In Study 4.2, we tested our predictions in an additional social context to ensure generalisability. We examined whether participants would perceive a typical British person (ingroup) as feeling emotions differently than a typical Irish Traveller (outgroup), retaining the same hypotheses as in Study 4.1.

4.4.1. Method

Participants

The power analysis conducted for Study 4.1 also informed the sample size for Study 4.2. We recruited 130 participants for Study 4.2. All participants were adult British (English, Scottish, or Welsh) nationals who currently reside in England, Scotland, or Wales. This strict

inclusion criteria of ”British” was to avoid any biases that might arise should individuals from Northern Ireland take part, regardless of whether they identify as Irish or British. All participants were fluent in English, had a right-leaning political orientation, and had a minimal approval rating of 95% on Prolific. Data submitted by one participant who indicated that they came from an Irish Traveller background were omitted and replaced due to the bias this could carry (though when analyses were run including this participant's data, results were not different from those reported). In the final sample of 130 participants, 70 identified as male, 58 as female, one as nonbinary, and one chose not to indicate their gender identity. The participants' ages ranged from 20 to 81 (M = 47.6, SD = 14.5). Participation in Study 4.2 took an average of 11 minutes to complete (excluding two outliers whose duration of participation was more than three standard deviations above the mean).

Design, Procedure, and Materials

Study 4.2 followed the same design and procedure as in Study 4.1, with the same measures adapted slightly to refer to the new intergroup context of British people and Irish Travellers.

4.4.2. Results

Emotion Attribution

In Study 4.2, we ran the same analysis as in Study 4.1, which was a 2 (group: ingroup and outgroup) X 2 (emotion humanness: uniquely human and shared with other animals) X 2 (emotion sociality: prosocial and antisocial) repeated measures ANOVA. There was no significant main effect of group, $F(1, 129) = .062$, $p = .804$, $\eta_p^2 < .001$, or emotion sociality $F(1, 129) = .009$, $p = .924$, $\eta_p^2 < .001$. A significant main effect of emotion humanness was found $F(1, 129) = 318.676$, $p < .001$, $\eta_p^2 = .71$. Emotions shared with other animals (M = 60.8, SE = 0.89) tended to be rated higher than uniquely human emotions (M = 53.7, SE = 0.85).

A significant two-way interaction was found between group and emotion sociality, $F(1, 129) = 58.198$, $p < .001$, $\eta_p^2 = .31$. Pairwise comparisons revealed that participants tended to rate a typical UK national as feeling prosocial emotions more strongly (M = 61.8,

SE = 1.30) than a typical Irish Traveller (*M* = 52.6, *SE* = 1.52), *p* < .001, 95% CI [6.35, 12.11], η_p^2 = .24. However, antisocial emotions were rated as being felt more strongly by a typical Irish Traveller (*M* = 62.1, *SE* = 1.26) than by a typical British person (*M* = 52.5, *SE* = 1.52), *p* < .001, 95% CI [6.82, 12.37], η_p^2 = .27.

A significant two-way interaction was also found between group and emotion humanness, *F*(1, 129) = 36.617, *p* < .001, η_p^2 = .22. Pairwise comparisons revealed that participants rated a typical UK national as feeling uniquely human emotions more strongly (*M* = 54.6, *SE* = 1.01) than a typical Irish Traveller (*M* = 52.7, *SE* = 0.88), *p* = .030, 95% CI [0.19, 3.54], η_p^2 = .04. Participants also tended to rate a typical Irish Traveller as feeling emotions shared with other animals more strongly (*M* = 62.0, *SE* = 0.95) than a typical British person (*M* = 59.7, *SE* = 0.99), *p* = .003, 95% CI [0.75, 3.69], η_p^2 = .07. Importantly, each of the two-way interactions was qualified by a significant three-way interaction between group, emotion humanness, and emotion sociality, *F*(1, 129) = 24.814, *p* < .001, η_p^2 = .16.

The pattern of pairwise comparisons for the three-way ANOVA more strongly resembled that predicted by the intergroup preference account. Participants rated a typical British person as feeling prosocial uniquely human emotions more strongly (*M* = 58.7, *SE* = 1.34) than a typical Irish Traveller (*M* = 45.9, *SE* = 1.37), *p* < .001, 95% CI [9.69, 15.89] η_p^2 = .34. Crucially, participants tended to rate a typical Irish Traveller as feeling antisocial uniquely human emotions more strongly (*M* = 59.6, *SE* = 1.35) than a typical British person (*M* = 50.5, *SE* = 1.55), *p* < .001, 95% CI [-12.23, -5.90], η_p^2 = .20. Participants tended to rate a typical ingroup member as feeling uniquely human emotions more strongly than a typical outgroup member by roughly 13 points on the 100-point scale when the emotions were prosocial, and about 9 points less strongly when they were antisocial. This result shows that an intergroup preference towards seeing members of one's ingroup as more prosocial and outgroup members as more antisocial underlay any evidence of infrahumanization occurring, as the two-way interaction between emotion humanness and target group might suggest.

The pairwise comparisons for emotions shared with other animals revealed similar trends to those for uniquely human emotions. Participants tended to rate a typical British

person as feeling prosocial emotions shared with other animals more strongly (M = 64.9, SE = 1.34) than a typical Irish Traveller (M = 59.2, SE = 1.37), p < .001, 95% CI [2.72, 8.63], η_p^2 = .10. Participants also rated Irish Travellers as feeling antisocial emotions shared with other animals more strongly (M = 64.7, SE = 1.26) than a typical UK national (M = 54.6, SE = 1.55), p < .001, 95% CI [7.51, 12.73], η_p^2 = .31. Participants tended to rate a typical ingroup member as feeling shared emotions more strongly than a typical outgroup member, by about 6 points on the 100-point scale, when the shared emotions were prosocial. When they were antisocial, participants tended to rate a typical outgroup member as feeling them more strongly than a typical ingroup member by about 10 points. This trend is evidence of an intergroup preference driven by the social desirability of emotions, similar to that which emerged for how uniquely human emotions were perceived between ingroup and outgroup members. The results of Study 4.2 are illustrated in Figure 14.

Explicit Dehumanization

A paired samples *t*-test was conducted to test whether participants rated members of their ingroup (British people) as more evolved on the 'Ascent of Human' scale (Kteily et al., 2015, see Figure 1) than members of the outgroup (Irish Travellers). The *t*-test revealed a significant difference in how evolved participants rated members of each group, $t(129) = 6.819$, p < .001, *Cohen's d* = .60. As can be seen in Figure 13**,** participants tended to rate Irish Travellers as significantly less evolved (M = 79.8, SD = 23.1) than British people (M = 92.0, SD = 13.7)**.**

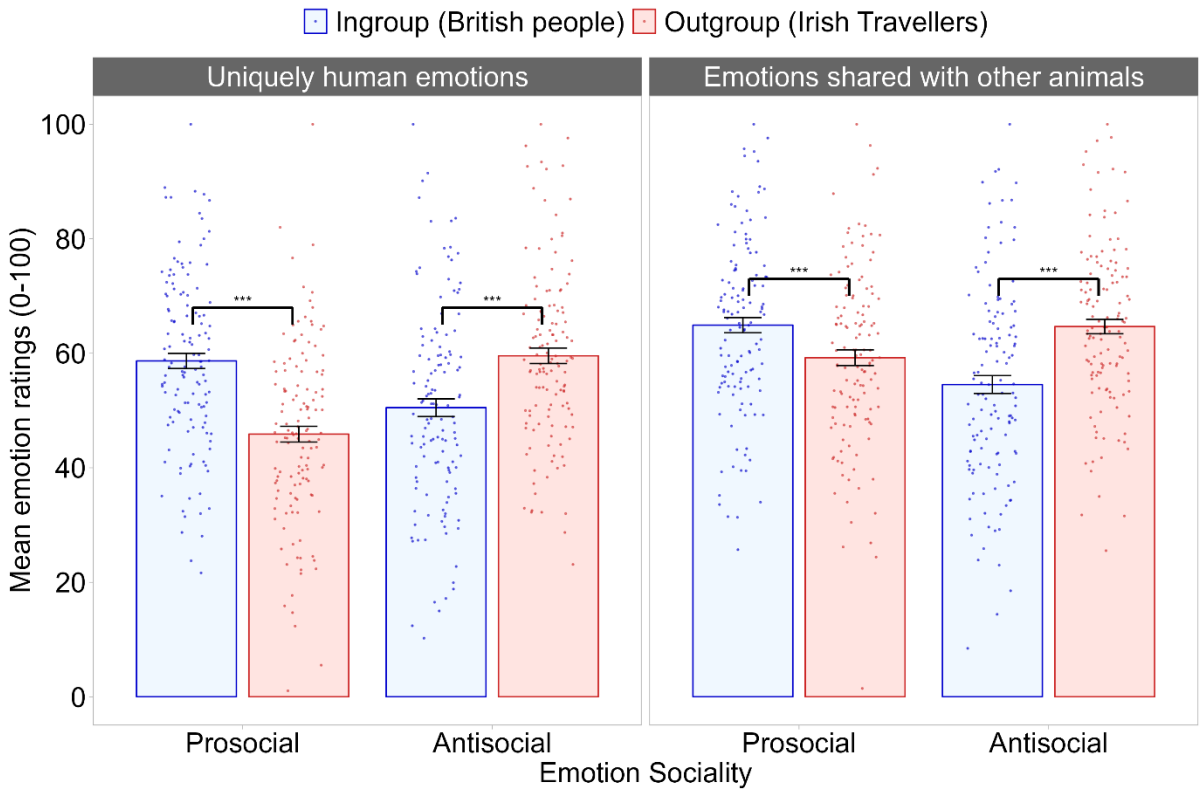

Figure 14. Results of Study 4.2

Participants attributed prosocial emotions more strongly to their ingroup (British people) and antisocial emotions more strongly to the outgroup (Irish Travellers), regardless of emotion humanness.

Note: Error bars represent ±1 *SE*. *** denotes $p < .001$.

4.5. Chapter 4 General Discussion

The two studies in this chapter converged on a similar pattern of results. Across two intergroup contexts, participants attributed prosocial emotions more strongly to their ingroup and antisocial emotions more strongly to their outgroup, regardless of whether or not they were uniquely human. We came to this conclusion using a study design that was more robust than most previous infrahumanization research by including a large set of prosocial and antisocial emotions in our measures. We also examined two different intergroup contexts. The first context involved nationals of a country (the UK) as the participant ingroup and immigrants as

the target outgroup, a context widely studied in dehumanization research (Banton et al., 2020; Hodson & Costello, 2007; Markowitz & Slovic, 2021; Prati et al., 2016). The second context involved British people as the participant ingroup and Irish Travellers as the target outgroup, who are one of many minority ethnic groups often labelled as “Gypsies”. Perceptions of such “Gypsy” groups have been examined in the infrahumanization literature (Dalsklev & Kunst, 2015; Kteily et al., 2015; Marcu & Chryssochoou, 2005; Martínez et al., 2012). However, it is worth noting that scarce research has examined the infrahumanization of Irish Travellers specifically.

Even though individuals in both samples explicitly dehumanized the outgroups on the 'Ascent of Human' scale (Kteily et al., 2015), our results provide no persuasive evidence of infrahumanization occurring. Although there was a two-way interaction between group membership and humanness in Study 4.2, this was qualified by a significant three-way interaction. Simple comparisons revealed a similar pattern as in Study 4.1, which suggested that when emotions are judged as antisocial, individuals attribute them more strongly to the outgroup regardless of emotion humanness. Given the larger number of emotions used in these studies and the replication across two intergroup contexts, it is unlikely that the particular influence of one or two emotions might have skewed our results (Enock & Over, 2022; Vaes, 2023). These results converge with recent challenges to infrahumanization theory and other social-psychological theories of subtle dehumanization. In particular, our results accord to those of Enock et al. (2021b) and Enock and Over (2022), who showed that apparent evidence for infrahumanization and its link to reduced prosocial behaviour towards the target is better explained by intergroup preference.

It is not our goal to suggest that infrahumanization never occurs. We used a commonly utilised design (Demoulin et al., 2004, 2005; Prati et al., 2015; Prati & Giner-Sorolla, 2018; Vaes et al., 2003) and two socially significant intergroup contexts in which there are strong theoretical reasons to predict that infrahumanization might occur (Banton et al., 2020; Esses et al., 2008; Hodson & Costello, 2007; Kteily et al., 2015; Markowitz & Slovic, 2020, 2021; Utych, 2018). However, it remains entirely possible that infrahumanization could be detected

with other experimental designs or in other intergroup contexts. One potential weakness of our studies was that they examined correlational, within-subjects data. While this allowed us to test our hypotheses with minimal noise within the data, a possible next step for this research would have been to add a between-subjects element. For instance, a mixed design could be adopted where half of the participants rate how a typical ingroup member experiences emotions and the other half a typical outgroup member, treating the target group as a between-subjects variable and making the comparison between groups less salient to participants (see Linville & Jones, 1980). Alternatively, emotion humanness and sociality could have been treated as between-subject variables, with experimental manipulation controlling which emotions are associated with which group using vignettes, for example. Including a between-subjects or mixed design could build upon the findings of Studies 4.1 and 4.2 as it may explain whether infrahumanization or intergroup preference might predict a third dependent variable, such as a measure of likability or intention to harm the target. Further research could examine the use of such mixed designs.

Partly due to the critiques raised by Enock and colleagues (2021b), researchers have suggested that it might be more appropriate to adopt an individual differences approach when investigating infrahumanization theory. Vaes (2023) suggests that rather than a consensus on which characteristics are considered uniquely human and shared with other species, there might be substantial variability in this judgment between individuals. In order to investigate whether British participants subtly dehumanize immigrants, he asked participants to rate the extent to which they perceived certain character traits to be unique to humans, typically prosocial, and characteristic of their ingroup and outgroup. He found a stronger correlation between the extent to which traits were perceived as unique to humans and typical of the ingroup than the extent to which they were perceived as unique to humans and typical of the outgroup, even while controlling for the social desirability of the traits. However, these results are difficult to interpret as participants reported liking the outgroup members more than the ingroup members, meaning the manipulation check was unsuccessful. This was likely due to the majority of participants in Vaes' study being left-leaning in their political orientation, who

tend to have more favourable views of the target outgroup, immigrants. We wanted to avoid this possible bias in our studies, thus, collecting data from only right-leaning participants seemed better suited when examining the occurrence of subtle dehumanization (see Crawford, 2012). Research suggests a right-leaning political orientation to be associated with more general prejudice, as well as towards the specific minority target outgroups examined (Banton et al., 2020; Duckitt & Sibley, 2007; Richards et al., 2023; Sánchez & García, 2016; Todosijević & Enyedi, 2002). Thus, infrahumanization should have had a better chance of being detected should it have occurred in our studies, as a right-leaning sample seemed to ensure our manipulation checks were successful. Furthermore, the test-retest reliability of the various individual differences is not yet clear. However, this general approach could be expanded from testing the dual model of dehumanization (Haslam, 2006) to other theories of dehumanization in the future, particularly infrahumanization theory.

Regardless of individual differences, our research places considerable constraints on what can be inferred from previous infrahumanization research. Combined with Enock et al.'s (2021b) findings, our results suggest that infrahumanization occurs considerably less frequently than generally assumed in the literature. When the sociality of the traits is controlled for, there is relatively little evidence for infrahumanization theory, at least across the multiple intergroup contexts explored in these papers. Importantly, our results show that the lack of evidence for infrahumanization in previous work is unlikely to be due to a small number of emotions being included in stimulus sets. The stimuli we developed in this chapter, incorporating a substantial number of emotions and appropriately controlling for sociality, could be used to test whether there is evidence for infrahumanization in other intergroup contexts.

Chapter 5: General Discussion

5.1. Thesis Scope and Aims

My doctoral research sought to critically examine theories of subtle dehumanization to better our understanding of intergroup perceptions and harm. The three key aims of the work presented in this thesis can be summarised as follows:

- Investigate the extent to which social-psychological findings on subtle forms of dehumanization replicate, specifically those relating to the dual model of trait-based dehumanization (Haslam, 2006) and infrahumanization theory (Leyens et al., 2000, 2001).
- Understand the extent to which earlier evidence of trait-based dehumanization and infrahumanization can be better explained by effects associated with the confound of social desirability, mainly intergroup preference and negative evaluation of the target.
- Critically evaluate the hypothesised causal relationship between trait-based dehumanization and harm.

I investigated these questions across 11 empirical studies (total $N = 1{,}748$). Taken together, these studies shed fresh light on our understanding of how subtle forms of dehumanization ought to be approached within social psychology.

5.2. Summary of Empirical Findings

5.2.1. Chapter 2: Is Outgroup Harm Predicted by Trait-Based Dehumanization or Negative Evaluation?

In Chapter 2, I sought to address the extent to which trait-based dehumanization predicts harm against an outgroup. Bastian et al. (2013) provided an empirical example of how trait-based dehumanization might predict harm while employing the dual model's framework (Haslam, 2006). These researchers found that the extent to which participants dehumanized criminals by denying them human character traits, such as *refined*, *sophisticated,* and *moral*, positively predicted the severity of the punishment endorsed for them. However, the selection

of traits included in Bastian and colleagues' measures were all socially desirable. Thus, apparent dehumanization cannot be disentangled from a negative evaluation of the criminal targets (Over, 2021a, 2021b). I first replicated the trend of greater dehumanization predicting harsher punishments endorsed in Studies 2.1A and 2.1B using only socially desirable human traits, as included by Bastian and colleagues, and as presented in the dual model framework (Haslam, 2006). In Studies 2.2A and 2.2B, I investigated whether this trend replicated while including uniquely human but socially undesirable traits in my measures, such as *jealous*, *spiteful*, and *corrupt*.

Contradicting conclusions drawn by Bastian et al. (2013), I found the relationship was reversed when undesirable human traits were included in the measures used. In the case of undesirable traits, the greater the extent to which participants attributed human qualities to outgroup members, the harsher the punishment endorsed. Using a combination of correlational and experimental methods, my results demonstrated that denying uniquely human traits predicted increased intentions to harm the target only when the traits included in the measures were socially desirable. When they were undesirable, denying uniquely human traits was instead associated with decreased intentions to harm the target. These trends emerged for traits shared with animals and traits shared with robots, which were included in the design of Studies 2.3A and 2.3B.

Overall, the findings of Chapter 2 evidence how trait desirability confounds the dual model's framework (Haslam, 2006) and the results Bastian et al. (2013) reported. Whether or not the character traits were uniquely human did not affect the endorsement of harm to criminal targets. Including a selection of traits balanced in their social desirability showed a negative evaluation of the target as the critical factor in making punishment decisions.

5.2.2. Chapter 3: Infrahumanization Following Harm or Recognising Negative Emotions when Harmed?

The studies detailed in Chapter 3 investigated whether subtle dehumanization can sometimes occur as a consequence rather than as a cause of intergroup harm. Previous

studies have reported infrahumanization following ham (Castano & Giner-Sorolla, 2006; Čehajić et al., 2009). These researchers suggested that people tend to see members of an outgroup that their ingroup has harmed as typically feeling uniquely human emotions to a lesser extent when reminded of their ingroup's role in the harm, compared to when they are not. I sought to better understand this effect by replicating it in novel intergroup contexts (the harm caused by climate change and the fast fashion industry). However, no evidence of infrahumanization occurring as a response to feeling responsible for harming members of an outgroup was found in the two contexts examined. I concluded that the previously reported trend of infrahumanization following harm did not replicate while using a factorial design with 16 emotions in the measures used, as in the design of Čehajić and colleagues (2009) in either context. The design of Čehajić and colleagues was used in Studies 3.1 and 3.2. Infrahumanization following harm also failed to replicate in Study 3.3, which had a more elaborate design with 64 emotion items in the measures used and emotion sociality and emotion humanness being treated as continuous variables - the approach taken by Castano and Giner-Sorolla (2006). The same participant sample scored each of the 64 emotions along these emotion dimensions in the latter design before they responded to the dependent measure of outgroup emotion attribution.

Across all three studies, there was no evidence of infrahumanization following harm. Instead, harmed outgroup members were seen to feel more negative emotions, regardless of whether ingroup responsibility for harm was made salient or the humanness of the emotions included. The trends observed in Chapter 3 suggest that previously reported effects of infrahumanization following harm do not replicate reliably or are not as generalisable across intergroup contexts as claimed previously. These findings highlight a need for caution when interpreting past results.

5.2.3. Chapter 4: Infrahumanization or Intergroup Preference?

The studies reported in Chapter 4 of this thesis revisited the classic paradigm of infrahumanization theory (Leyens et al., 2000, 2001), which posits that outgroup members are perceived as experiencing uniquely human emotions to a lesser extent than members of one's

ingroup. I investigated a methodological weakness in previous research on infrahumanization, which is that previous work typically measures attributions of uniquely human emotions that are prosocial, omitting those that are antisocial (Over, 2021a, 2021b). Infrahumanization research has also been criticised for incorporating a small number of emotion items in the measures used, typically around 12 in total (e.g., Cortes et al., 2005; Kteily et al., 2015; Paladino et al., 2002; Vaes et al., 2002; Viki & Calitri, 2008). A small selection of emotions could be an issue, as some critics have suggested that the influence of just one or two items in a small selection could disproportionately skew the results (Enock & Over, 2022; Vaes, 2023). Thus, I sought to investigate the replicability of infrahumanization while controlling for emotion sociality and including 68 emotions in my measures, roughly four times more than is standard in the field.

The first study examined emotion attributions to UK nationals (ingroup) and immigrants to the UK (outgroup), which is a context in which infrahumanization has previously been reported (Banton et al., 2020; Prati et al., 2016). In the second study, I examined the responses of British people (ingroup) to Irish Travellers (outgroup). I did not find evidence for the infrahumanization of outgroup members in either study. Each study employed a factorial design, as standard in infrahumanization research, with 68 emotions in the ingroup and outgroup emotion attribution measures. In both studies, participants attributed prosocial uniquely human emotions more to their ingroup and antisocial uniquely human emotions more to the outgroup, regardless of emotion humanness. Corresponding to the trends reported by (Enock et al., 2021b; Enock & Over, 2022), these findings lend empirical support to recent critiques of dehumanization research more broadly (Over, 2021a, 2021b). By accounting for emotion sociality as a confound to prior infrahumanization research, I showed how ingroup preference better explains emotion attributions to ingroup and outgroup members than infrahumanization.

5.3. Interim Conclusions

5.3.1. Social Desirability Confounds Social-Psychological Models of Subtle Dehumanization.

The works detailed in Chapters 2 and 4 of this thesis demonstrate the confounding role of social desirability in current models of subtle dehumanization, specifically the dual model's characterisation of trait-based dehumanization (Haslam, 2006) and infrahumanization theory's characterisation of emotion-based dehumanization (Leyens et al., 2000, 2001). Practically speaking, including traits or emotions that varied in terms of their sociality negated any trends indicative of subtle dehumanization.

By demonstrating the confounding role of social desirability in existing models of the dual model and infrahumanization theory, my doctoral work provides support to broader critiques of social-psychological literature on dehumanization (Bloom, 2022; Over, 2021a, 2021b; Smith, 2021a, 2023). Over (2021a, 2021b) previously pointed out that measures of dehumanization developed from the dual model and infrahumanization theory tend to rely on prosocial or desirable aspects of humanity while omitting undesirable or antisocial aspects. As a result, Over notes, we cannot confidently claim that dehumanization occurs independently from effects driven by social desirability, such as intergroup preference (see Brewer, 1999; Turner & Tajfel, 1979). The findings of Chapters 2 and 4 support this critique, demonstrating that social desirability does indeed confound these frameworks.

Discussing explicit forms of dehumanization, Smith (2021a) notes that current theories do not adequately distinguish between genuine dehumanization (seeing another person or group as less than human or nonhuman) and simply degrading rhetoric (see also Smith, 2020). In response to a framework of dehumanization proposed by Kteily and Landry (2022), Bloom (2022) similarly calls upon social psychology to distinguish between genuinely perceiving and treating people as subhuman and seeing them as lacking in prosocial qualities compared to a “human ideal”. I argue that these critiques also apply to subtle forms of dehumanization, as my work shows trait-based and emotion-based forms of it to be underpinned by the perceiver's

tendency to derogate members of an outgroup (seeing them as possessing more antisocial and less prosocial aspects of humanity) rather than seeing them as subhuman.

A recent review of the infrahumanization literature by Bunce et al. (2024) further highlights a need for the field to re-evaluate apparent empirical evidence of subtle dehumanization. In line with the findings of Chapter 4 and other recent empirical works (Enock et al., 2021b; Enock & Over, 2022), Bunce and colleagues conducted a systematic review of the literature, combined with novel data collection, and concluded that social desirability confounds close to 80% of published infrahumanization studies, whereby the uniquely human emotions in the measures used are significantly more prosocial than the emotions shared with other animals. Considering my findings in Chapter 2 and other recent evidence of trait desirability confounding the dual model's framework (Enock et al., 2021a; Enock & Over, 2023), it is reasonable to assume the same issue might permeate published works claiming to evidence trait-based dehumanization. Thus, the challenge remains for the field to address the confound of social desirability in models of subtle dehumanization and the validity of related findings.

5.3.2. Subtle Dehumanization Models Contribute less to our Understanding of Intergroup Harm than Previously Thought.

Turning to dehumanization as a construct more broadly, seeing someone as subhuman has often been associated with harming the target, whether through active transgression, support for violence, or refusing them help (Bar-Tal, 2000; M. L. Fisher, 2012; Harris & Fiske, 2011; Haslam, 2006, 2019; Haslam & Loughnan, 2016; Kteily & Landry, 2022; Luft, 2015, 2019; Maoz & McCauley, 2008; Rudman & Mescher, 2012; Savage, 2013; Smith, 2011, 2020, 2021b; Stanton, 1998; Staub, 1989; Viki et al., 2013; Zimbardo, 2007). Studies employing models of subtle dehumanization have supported this view (Bastian et al., 2013; Cuddy et al., 2007; Vaes et al., 2003). However, the conceptual shortcomings of the dual model (Haslam, 2006) and infrahumanization theory (Leyens et al., 2000, 2001) highlighted in my work challenge the hypothesised causal link between subtle forms of dehumanization and harm (see Haslam, 2019; Haslam & Loughnan, 2015).

My work in Chapter 2 demonstrates that someone is at greater risk of harm if seen as possessing undesirable human traits rather than lacking in human traits altogether. Other work has similarly demonstrated how apparent evidence for a link between infrahumanization and harm is similarly confounded, where recognising antisocial human emotions predicted harm towards the target rather than infrahumanization (Enock & Over, 2022, 2023a). Together, my findings and those of previous research provide support for arguments made against the causal relationship between these characterisations of dehumanization and harm. Critics have noted how this relationship is inconsistent and that intentions to harm someone might be better explained by recognising their humanity (Bloom, 2017; Lang, 2010, 2020; Manne, 2016, 2018; Over, 2021a, 2021b). In other words, those we appear to dehumanize might instead be perceived as “the bad guys”, exhibiting the less accepted human qualities like corruption or resentment, making us more likely to be hostile or harmful towards them. There was apparent evidence of such a trend in Study 2.3A, where the attribution of undesirable human traits predicted greater harm than shared undesirable traits. Relatedly, critics have also noted plenty of examples where animals, objects, and people considered lacking in uniquely human qualities do not tend to be harmed for that reason (Over, 2021a; Smith, 2021b).

My findings support the view that recognising certain aspects of humanity in another, rather than dehumanization, can motivate harm towards them through derogation and negative evaluation. Such an observation can help explain the underlying processes behind much of the social inequalities and intergroup hostilities we see today, which Smith (2023) suggests dehumanization research tends to minimize or overlook (see also Ng, 2021). Lang (2010, 2020) notes that one group's intentions to dominate another and maintain a hierarchy involve a social relationship that requires the perpetrator of harm to recognise the target's humanity. On a similar note, Davis (1981) suggests that the recognition of Black US Americans' humanity, rather than their imposed lesser-human status relative to their White counterparts, explained historic efforts to bar them from formal education (p. 101). In the context of gendered violence, Manne (2016, 2018) argues that a man’s misogynist motivation to enforce patriarchal superiority through violence and oppression is also rooted in the

recognition of the humanity of women. These motivators behind intergroup harm could help explain our findings and those of Rai et al. (2017), who concluded that “dehumanization removes the very qualities that make moral violence meaningful” (p. 8514). Aligning with my findings, these alternative approaches to intergroup harm seem to provide a better explanation and more practical insight into its causes than apparent dehumanization.

Consequentially, the importance of the dual model and infrahumanization theory in improving our understanding of urgent social phenomena such as prejudice and discrimination seems more limited than previously thought. A large body of social-psychological literature has already established that intergroup preference plays a decisive role in how we perceive and treat members of outgroups. For instance, the highly influential social identity theory (Tajfel & Turner, 1979) suggests that we are implicitly motivated to perceive negative aspects of outgroup members to see our ingroup as more virtuous and prosocial, boosting our self-image. Such a paradigm has been widely employed to explain intergroup prejudice and harm (see also Brewer, 1999; Reicher et al., 2008). While no theory is without limitations, trends aligning with this view are well-documented and often replicated within social psychology and beyond (Hornsey, 2008). As my work demonstrates, models of trait- based and emotion-based forms of dehumanization are confounded, often measuring intergroup preference rather than dehumanization. Therefore, the field must re-assess the internal validity and unique contributions of such models of subtle dehumanization to the social-psychological literature.

5.3.3. Previous Findings Based on Subtle Dehumanization Models have Replication Issues.

Across the 11 studies reported in this thesis, I did not find evidence of trait-based or emotion-based subtle dehumanization as characterised by the dual model (Haslam, 2006) or infrahumanization theory (Leyens et al. 2000, 2001), respectively. Previous findings are initially replicable, such as those of Bastian et al. (2013), but only when using designs confounded by social desirability. Controlling for this confound, in turn, nullified any evidence of dehumanization occurring. Aside from the confound of social desirability, I also show that

some findings fail to replicate altogether in at least some intergroup contexts, as in the case of infrahumanization following harm (Castano & Giner-Sorolla, 2006; Čehajić et al., 2009).

My doctoral work provides compelling evidence of internal validity and replication issues within the frameworks of subtle dehumanization and associated findings. Whether due to a general replication issue or a reliance on confounded designs to be replicable, the key findings regarding subtle forms of dehumanization remain questionable. As psychological science has been forced to reckon with a replication crisis over the last decade, social psychology has been at the centre of the storm (Open Science Collaboration, 2015). There has been pressure on social psychologists to address the many theories and effects developed in the age before open science was standard practice, with Kahneman (2012, as cited in Diener & Biswas-Diener, 2024) stressing the need for social psychologists to "act as a group and avoid defensiveness" (p. 2). Thus, future research needs to inspect claims made in the field concerning these phenomena, particularly those that pre-date the era of open science, while accounting for the confounds evidenced in this thesis and related work (Bunce et al., 2024; Enock et al., 2021a, 2021b; Enock & Over, 2022, 2023; see also Bloom, 2017, 2022; Over, 2021a, 2021b; Smith, 2023). Providing sound and replicable research to tackle pressing real-world social issues like prejudice and intergroup harm is essential for social-psychological theories.

5.4. Suggestions for Future Research

5.4.1. Summary of Limitations and Methodological Recommendations for Future Research

Several limitations were noted across the three empirical chapters of this thesis, and methodological recommendations for future research were made, which can help inform further research to examine subtle forms of dehumanization. Research informed by the limitations and conclusions of this thesis could help address the theoretical and methodological redundancies in subtle dehumanization research, steering future work on intergroup attitudes and harm towards more robust and replicable findings.

The most crucial recommendation is that researchers control for the confound of social desirability in the current frameworks of the dual model and infrahumanization theory (Over, 2021a, 2021b). Accounting for this confound would better allow us to assess apparent dehumanization independently of intergroup preference or a negative evaluation of the target (see Enock et al., 2021a, 2021b). As recommended by Enock and Over (2022) and Vaes (2023), including a large number of traits or emotions in the measures of subtle dehumanization would minimise the risk of one highly influential item skewing results. Ensuring a balanced selection of emotions or traits is essential when including many items and controlling for social desirability. A balanced selection would necessitate that, on average, uniquely human items are significantly higher in humanness than non-uniquely human items but do not differ in social desirability. Likewise, socially desirable items should be significantly more desirable than undesirable items but do not differ in humanness.

Manipulation checks such as the ‘Ascent of Human’ scale (Kteily et al., 2015) are encouraged when examining subtle forms of dehumanization to show that the examined context is appropriate for detecting dehumanization. Including manipulation checks relevant to the stimuli, such as perceived harm and sense of collective responsibility in Chapter 3, is particularly important when replicating previous findings as they ensure that the design and intergroup context being examined are suitable for detecting subtle forms of dehumanization. However, the lack of an explicit dehumanization measure in this research was a weakness, as it would have helped to clarify the appropriateness of the examined intergroup contexts further. Further research should examine how a humanizing perception of another person or group might encourage harm towards them (Lang, 2010, 2020; Manne, 2016, 2018; Over, 2021a, 2021b). Accounting for broader societal influences on intergroup attitudes, such as social hierarchies or stereotypes of the target (Ng, 2021; Smith, 2023), is encouraged to understand further how humanization might predict harm rather than dehumanization.

As dehumanization research usually examines advantaged group members' perceptions of lesser-privileged group members in a society, it seems necessary to account for the possible power distance, social hierarchies, and hostilities in each context. The

omission of these societal influences in the frameworks used to examine trait and emotion attributions to outgroup members might have predicated previous evidence of dehumanization being found. My doctoral research, in turn, inherits this limitation as I sought to replicate previous designs conceptually. Future research should attempt to account for these societal influences to help clarify the role of intergroup attitudes and negative evaluation in intergroup perceptions. It also remains possible that evidence for a dehumanizing perception in line with the dual model or infrahumanization theory might arise should such variables be accounted for, although it could also further rebut the assumptions of these frameworks. It is also important to note that, like the empirical works I sought to replicate, the research presented in this thesis is confined to lab-based testing. Thus, my findings and conclusions concerning intergroup biases that may emerge in such conditions have limited generalisability and would be inappropriate if applied to cases of extreme intergroup harm in the real world.

One possible way to overcome the limitations of lab-based testing could be incorporating archival research into the toolkit used to examine the social world. These could include existing natural experiments, true experiments, quasi-experiments, and correlational data, which Heng et al. (2018) note could help social psychologists examine the real world more directly than artificially simulating them in a solely lab-based approach. Heng and colleagues highlight how this could be particularly beneficial for research concerning ethically sensitive phenomena, an evident concern when examining intergroup attitudes and harm (see also Reis & Gosling, 2010). The use of virtual environments in assessing social interactions could carry similar strengths when researching intergroup attitudes and behaviours, retaining a lab-based approach while allowing researchers to have greater control over the social contexts and confederate identity and, to a degree, participant identity (McCall & Blascovich, 2009). Incorporating virtual reality into study designs could allow researchers to better create and capture real-world social behaviour than current designs in dehumanization research based on self-report measures and hypothetical intergroup contexts (see Bailenson et al., 2004; Sterna & Zibrek, 2021).

These methodological recommendations could help future research delineate the effects of subtle dehumanization from those of social desirability and address the replicability issues highlighted in my doctoral research. Testing for evidence of subtle dehumanization in more intergroup contexts using more robust study designs with appropriate controls and pre-registration is needed to better understand the usefulness of subtle dehumanization theories.

5.4.2. Using a Wider Range of Designs

Individual Differences. Whether an emotion or trait is definitively unique to humans is not always clear. Even among published studies, there is variability in this regard. For example, *disgust* has been included in some study designs as an emotion shared with other animals (Prati & Giner-Sorolla, 2018; Rodríguez-Torres et al., 2005) but in other studies as a uniquely human emotion (Bain et al., 2009; DeLuca-McLean & Castano, 2009). Individual differences may exist in what traits and emotions participants consider uniquely human. Adopting an approach that accounts for these individual differences in categorisation might help account for such variations. For example, Vaes (2023) tested for trait-based dehumanization by examining how within-participant correlations between ingroup trait typicality and trait humanness ratings compared to correlations between outgroup typicality and the same humanness ratings. Vaes notes that while within-participant correlations might be small due to them relying on individual judgements rather than group means, they are better powered due to more observations being analysed. Vaes concluded that this more elaborate testing method for trait-based dehumanization provided evidence of its occurrence independent of the effects of intergroup preference (Enock et al., 2021a). However, this finding is difficult to fully interpret because participants perceived the outgroup as more positive than the ingroup overall, meaning they apparently dehumanized the group they liked more.

Accounting for individual-level differences in variables when testing for trait-based dehumanization and infrahumanization seems a valuable addition to the current methods used in the field. It will also be necessary to examine the replicability of Vaes' (2023) findings using this method across other intergroup contexts. Further replication would also help us

understand the test-re-test reliability of an individual measures approach compared to the standard group means approach.

Implicit Measures. Most dehumanization research has examined explicit judgements, where participants are asked to rate their ingroup and outgroups in terms of their characteristics or experiences. The research reported in my thesis relied on explicit measures of this kind, specifically trait attributions and perceived emotional experiences. While measures of subtle dehumanization are said to capture unconscious perceptions of the target as less human, they are still vulnerable to social desirability biases in practice. For example, someone might be aware that their views of a particular minority group are racist, which they might try to avoid showing when responding to questions about that group. Implicit methods are less vulnerable to such response biases as they measure dehumanizing mental associations without the participant necessarily being aware of them (Haslam, 2021, p. 135). Implicit association tests (IATs) typically involve participants making quick judgements about whether traits or other stimuli presented on a monitor are typical or atypical of a social group. Some research that has used IATs suggests that certain social groups are mentally associated with animalistic traits (Saminaden et al., 2010) or animal-related words (Viki et al., 2006). Other research employed priming effects to see if participants' mental associations between a social group and animals could be revealed through experimental manipulation (Goff et al., 2008).

Future research could investigate implicit measures in more detail to determine how much we might mentally associate others with being less or nonhuman outside our awareness. However, such designs need to control for intergroup preference. My thesis demonstrates the importance of having a selection of traits or emotions balanced regarding their social desirability when examining subtle forms of dehumanization using explicit measures. Similar controls should be in place with implicit measures, as the extent to which social desirability might similarly confound them is unclear.

Qualitative Approaches. The research presented in this thesis relied entirely on quantitative approaches to measuring and analysing data concerning the extent to which participants thought other people possessed character traits or felt emotions. Although less

common than quantitative methods in the field, qualitative approaches have been used in empirical examinations of dehumanization through thematic analysis (Durrheim et al., 2016), discourse analysis (Sakki & Castrén, 2022), thematic synthesis (Nielsen et al., 2023), as well as mixed methods such as computational linguistic analysis (Mendelsohn et al., 2020), and similitude analysis (Diniz et al., 2020). Scarce empirical work has incorporated recent critiques of dehumanization research into qualitative designs to differentiate between a genuine denial of humanity to others and attributing them negative human qualities (Smith, 2016, 2021b). It is worth noting that a few empirical studies of this kind have found the latter to be more predominant in supposedly dehumanizing propaganda (Enock & Over, 2023b; Landry et al., 2022). Incorporating a qualitative or mixed methods approach when examining subtle dehumanization could also improve the external validity of the measures and associated findings, helping to overcome the restrictions of lab-based findings and better understand real-world cases of intergroup perceptions and harm. Broadening the methods in which trait-based dehumanization and infrahumanization are measured might also provide further evidence of intergroup preference and negative evaluation, in line with my findings, rather than denying human qualities overall: an avenue for future research to explore.

5.4.3. Investigating Other Intergroup Contexts

In my thesis, I explored a range of intergroup contexts, both novel and previously examined in dehumanization literature. The respective ingroup and outgroup contexts I examined were people who have never been in prison and criminals, residents of the UK and residents of Lagos, UK consumers and textile workers in fast fashion supply chains, UK nationals and immigrants to the UK, and British people and Irish Travellers. However, social-psychological accounts of dehumanization have examined many more intergroup contexts. For example, targets of subtle forms of dehumanization in the literature have included those that differ from us in terms of age (Boudjemadi et al., 2017; Wiener et al., 2014), body characteristics (Deska et al., 2018; Kunst et al., 2017), national identity (Andrighetto et al., 2014; Bain et al., 2009), disability (Capozza et al., 2016; Rodríguez et al., 2016), political beliefs (Cassese, 2020; Pacilli et al., 2015), sexuality (MacInnis & Hodson, 2012), gender

(Gaunt, 2013; Viki & Abrams, 2003), and ethnicity (DeLuca-McLean & Castano, 2009; Vala et al., 2009). Studies have also examined inter-individual rather than intergroup contexts, suggesting we may subtly dehumanize those with whom we share intimate relationships (Karantzas et al., 2022; Rodrigues et al., 2021).

Turning to dehumanization research more broadly, examining novel contexts where subtle forms of dehumanization have not been empirically studied could help to determine the universalism of the conclusions drawn in previous work. Similarly, trying to replicate previous findings in contexts already examined would lead to more certainty as to whether members of a particular group indeed see members of another group as less human or as having more antisocial or undesirable aspects of humanity than ingroup members. The more replications of such findings in the field, the more contextually nuanced our understanding of these pressing intergroup phenomena will become.

5.4.4. Investigating Other Accounts of Dehumanization

Mental State Attribution. The research reported in this thesis focused on two prominent accounts of subtle dehumanization: the dual model (Haslam, 2006) and infrahumanization theory (Leyens et al., 2000, 2001). Another prominent characterisation of subtle dehumanization that I did not investigate is the mental state attribution account. According to this framework, dehumanization occurs when another person or group are perceived as lacking in higher-order cognition and socio-emotional capacity (Harris & Fiske, 2011). Gray et al. (2007) suggested that we perceive the mental states of others along two dimensions: *agency* (one's autonomy and capacity to plan) and *experience* (one's subjective ability to sense and feel). A failure to recognise the mental capacities of others is seen as a subtle form of dehumanization in this line of research (Epley et al., 2007). The stereotype content model (Fiske et al., 2002) similarly posits that seeing others lacking interpersonal warmth or competence is a subtle form of dehumanization (see also Harris & Fiske, 2006).

Whether through intergroup preference or a negative evaluation of the target, I found that members of an outgroup were not denied human traits or seen as experiencing human emotions to a lesser degree, contrary to the dual model and infrahumanization theory.

Therefore, it could be possible that research on mental state attribution might be similarly confounded, whereby attempts to measure the perception of human mental states might instead measure how socially desirable or normatively acceptable mental states are perceived in others. This hypothesis concurs with critiques of dehumanization research. Over (2021) notes that it is unclear why denying mental states such as those mentioned above to outgroup members equates to dehumanization, whereas attributing them with antisocial mental states such as *cunning*, *scheming* or *lacking in self-control* does not (see also Manne, 2016, 2018).

It would be interesting for future research to investigate these challenges to the mental state attribution account, as well as how critiques of broader dehumanization research might apply to this approach (Bloom, 2017, 2022; Lang, 2010, 2020; Over, 2021a, 2021b; Rai et al., 2017; Smith, 2023). Incorporating tasks into designs could be one method of tapping into the extent of mental state attribution to ingroup and outgroup members. The director task (Krauss & Glucksberg, 1977) has previously been used to examine the mentalisation of others. Despite some uncertainty about the nature of the mentalisation strategies explained by the director task (see Rubio-Fernández, 2017), it could be an interesting option to explore in the context of mental state attribution. Methods that account for the confounds of social desirability also need to be considered in future research examining mental state attributions.

Other Conceptualisations of Dehumanization in Social Psychology. There are several ways dehumanization has been characterised in the social-psychological literature that I did not examine. Prominent examples include explicit animalistic slurs, comparisons to an “ideal human”, and perceptual dehumanization. Animalistic slurs are the use of metaphorical language or images that draw explicit comparisons between a person or social group and animals, such as Nazi descriptions of the Jewish population as invasive insects or rats in 1940s Germany (Smith, 2011; Tirrell, 2012). Kteily and Landry (2022) suggested that a person might be explicitly described as less human to the extent that they are seen as lacking in traits applicable to the “ideal human”, mainly *intelligent*, *emotional*, *moral*, *rational*, and *civilised*. Perceptual dehumanization can be summarised as the activation of neural networks associated with perceiving objects and nonhuman animals rather than those associated with

the perception of humans when viewing a person's image (Deska et al., 2018; Fincher & Tetlock, 2016).

However, concerns have been raised regarding each of these frameworks. Animalistic slurs were shown by Enock and Over (2023) to influence a negative evaluation, rather than dehumanization, of the target. Comparisons to an “ideal human” have been criticised for conflating dehumanization with seeing someone as lacking in positive attributes, essentially rendering dehumanization a redundant term as “if everything is dehumanization, then nothing is dehumanization” (Bloom, 2022). Perceptual dehumanization research has been criticised for relying on contested methods and stimuli within vision and cognitive sciences, casting doubt on the conclusions drawn (Eggleston et al., 2023; Over & Cook, 2022).

These characterisations of dehumanization are beyond the scope of this thesis. However, they illustrate the current rapid pace at which developments in the field occur beyond the cornerstone theories of trait-based dehumanization and infrahumanization. For instance, Phillips (2022, 2023) has recently proposed a dual framework of blatant dehumanization in response to the critiques of prior dehumanization research. Phillips differentiates between *descriptive dehumanization*, referring to someone as biologically nonhuman or less human, and *normative dehumanization*, referring to someone as lacking a commitment to act morally or humanely.

Novel approaches to the construct of dehumanization and constructive critiques that challenge earlier conceptualisations each provide opportunities to better our understanding of intergroup and inter-individual perceptions of humanity. Future research could account for the impact of animalistic slurs, descriptive and normative dehumanization, or measures of perceptual dehumanization alongside measures of subtle dehumanization. Such a multi-faceted approach to understanding when and if it is possible to see another person as less or nonhuman at the psychological level would be an exciting direction for future research. It would also be necessary to examine how the confounds and shortcomings concerning research on subtle dehumanization identified in my thesis might overlap with these other conceptualisations.

5.5. Conclusion

It is important to emphasise that the findings and points made in this thesis concerning dehumanization refer exclusively to social-psychological models that propose subtle forms of the construct (Haslam, 2006; Leyens et al., 2000, 2001). I do not intend for any content to be interpreted in the broader philosophical, legal, and sociological interpretations of dehumanization, nor do I intend to minimise the urgency of addressing the pressing societal issues of intergroup hostility, prejudice, and harm. Rather, this body of work intends to improve our understanding of intergroup perceptions by assessing the validity of the claims made regarding subtle forms of dehumanization, considering recent critiques. As articulated by Over (2021b), this is to improve our overall understanding of the psychological underpinnings of discriminatory behaviour, bringing us closer to tackling its prevalence in today's world.

Whether research on trait-based dehumanization and infrahumanization theory uniquely contribute to our understanding of intergroup perception beyond ingroup preference and negative evaluation remains unclear. My doctoral research addressed conceptual and replicability concerns for this area of research, highlighting the confound of social desirability in the measures used when examining such constructs. Future research should examine subtle forms of dehumanization as a mechanism of intergroup perceptions while taking steps to control for the confounds highlighted in this thesis. If and how subtle dehumanization is better explained by social phenomena already well documented in the literature, such as intergroup preference, also warrants more attention. Such steps are especially pertinent given the empirical influence of the dual model and infrahumanization theory within and beyond social psychology, such as informing interventions to reduce intergroup antagonism and prejudice (for a recent example, see Sin et al., 2023). Understanding the psychological mechanisms underlying intergroup hostility and harm is paramount for social psychology. Thus, the theories we use must be replicable and robust to ensure they accurately reflect the psychological determinants of the social world around us, lest we risk hampering the field’s ability to tackle pressing social issues by metaphorically barking up the wrong tree.

Appendices

Appendix A: Supplementary Materials for Chapter 2

A.1. Method: Vignettes in Full

Studies 2.1A and 2.2A: Violent Criminals Vignettes

Note: Vignettes 1 and 2 are adapted versions of those used in Bastian et al. (2013).

1) Alex attempted to hijack a bus carrying 50 passengers by threatening the driver with a chisel. After boarding the bus, Alex forced the driver off and told passengers to stay put. Police arrived shortly after and Alex injured a bystander with the chisel while fleeing. When a local resident phoned the police saying they had been threatened by someone hiding in their garage, Alex was found with the chisel still in hand.

2) Sam recently threatened seven young children and injured two adults with a pocket knife at a preschool. Sam had rented a house to the proprietor of the preschool and was upset that the property was not vacated in April when the lease expired. Sam carried out the attacks at the start of a school day before being taken into custody.

3) Robin physically assaulted someone in an alley of a residential suburb last night. A witness saw Robin corner the victim with a pistol, holding them at gunpoint and hitting them across the head with the handle of the gun. Robin eventually shot the victim in the leg as they ran away, before also fleeing the scene. Police soon caught up with Robin and took them into custody.

4) Charlie was arrested after a fight broke out in a pub soon after opening time, apparently triggered by a minor disagreement. Charlie smashed a pint glass and used it to stab another customer. Two additional customers received cuts as they tried to hold Charlie back until the Police arrived.

5) Jamie was recently arrested after stabbing their partner with a kitchen knife and chasing them out of their shared flat. Their partner escaped and found safety with a neighbour who heard the commotion and ushered them inside. As they waited for the police and an ambulance to arrive, Jamie threatened both their partner and neighbour while attempting to break down the neighbour's front door.

Studies 2.1B and 2.2B: Thieves Vignettes

1) Alex was caught breaking into a student-rented house over the Christmas holidays. Having gained access by smashing the kitchen window, Alex broke into all five bedrooms, stealing any valuables left there over the holidays (including TVs, electronic devices and jewellery). Police apprehended Alex before they left the house with a car-load of the students' belongings after a passer-by reported the suspicious activity.

2) Sam has stolen close to £8,000 from school grants and fundraisers during their eight years working at a local school. Suspicions were raised when there was an unexplained shortfall in their target amount raised at a recent fundraiser for disadvantaged families in the community. Police soon found that Sam had stolen funds raised at over 30 similar events, as well as multiple council grants such as those intended to help accommodate students with disabilities.

3) Robin is a car dealer and garage owner who committed motor vehicle theft on multiple occasions. Robin targeted a range of vehicles and often modified their features before putting them up for resale at their dealership. Robin was eventually caught attempting to pick car keys out of a car owner's backpack while in a shop queue. This method had been successful for Robin on at least 20 previous occasions.

4) Until their recent arrest, Charlie had worked as a till operator at a local charity shop supporting individuals experiencing homelessness. Charlie had been stealing cash amounts varying from £5 to £50 from the tills almost daily over a five-year period. Police revealed that Charlie had stolen several thousand pounds from the charity shop while working there.

5) Jamie was recently made aware of a considerable amount of valuable jewellery in the possession of one of their neighbours. After studying their neighbours' movements, Jamie broke into their apartment while they were out one day and found where the valuables were stored. After noticing that they had been robbed, the neighbour checked their CCTV footage and reported Jamie to the Police.

Studies 3A and 3B: Parole vignettes

Blank spaces within paragraphs were filled in by traits according to trait category (see Study 2.3A Materials).

1) Alex, known by locals in their hometown as having always been a/an_____ individual, has recently begun their first parole hearing at the local courthouse. After being tried and convicted 36 months ago, Alex's behaviour has been reported as_____ during their time in prison so far. In a behavioural report by one of the prison's counsellors, Alex was noted as exhibiting a/an _____

demeanour since their arrival. The counsellor also explains that other prisoners often refer to Alex as being _____ in character.

2) Sam is currently applying for parole after being convicted of a crime just over three years ago. In assessing their suitability for parole, the parole committee gathered reports from prison staff. Guards patrolling the prison grounds noted Sam as being a _____ individual, exhibiting _____ behaviour towards other prisoners for the most part. The prisoner who shares a cell with Sam has referred to them as the most _____ cell-mate they have ever had. In last week's parole hearing, Sam's responses indicated a _____ character. A decision will be made within the next week regarding Sam's parole.

3) Robin is a recently convicted criminal who has agreed to a behavioural assessment in the hopes of it helping with their parole hearing next month. "Having always appeared to be _____ and _____, I feel certain that Robin is _____ in character", noted the psychologist hired to assess Robin. This appears to be similar to what Robin's fellow prisoners say, with one fellow inmate mentioning Robin is known as _____ amongst the other prisoners ever since they were first sentenced 32 months ago.

4) An inmate at a local prison named Jamie has just finished their parole hearing following being sentenced three years ago. According to a prison guard who has got to know Jamie well in that time, Jamie usually appears to be _____, shows a/an _____ approach in helping distribute food and other resources throughout the prison, and is generally known as _____ by the other prisoners. Jamie's cell-mate supports this assessment, calling Jamie a/an _____ individual with which to share their space. A counsellor who has been meeting fortnightly with Jamie concurs with these appraisals.

Appendix B: Supplementary Materials for Chapter 3

B.1. Method: Vignettes in Full

Studies 3.1 and 3.3: Climate Change Context

Initial vignettes for both the **harm responsible** and **harm not responsible** conditions detailing the harm climate change has on residents of Lagos are displayed in Table B1.

Participants in the **harm responsible** condition then read the paragraphs in Table B2., detailing the contributions of UK residents to climate change. Participants in the **harm not responsible** condition instead read the paragraphs in Table B3., emphasizing the contribution of multinational Big Oil companies.

Participants in the **no harm** condition read only four paragraphs in total, which are displayed in Table B4. These contained no mention of climate change or harm and instead detailed aspects of the local culture, economic activities, and lifestyle of residents of Lagos.

All paragraphs were displayed one at a time, and an 8 to 10-second delay was applied before participants could move on to the next.

Table B1. Initial climate change harm vignettes for the harm responsible and harm not responsible conditions

Paragraph 1:	Although climate change is expected to affect all countries, some regions are particularly vulnerable. Lagos in Nigeria is one such region that has already been impacted. The largest city in Nigeria, as well as the most populous on the African continent, the residents of Lagos are considered to be at extreme risk from the harmful effects of climate change.
Paragraph 2:	Lagos currently hosts a thriving economy and reached a Gross Domestic Product (GDP) of USD 136 billion in 2017, about a third of the GDP of the whole of Nigeria. The city is a major transportation hub, with multiple ports and a major international airport. Lagos is also a regional hub for the high-tech industry.
Paragraph 3:	As climate change progresses, residents of Lagos have experienced increasingly more hot days and droughts, as well as higher humidity. Rising temperatures have enhanced the capacity of Anopheles mosquitoes to spread malaria and it is thought that the prevalence of other diseases has also increased. Ailments like measles, meningitis, heat rashes, dehydration and respiratory problems have all been linked to severe heat and high relative humidity.
Paragraph 4:	Climate change has already impacted the city's economy. Lagos is especially vulnerable because it is located on the Gulf of Guinea. Rising sea levels are causing coastal erosion and contaminating drinking water sources. This further harms local agriculture and has damaged the country's important fishing industry.
Paragraph 5:	The resulting job losses have caused lower levels of income and a poorer standard of living for those whose livelihoods depend on such industries. Experts fear the extent of the damage that climate change has on Lagos, with displacement and loss of life already being experienced by some of its residents.

Table B2. Harm responsible vignettes in the climate change context

Paragraph 1:	The UK is historically one of the world's greatest contributors to global warming. This is because of our long history of generating greenhouse gas emissions which began during the Industrial Revolution in the eighteenth and nineteenth centuries.
Paragraph 2:	Richer countries like the UK are responsible for far more carbon pollution than the world's poorer countries. Today, the average carbon footprint per person in the UK is at around 12.7 metric tonnes of carbon dioxide emissions per year, nearly three times more than the worldwide average.
Paragraph 3:	Considering the significant impact that residents of high-polluting countries such as the UK have, global average temperatures could increase by 4°C by the end of the century. Despite recent efforts to reduce carbon emissions in the UK, there is still much room for improvement and we are all collectively responsible.
Paragraph 4:	The average UK resident contributes a substantial amount of support for industries that are spearheading global warming. With high rates of meat and dairy consumption as well as the excessive prevalence of fast fashion and single-use products, the consumer habits of UK residents are typically unsustainable. Moreover, 61% of surface transport emissions in the UK are from private cars and UK residents also create high CO2 emissions from international flights.

Table B3. Harm not responsible vignettes in the climate change context

Paragraph 1:	A recent report found that, since 1988, just 100 international companies have been collectively responsible for 71% of global greenhouse gas emissions. Further, just 25 corporations and state-owned entities were responsible for more than half of global industrial emissions in that same period.
Paragraph 2:	The Carbon Disclosure Project has recently presented data that "pinpoints how a relatively small set of fossil fuel producers may hold the key to systemic change on carbon emissions". Considering the significant impact that high polluting companies have, global average temperatures could increase by 4°C by the end of the century.
Paragraph 3:	Despite recent calls to reduce carbon emissions globally, there is still much room for improvement and companies with the highest global greenhouse gas emissions must take responsibility. Most of these are coal and oil-producing companies and include ExxonMobil, Shell, Chevron, Gazprom, and the Saudi Arabian Oil Company. Such companies emit large amounts of greenhouse gases at each stage of their supply chain. Locating fossil fuel reserves often involves offshore seismic shocks and seabed drilling.
Paragraph 4:	The processing and transportation of fossil fuels also generate substantial levels of emissions. The scale of historical emissions associated with these fossil fuel producers is large enough to have already contributed significantly to climate change, according to the report.

Table B4. No harm vignettes in the climate change context

Paragraph 1:	Many countries across the world have seen considerable economic development in recent decades. Lagos in Nigeria is one such region. The largest city in Nigeria, as well as the most densely populated on the African continent, the residents of Lagos are now thought to number over 10 million people.
Paragraph 2:	Lagos is often considered Africa's foremost urban centre and hub of regional, national, and global socio-economic and political activities. Lagos currently hosts a thriving economy and reached a Gross Domestic Product (GDP) of USD 136 billion in 2017, about a third of the GDP of the whole of Nigeria. The city is a major transportation hub, with multiple ports and a major international airport. Lagos is also a regional hub for high-tech industry.
Paragraph 3:	The residents of Lagos experience a tropical climate that tends to be hot all year round. The dry season is from November to March and the rainy season is from April to October. The hottest month tends to be March, with average temperatures around 30°C, and the coldest month tends to be August, with average temperatures around 26°C. Humidity in Lagos is generally high all year round, with an annual average relative humidity of 84.7%, ranging from 80% in March to 88% in June.
Paragraph 4:	Given that it is located on the Gulf of Guinea, local fishing and agricultural industries are central to the economy of the city and surrounding region, providing many jobs for local residents. The food and cultural scene also provide local jobs. Recreational activities take place along the Atlantic coastline of Lagos, where there are a number of beaches, including Elegushi Beach and Alpha Beach.

Study 2: Fast Fashion Context

Initial vignettes for both the **harm responsible** and **harm not responsible** conditions detailing the harm experienced by textile workers in fast fashion supply chains are displayed in Table B5.

Participants in the **harm responsible** condition moved on to read the paragraphs in Table B6, detailing the high rates of fast fashion consumption by UK residents, emphasising how this is complacent in the harm experienced by textile workers, allowing it to continue. Participants in the **harm not responsible** condition instead read the paragraphs in Table B7, emphasizing how fast fashion companies' profit-driven CEOs and boards of directors are responsible for the harm experienced by textile workers in supply chains.

Participants in the **no harm** condition read only four paragraphs displayed in Table B8. These contained no mention of the exploitation or harm experienced by textile workers and instead described the history and prominence of fast fashion globally.

All paragraphs were displayed one at a time, and an 8 to 10-second delay was applied before participants could move on to the next.

Table B5. Initial fast fashion harm vignettes for the harm responsible and harm not responsible conditions

Paragraph 1:	Fast fashion refers to the production of high volumes of clothing throughout the year. The fast fashion business model relies on a continuous demand for new clothes, offering cheap garments and ever-changing new ranges. The fast fashion industry allows extreme harm to be inflicted on the textile workers who work along the supply chains.
Paragraph 2:	So-called 'sweatshops' are particularly common in South and Southeast Asia. These factories manufacture products for the world's largest clothing brands. The workers in these factories are not typically offered legal protections or fair wages and basic human rights may be neglected. Moreover, gender-based violence, child labour and slavery have been documented within supply chains for fast fashion companies.
Paragraph 3:	Workers for fast fashion supply chains may live in dire socioeconomic conditions, struggling to afford necessities such as food and electricity, even when being expected to work an illegal number of hours. Physical and psychological injuries are widely documented, with no compensation schemes in place. It is estimated that 27 million people working in the fast fashion industry, from cotton farming to sweatshops, suffer work-related illnesses.
Paragraph 4:	Exposure to chemicals in farming and textile processing can cause skin and respiratory conditions. Hearing damage, repetitive strain injury, musculoskeletal injury, eye strain and lung disease from lint inhalation are all widespread amongst textile workers today. Furthermore, instances such as the collapse of Rana Plaza in Bangladesh which killed over 1,100 textile workers highlight the vulnerability of these workers who have no other choice but to work in unsafe, harmful conditions.

Table B6. Harm responsible vignettes in the fast fashion context

Paragraph 1:	One might wonder how fast fashion continues to dominate the UK fashion trade despite harmful practices being well-documented in its supply chain. As supplies aim to meet demand, it is the consumer that supports and maintains the fast fashion industry. The contribution of UK consumers is enormous. Despite exploitative practices being widely reported in the British media, most British people continue to buy fast fashion.
Paragraph 2:	UK consumers tend to buy clothes for the short term. It has been reported that a large number of people in the UK consider a garment worn only for a couple of months to be 'old'. The cycle of frequent purchasing, wearing, and disposing of clothes and other garments supports the high demand for the fast fashion model. This in turn enables the exploitation, injury, and deaths of textile workers to continue.
Paragraph 3:	A cross-party report by the House of Commons noted that UK residents buy more clothes per person than any other European country. These purchasing patterns perpetuate the continued exploitation of labourers in fast fashion supply chains.

Table B7. Harm not responsible vignettes in the fast fashion context

Paragraph 1:	One might wonder how fast fashion continues to dominate the global fashion trade despite such harmful practices within in its supply chain. Like many global markets, it is the CEOs and directors of the companies involved that support and maintain the fast fashion industry. The contribution of each company is enormous, and the exploitation of their workers is often completely hidden from consumers.
Paragraph 2:	Boards of directors for fast fashion companies carefully market their products to normalise cheap, low-quality clothes. UK consumers are often unaware of the problems and have little option but to buy at least some of their clothes from these chains. The prioritization of profits by CEOs of multinational fast fashion companies allows for the exploitation, injury, and deaths of textile workers to continue.
Paragraph 3:	Multinational companies continue to push onwards, hiding workers' conditions from UK consumers, and increasing profits by each year. These practices perpetuate the continued exploitation of labourers in fast fashion supply chains.

Table B8. No harm vignettes in the fast fashion context

Paragraph 1:	Fast fashion refers to the production of high volumes of clothing throughout the year. The fast fashion business model relies on an endless demand for new clothes, offering inexpensive garments and ever-changing new ranges. Fast fashion brands are able to mass produce clothes at a low cost, meaning that consumers can update their wardrobes quickly and affordably.
Paragraph 2:	The accessibility of inexpensive new clothes is relatively new. Up until the about 1950s, a large amount of clothing was still made in the home, or in local factories and couture houses for families that could afford it. In the 1960s young people embraced cheaply made clothing to follow new trends. Fashion brands had to find ways to keep up with this increasing demand for affordable clothing, leading them to open textile mills across the world.
Paragraph 3:	Today, fast fashion is boosted even further with the rise of social media and celebrity culture. When a celebrity posts a photo wearing a new outfit, fast fashion brands rush to be the first to make a similar style accessible to the general public.

B.2. Results: Main Effects and Two-Way Interactions

Study 3.1: Testing for Infrahumanization of Outgroup Members Harmed by Climate Change

Emotion Ratings.

Main Effects. There was a significant main effect of emotion humanness, $F(1, 249) = 307.608$, $p < .001$, $\eta_p^2 = .55$. Residents of Lagos were seen as typically

experiencing emotions shared with other animals (M = 57.4, SE = 0.64) to a greater extent than uniquely human emotions (M = 48.6, SE = 0.72), $p < .001$. There was also a significant main effect of emotion valence, $F(1, 249) = 17.691$, $p < .001$, $\eta_p^2 = .07$, whereby participants rated residents of Lagos as experiencing positive emotions (M = 55.8, SE = 0.93) to a greater extent than negative emotions (M = 50.2, SE = 0.92), $p < .001$. There was no significant main effect of harm condition on emotion typicality ratings, $F(2, 249) = 1.63$, $p = .198$, $\eta_p^2 = .01$.

Condition*Humanness. A significant two-way interaction between harm condition and emotion humanness was found $F(2, 249) = 20.27$, $p < .001$, $\eta_p^2 = .14$. However, the pattern of results did not follow that predicted by infrahumanization theory. Residents of Lagos were not seen as experiencing uniquely human emotions any differently between the harm responsible condition (M = 48.7, SE = 1.24) and the harm not responsible condition (M = 47.8, SE = 1.24), $p = 1$. Uniquely human emotion ratings in the no harm condition (M = 49.4, SE = 1.24) did not differ from those in either of the harm conditions, both ps = 1. No difference in ratings of emotions shared with animals was found between the harm responsible condition (M = 60.0 SE = 1.11) and the harm not responsible condition (M = 58.3, SE = 1.11), $p = .842$. Compared to participants in the no harm condition (M = 53.7, SE = 1.11), emotions shared with other animals were rated significantly higher by those in both the harm responsible condition ($p < .001$) and the harm not responsible condition ($p = .011$).

Condition*Valence. A significant two-way interaction was found between harm condition and emotion valence, $F(2, 246) = 57.714$, $p < .001$, $\eta_p^2 = .32$. No difference was found in ratings of positive emotions between participants in the harm responsible condition (M = 53.14, SE = 1.61) and those in the harm not responsible condition (M = 49.80, SE = 1.61), $p = .433$. Participants in the no harm condition (M = 64.53, SE = 1.61) rated positive human emotions higher than participants in either of the harm conditions, both ps < .001.

No difference in ratings of negative emotions was found between the harm responsible condition (M = 55.6, SE = 1.59) and the harm not responsible condition (M = 56.3, SE = 1.59),

p = 1. Participants in both harm conditions rated negative emotions higher than those in the no harm condition (M = 38.6, SE = 1.59), both ps < .001.

Humanness*Valence. A significant two-way interaction between emotion humanness and emotion valence was found $F(1, 249) = 275.124$, $p < .001$, $\eta_p^2 = .53$. No difference was found in how participants perceived residents of Lagos as experiencing positive emotions when they were uniquely human (M = 56.4, SE = 0.93) compared to when they were shared with other animals (M = 55.2, SE = 1.05), p = .074. In contrast, participants perceived residents of Lagos as experiencing negative emotions to a greater extent when they were shared with other animals (M = 59.5, SE = 1.07) than when they were uniquely human (M = 40.9, SE = 0.96), p < .001.

Study 3.2: Testing for Infrahumanization of Outgroup Members Harmed by Fast Fashion

Emotion Ratings.

Main Effects. A significant main effect of emotion humanness was found, $F(1, 249) = 112.095$, $p < .001$, $\eta_p^2 = .31$. Emotions shared with other animals (M = 47.0, SE = 0.63) tended to be rated higher than uniquely human emotions (M = 41.1, SE = 0.77), p < .001. A significant main effect of emotion valence was also found, $F(1, 249) = 348.403$, $p < .001$, $\eta_p^2 = .58$. Participants tended to rate textile workers as experiencing negative emotions (M = 58.0, SE = 0.90) to a greater extent than positive emotions (M = 30.2, SE = 1.06), p < .001. No main effect of harm condition was found, $F(2, 249) = 0.63$, p = .533, $\eta_p^2 = .01$.

Condition*Humanness. A significant two-way interaction between harm condition and emotion humanness was found, $F(2, 249) = 5.647$, p = .004, $\eta_p^2 = .04$. Contrary to the previous research on a infrahumanization following harm, there was no significant difference in the extent to which participants attributed uniquely human emotions to textile workers between the harm responsible condition (M = 41.4, SE = 1.33) and the harm not responsible condition (M = 39.2, SE = 1.33), p = .74. Unlike in our Study 3.1 results, the extent to which participants attributed uniquely human emotions to textile workers did not differ between the

no harm condition (M = 42.8, SE = 1.33) and either the harm responsible condition (p = 1) or the harm not responsible condition (p = .170). Similarly, the extent to which participants attributed emotions shared with other animals to textile workers did not differ between the harm responsible condition (M = 48.0, SE = 1.08) and the harm not responsible condition (M = 46.9, SE = 1.08), p = 1. Participants in the no harm condition (M = 46.1, SE = 1.08) did not rate shared emotions any differently than those in either the harm responsible condition (p = .627) or the harm not responsible condition (p = 1).

Condition*Valence. A significant two-way interaction was also found between harm and emotion valence, $F(2, 246) = 29.730$, $p < .001$, $\eta_p^2 = .19$. No significant difference was found in the extent to which participants perceived textile workers as typically experiencing positive emotions between the harm responsible condition (M = 27.7, SE = 1.84) and the harm not responsible condition (M = 24.2, SE = 1.84), p = .524. Participants in no harm condition (M = 38.6, SE = 1.84) rated textile workers as typically experiencing positive emotions to a greater extent than those in either of the two harm conditions, both ps < .001. No significant difference was found in the extent to which participants believed textile workers typically experience negative emotions between the harm responsible condition (M = 61.7, SE = 1.55) and the harm not responsible condition (M = 61.9, SE = 1.55), p = 1. Participants in each of the two harm conditions rated textile workers as experiencing negative emotions to a greater extent than those in the no harm condition (M = 50.3, SE = 1.55), both ps < .001.

***Humanness*Valence*.** A significant two-way interaction between emotion humanness and emotion valence was found $F(1, 249) = 563.328$, $p < .001$, $\eta_p^2 = .68$. Participants perceived textile workers in fast fashion supply chains as experiencing positive uniquely human emotions (M = 34.6, SE = 1.15) to a greater extent than positive emotions shared with other species (M = 25.7, SE = 1.10), $p < .001$. Participants also perceived textile workers as typically experiencing negative emotions shared with other animals (M = 68.3, SE = 1.02) to a greater extent than negative uniquely human emotions (M = 47.6, SE = 1.00), $p < .001$.

B.3. Comparing Alternative Analyses for Study 3.3

Alternative Analyses When Testing for Infrahumanization Following Harm

Alternative results when deciding which highly influential emotions were excluded (if any) are displayed in Table B9, and associated main results are displayed in Table B10.

Table B9. Results of regression models when testing for infrahumanization following harm using different analyses

Analysis	Condition	*F*	*df*	*b*	*p*	R^2	*r**
64 emotions in all models	**HR**	4.334	1,62	-.112	.042	.07	-.256
	HNR	3.18	1,62	-.126	.08	.05	-.221
	NH	2.581	1,62	-.107	.113	.04	-.2
Model-by-model	**HR**	5.667	1,61	-.128	.02	.09	-.292
	HNR	7.84	1,59	-.202	.007	.12	-.343
	NH	5.168	1,60	-.15	.027	.08	-.282
59 emotions in all models	**HR**	7.231	1,57	-.156	.009	.11	-.336
	HNR	5.636	1,57	-.18	.021	.09	-.3
	NH	2.878	1,57	-.123	.095	.05	-.219

Notes: HR = Harm responsible condition; *HNR* = Harm not responsible condition; *NH* = No harm condition*; b* = coefficient of the predictor variable (emotion humanness); r* = Pearson correlation coefficient between emotion humanness and outgroup emotion rating. No meaningful difference can be seen between the different analyses. Analysis with 64 emotions in all models included all emotions from our selection in each model. The model-by-model analysis excluded highly influential emotions from each respective model; thus, the models differed in which and how many emotions were included. Analysis with 59 emotions excluded emotions identified as highly influential in any model from all three models (*Bitterness*, *Disenchantment*, *Disillusion*, *Optimism* and *Terror* removed).

Table B10. Comparing the regression coefficients between the different conditions when testing for infrahumanization following harm using different analyses

Analysis	Models being compared	*z*	*p*	*Cohen's q*	Same result using Pearson *r*?	Same result with partial regression?
64 emotions	**HR & HNR**	.082	.934	.02	Yes	Yes
	HR & NH	.028	.978	.01	Yes	Yes
	HNR & NH	.11	.913	.02	Yes	Yes
Model-by-model	**HR & HNR**	.413	.68	.08	Yes	Yes
	HR & NH	.122	.903	.02	Yes	Yes
	HNR & NH	.291	.771	.05	Yes	Yes
59 emotions	**HR & HNR**	.132	.90	.03	Yes	Yes
	HR & NH	.176	.861	.03	Yes	Yes
	HNR & NH	.307	.759	.06	Yes	Yes

Notes: HR = Harm responsible condition; *HNR* = Harm not responsible condition; *NH* = No harm condition. Results of Fisher Z transformations performed using the cocor package in R are displayed. The theoretically relevant comparison is between HR & HNR, while the comparisons between HR & NH, and HNR & NH are not theoretically relevant. Comparing Pearson correlation coefficients between emotion humanness and outgroup emotion ratings, rather than regression coefficients, yielded the same results. Comparing partial regression coefficients while controlling for emotion valence yielded the same results. No meaningful difference can be seen between the different analyses.

Alternative Analyses when Testing for Negative Affect when Harmed.

Alternative results when deciding which highly influential emotions were excluded (if any) are displayed in Table B11, and associated main results are displayed in Table B12.

Table B11. Results of regression models using different analyses when testing for negative affect when harmed

Analysis	Condition	*F*	*df*	*b*	*p*	R^2	*r**
64 emotions in all models	**HR**	25.41	1,62	-.161	.001	.29	-.54
	HNR	67.94	1,62	-.277	.001	.52	-.723
	NH	236.7	1,62	.314	.001	.79	.891
Model-by-model	**HR**	52.51	1,59	-.198	.001	.47	-.686
	HNR	116.3	1,59	-.314	.001	.66	-.815
	NH	283.9	1,60	.329	.001	.83	.91
59 emotions in all models	**HR**	44.95	1,57	-.187	.001	.44	-.664
	HNR	104.9	1,57	-.307	.001	.65	-.805
	NH	255	1,57	.318	.001	.82	.904

Notes: HR = Harm responsible condition; *HNR* = Harm not responsible condition; *NH* = No harm condition; *b* = coefficient of the predictor variable (emotion valence); r* = Pearson correlation coefficient between emotion humanness and outgroup emotion rating. Analysis with 64 emotions in all models included all emotions from our selection in each model. The model-by-model analysis excluded highly influential emotions from each respective model; thus, the models differed in which emotions were included. Analysis with 59 emotions excluded emotions identified as highly influential in any model from all three models (*Guilt, Love, Shame, Euphoria* and *Grief* removed). No meaningful difference can be seen between the different analyses.

Table B12. Comparing the regression coefficients between the different conditions when testing for negative affect when harmed, using different analyses

Analysis	Models being compared	*z*	*p*	*Cohen's q*	Same result using Pearson r?	Same result with partial regression?
64 emotions	**HR & HNR**	0.678	.498	.12	Yes	Yes
	HR & NH	2.691	.007	.49	Yes	Yes
	HNR & NH	3.368	< .001	.61	Yes	Yes
Model-by-model	**HR & HNR**	0.672	.502	.13	Yes	Yes
	HR & NH	2.934	.003	.54	Yes	Yes
	HNR & NH	3.609	< .001	.67	Yes	Yes
59 emotions	**HR & HNR**	0.677	.498	.13	Yes	Yes
	HR & NH	2.749	.006	.52	Yes	Yes
	HNR & NH	3.426	< .001	.65	Yes	Yes

Notes: Cond 1 = Harm responsible condition; *Cond 2* = Harm not responsible condition; *Cond 3* = No harm condition. Results of Fisher Z transformations performed using the cocor package in R are displayed. The theoretically relevant comparisons are between Conditions 1 & 3, and 2 & 3, whereas the comparison between HR & HNR is not theoretically relevant. Comparing Pearson correlation coefficients between emotion valence and outgroup emotion ratings, rather than regression coefficients, yielded the same results. Comparing partial regression coefficients while controlling for emotion humanness yielded the same results. No meaningful differences can be seen between the different analyses.

References

Alliance Against Antigypsyism. (2017). *Alliance Against Antigypsyism - A Reference Paper*. https://antigypsyism.eu/reference-paper-on-antigypsyism/

Altemeyer, B. (1981). *Right-wing authoritarianism*. https://www.google.co.uk/books/edition/Right_Wing_Authoritarianism/-mFaEAAAQBAJ?hl=en&gbpv=0

Amnesty International. (2021, December 1). *EU: 'Exceptional measures' normalize dehumanization of asylum seekers*. https://www.amnesty.org/en/latest/news/2021/12/eu-exceptional-measures-normalize-dehumanization-of-asylum-seekers/

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. *British Journal of Social Psychology*, *53*(3), 573–584. https://doi.org/10.1111/BJSO.12066

Andrighetto, L., Mari, S., Volpato, C., & Behluli, B. (2012). Reducing Competitive Victimhood in Kosovo: The Role of Extended Contact and Common Ingroup Identity. *Political Psychology*, *33*(4), 513–529. https://doi.org/10.1111/J.1467-9221.2012.00887.X

Azevedo, R. T., De Beukelaer, S., Jones, I. L., Safra, L., & Tsakiris, M. (2021). When the lens is too wide: The political consequences of the visual dehumanization of refugees. *Humanities and Social Sciences Communications 2021 8:1*, *8*(1), 1–16. https://doi.org/10.1057/s41599-021-00786-x

Bailenson, J., Guadagno, R., Aharoni, E., Dimov, A., Beall, A., & Blascovich, J. (2004). Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments. *Proceedings of the 7th Annual International Workshop on PRESENCE*.

Bailey, A. H., LaFrance, M., & Dovidio, J. F. (2018). Is Man the Measure of All Things? A Social Cognitive Account of Androcentrism. *Personality and Social Psychology Review*,

23(4), 307–331. https://doi.org/10.1177/1088868318782848

Bain, P., Park, J., Kwok, C., & Haslam, N. (2009). Attributing human uniqueness and human nature to cultural groups: Distinct forms of subtle dehumanization. *Group Processes and Intergroup Relations*, *12*(6), 789–805. https://doi.org/10.1177/1368430209340415

Bain, P., Vaes, J., Kashima, Y., Haslam, N., & Guan, Y. (2012). Folk conceptions of humanness: Beliefs about distinctive and core human characteristics in Australia, Italy, and China. *Journal of Cross-Cultural Psychology*, *43*(1), 53–58. https://doi.org/10.1177/0022022111419029

Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit Stereotyping in Person Judgment. *Journal of Personality and Social Psychology*, *65*(2), 272–281. https://doi.org/10.1037/0022-3514.65.2.272

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

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of Moral Disengagement in the Exercise of Moral Agency. *Journal of Personality and Social Psychology*, *71*(2). https://doi.org/10.1037/0022-3514.71.2.364

Banton, O., West, K., & Kinney, E. (2020). The surprising politics of anti-immigrant prejudice: How political conservatism moderates the effect of immigrant race and religion on infrahumanization judgements. *British Journal of Social Psychology*, *59*(1), 157–170. https://doi.org/10.1111/BJSO.12337

Bar-Tal, D. (2000). Shared beliefs in a society: Social psychological analysis. In *Society: Social Psychological Analysis BAR*. https://books.google.com/books?hl=en&lr=&id=R-KnGmLKo6UC&oi=fnd&pg=PP1&dq=shared+beliefs+in+society+bar+tal+2000&ots=LNvXydlFWq&sig=83EjnIdaUYmj66yryPoen--seQs

Barber, M., & Davis, R. (2022). Partisanship and the trolley problem: Partisan willingness to sacrifice members of the other party. *Research and Politics*, *9*(4). https://doi.org/10.1177/20531680221137143

Bastian, B. (2019). A dehumanization perspective on dependence in low-satisfaction (abusive) relationships. *Journal of Social and Personal Relationships*, *36*(5), 1421–1440. https://doi.org/10.1177/0265407519835978

Bastian, B., Denson, T. F., & Haslam, N. (2013). The Roles of Dehumanization and Moral Outrage in Retributive Justice. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061842

Bastian, B., & Haslam, N. (2010). Excluded from humanity: The dehumanizing effects of social ostracism. *Journal of Experimental Social Psychology*, *46*(1), 107–113. https://doi.org/10.1016/j.jesp.2009.06.022

Batson, C. D., & Ahmad, N. Y. (2009). Using Empathy to Improve Intergroup Attitudes and Relations. *Social Issues and Policy Review*, *3*(1), 141–177. https://doi.org/10.1111/J.1751-2409.2009.01013.X

Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: An Interpersonal Approach. *Psychological Bulletin*, *115*(2). https://doi.org/10.1037/0033-2909.115.2.243

Bernard, P., Cogoni, C., & Carnaghi, A. (2020). The Sexualization–Objectification Link: Sexualization Affects the Way People See and Feel Toward Others. *Current Directions in Psychological Science*, *29*(2), 134–139. https://doi.org/10.1177/0963721419898187

Bernard, P., Gervais, S. J., Allen, J., Delmée, A., & Klein, O. (2015). From Sex Objects to Human Beings: Masking Sexual Body Parts and Humanization as Moderators to Women's Objectification. *Psychology of Women Quarterly*, *39*(4), 432–446. https://doi.org/10.1177/0361684315580125

Bernard, P., Loughnan, S., Marchal, C., Godart, A., & Klein, O. (2015). The Exonerating Effect of Sexual Objectification: Sexual Objectification Decreases Rapist Blame in a Stranger Rape Context. *Sex Roles 2015 72:11*, *72*(11), 499–508. https://doi.org/10.1007/S11199-015-0482-0

Billig, M. (2012). The notion of 'prejudice': Some rhetorical and ideological aspects. In *Beyond Prejudice: Extending the Social Psychology of Conflict, Inequality and Social Change* (pp. 139–157). Cambridge University Press.

https://doi.org/10.1017/CBO9781139022736.009

Bloom, P. (2017, November 20). The Root of All Cruelty? Perpetrators of violence, we're told, dehumanize their victims. The truth is worse. *The New Yorker*. https://www.newyorker.com/magazine/2017/11/27/the-root-of-all-cruelty

Bloom, P. (2022). If everything is dehumanization, then nothing is. *Trends in Cognitive Sciences*, *26*(7), 539. https://doi.org/10.1016/J.TICS.2022.03.001

Boucher, D. (2019). Reclaiming history: dehumanization and the failure of decolonization. *International Journal of Social Economics*, *46*(11), 1250–1263. https://doi.org/10.1108/IJSE-03-2019-0151

Boudjemadi, V., Demoulin, S., & Bastart, J. (2017). Animalistic dehumanization of older people by younger ones: Variations of humanness perceptions as a function of a target's age. *Psychology and Aging*, *32*(3), 293–306. https://doi.org/10.1037/PAG0000161

Brennan, R. A., Enock, F. E., & Over, H. (2024). *The attribution of socially undesirable traits better explains increased harm towards criminals than trait-based dehumanization [Manuscript submitted for publication]*.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues*, *55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Bruneau, E., Kteily, N., & Laustsen, L. (2018). The unique effects of blatant dehumanization on attitudes and behavior towards Muslim refugees during the European 'refugee crisis' across four countries. *European Journal of Social Psychology*, *48*(5), 645–662. https://doi.org/10.1002/EJSP.2357

Bruneau, E., Szekeres, H., Kteily, N., Tropp, L. R., & Kende, A. (2020). Beyond dislike: Blatant dehumanization predicts teacher discrimination. *Group Processes and Intergroup Relations*, *23*(4), 560–577. https://doi.org/10.1177/1368430219845462

Bunce, C., Eggleston, A., Brennan, R. A., & Over, H. (2024). *To what extent is research on infrahumanization confounded by intergroup preference? [Manuscript submitted for publication]*. Department of Psychology, University of York.

Capozza, D., Di Bernardo, G. A., Falvo, R., Vianello, R., & Calò, L. (2016). Individuals with intellectual and developmental disabilities: Do educators assign them a fully human status? *Journal of Applied Social Psychology*, *46*(9), 497–509. https://doi.org/10.1111/JASP.12377

Cassese, E. C. (2020). Dehumanization of the Opposition in Political Campaigns. *Social Science Quarterly*, *101*(1), 107–120. https://doi.org/10.1111/SSQU.12745

Castano, E., & Giner-Sorolla, R. (2006). Not quite human: Infrahumanization in response to collective responsibility for intergroup killing. *Journal of Personality and Social Psychology*, *90*(5), 804–818. https://doi.org/10.1037/0022-3514.90.5.804

Čehajić, S., Brown, R., & González, R. (2009). What do I Care? Perceived Ingroup Responsibility and Dehumanization as Predictors of Empathy Felt for the Victim Group: *Group Processes & Intergroup Relations*, *12*(6), 715–729. https://doi.org/10.1177/1368430209347727

Charalambides, N. (2021). We recently went viral on TikTok - here's what we learned. *Prolific.Com*. https://www.prolific.com/blog/we-recently-went-viral-on-tiktok-heres-what-we-learned

Chen-Xia, X. J., Betancor, V., Rodríguez-Gómez, L., & Rodríguez-Pérez, A. (2023). Cultural variations in perceptions and reactions to social norm transgressions: a comparative study. *Frontiers in Psychology*, *14*, 1243955. https://doi.org/10.3389/fpsyg.2023.1243955

Cikara, M., Bruneau, E. G., & Saxe, R. R. (2017). Us and Them: Intergroup Failures of Empathy. *Current Directions in Psychological Science*, *20*(3), 149–153. https://doi.org/10.1177/0963721411408713

Civitillo, S., Ialuna, F., Lieck, D., & Jugert, P. (2022). Do infrahumanization and affective prejudice drive teacher discrimination against Romani students? A conceptual replication of Bruneau et al. (2020) in Germany. *Peace and Conflict: Journal of Peace Psychology*, *28*(3), 340–344. https://doi.org/https://doi.org/10.1037/pac0000609

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.

https://doi.org/10.4324/9780203771587

Cojocaru, M. (2022). Language Matters: Representation of Gypsies and Travellers in British media - Minority Rights Group. *Minority Rights Group*. https://minorityrights.org/language-matters-representation-of-gypsies-and-travellers-in-british-media/

Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the Thin Blue Line: Police Officers and Racial Bias in the Decision to Shoot. *Journal of Personality and Social Psychology*, *92*(6), 1006–1023. https://doi.org/10.1037/0022-3514.92.6.1006

Cortes, B. P., Demoulin, S., Rodriguez, R. T., Rodriguez, A. P., & Leyens, J. P. (2005). Infrahumanization or familiarity? Attribution of uniquely human emotions to the self, the ingroup, and the outgroup. *Personality & Social Psychology Bulletin*, *31*(2), 243–253. https://doi.org/10.1177/0146167204271421

Costello, K., & Hodson, G. (2009). Exploring the roots of dehumanization: The role of animal—human similarity in promoting immigrant humanization. *Group Processes & Intergroup Relations*, *13*(1), 3–22. https://doi.org/10.1177/1368430209347725

Costello, K., & Hodson, G. (2011). Social dominance-based threat reactions to immigrants in need of assistance. *European Journal of Social Psychology*, *41*(2), 220–231. https://doi.org/10.1002/EJSP.769

Crawford, J. T. (2012). The ideologically objectionable premise model: Predicting biased political judgments on the left and right. *Journal of Experimental Social Psychology*, *48*(1), 138–151. https://doi.org/10.1016/J.JESP.2011.10.004

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*(1), 107–118. https://doi.org/10.1177/1368430207071344

Dalsklev, M., & Kunst, J. R. (2015). The effect of disgust-eliciting media portrayals on outgroup dehumanization and support of deportation in a Norwegian sample. *International Journal of Intercultural Relations*, *47*, 28–40.

Davis, A. Y. (1981). *Women, Race & Class*. Random House.

De Dreu, C. K. W., Greer, L. L., Van Kleef, G. A., Shalvi, S., & Handgraaf, M. J. J. (2011). Oxytocin promotes human ethnocentrism. *Proceedings of the National Academy of Sciences*, *108*(4), 1262–1266. https://doi.org/10.1073/PNAS.1015316108

de Ruiter, A. (2021). To be or not to be human: Resolving the paradox of dehumanisation. *European Journal of Political Theory*, *0*(0), 1–23. https://doi.org/10.1177/1474885120984605

Decker, K. A., & Lord, C. G. (2023). Self-Polarization: Lionizing Those Who Agree and Demonizing Those Who Disagree. *Basic and Applied Social Psychology*, *45*(5), 125–137. https://doi.org/10.1080/01973533.2023.2234534

Delbosc, A., Naznin, F., Haslam, N., & Haworth, N. (2019). Dehumanization of cyclists predicts self-reported aggressive behaviour toward them: A pilot study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, 681–689.

Delgado, N., Rodríguez-Pérez, A., Vaes, J., Leyens, J. P., & Betancor, V. (2009). Priming Effects of Violence on Infrahumanization. *Group Processes & Intergroup Relations*. https://doi.org/10.1177/1368430209344607

Delgado Rodríguez, N., Rodríguez-Pérez, A., Vaes, J., Rodríguez, V. B., & Leyens, J. P. (2012). Contextual Variations of Infrahumanization: The Role of Physical Context and Territoriality. *Basic and Applied Social Psychology*, *34*(5). https://doi.org/10.1080/01973533.2012.712020

DeLuca-McLean, D., & Castano, E. (2009). Infra-Humanization of Ethnic Minorities: The Moderating Role of Ideology. *Basic and Applied Social Psychology*, *31*(2), 102–108. https://doi.org/10.1080/01973530902880258

Demoulin, S., Cortes, B. P., Viki, T. G., Rodriguez, A. P., Rodriguez, R. T., Paladino, M. P., & Leyens, J. P. (2009). The role of in-group identification in infra-humanization. *International Journal of Psychology*, *44*(1), 4–11. https://doi.org/10.1080/00207590802057654

Demoulin, S., Leyens, J.-P., Rodríguez-Torres, R., Rodríguez-Pérez, A., Paladino, P. M., &

Fiske, S. T. (2005). Motivation to support a desired conclusion versus motivation to avoid an undesirable conclusion: The case of infra-humanization. *International Journal of Psychology*, *40*(6), 416–428. https://doi.org/10.1080/00207590500184495

Demoulin, S., Leyens, J. P., Paladino, M. P., Rodriguez-Torres, R., Rodriguez-Perez, A., & Dovidio, J. F. (2004). Dimensions of “uniquely” and “non-uniquely” human emotions. *Cognition and Emotion*, *18*(1), 71–96. https://doi.org/10.1080/02699930244000444

Deska, J. C., Lloyd, E. P., & Hugenberg, K. (2018). Facing humanness: Facial width-to-height ratio predicts ascriptions of humanity. *Journal of Personality and Social Psychology*, *114*(1), 75–94. https://doi.org/10.1037/PSPI0000110

Devine, P. G. (1989). Stereotypes and Prejudice: Their Automatic and Controlled Components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLoS ONE*, *10*(4). https://doi.org/10.1371/JOURNAL.PONE.0121945

Diener, E., & Biswas-Diener, R. (2024). The replication crisis in psychology. In R. Biswas-Diener & E. Diener (Eds.), *Noba textbook series: Psychology.* DEF publishers. http://noba.to/q4cvydeh

Diniz, E., Castro, P., Bousfield, A., & Figueira Bernardes, S. (2020). Classism and dehumanization in chronic pain: A qualitative study of nurses’ inferences about women of different socio-economic status. *British Journal of Health Psychology*, *25*(1), 152–170. https://doi.org/10.1111/BJHP.12399

Dovidio, John F., Gaertner, S. L., & Kawakami, K. (2010). Racism. In J.F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 312–327). Sage. https://uk.sagepub.com/en-gb/eur/the-sage-handbook-of-prejudice-stereotyping-and-discrimination/book230685

Duckitt, J. (2001). A dual-process cognitive-motivational theory of ideology and prejudice. *Advances in Experimental Social Psychology*, *33*, 41–113.

https://doi.org/10.1016/S0065-2601(01)80004-6

Duckitt, J., & Sibley, C. G. (2007). Right wing authoritarianism, social dominance orientation and the dimensions of generalized prejudice. *European Journal of Personality*, *21*(2), 113–130. https://doi.org/10.1002/PER.614

Durante, F., Fiske, S. T., Kervyn, N., Cuddy, A. J. C., Akande, A. D., Adetoun, B. E., Adewuyi, M. F., Tserere, M. M., Ramiah, A. Al, Mastor, K. A., Barlow, F. K., Bonn, G., Tafarodi, R. W., Bosak, J., Cairns, E., Doherty, C., Capozza, D., Chandran, A., Chryssochoou, X., … Storari, C. C. (2013). Nations' income inequality predicts ambivalence in stereotype content: How societies mind the gap. *British Journal of Social Psychology*, *52*(4), 726–746. https://doi.org/10.1111/BJSO.12005

Durrheim, K., Quayle, M., & Dixon, J. (2016). The Struggle for the Nature of "Prejudice": "Prejudice" Expression as Identity Performance. *Political Psychology*, *37*(1), 17–35. https://doi.org/10.1111/POPS.12310

Eggleston, A., Cook, R., & Over, H. (2023). Are upside-down faces perceived as "less human"? *Journal of Experimental Psychology. Human Perception and Performance*, *49*(12), 1503–1517. https://doi.org/10.1037/XHP0001167

Ekman, P. (2008). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ellemers, N. (2018). Gender Stereotypes. *Annual Review of Psychology*, *69*, 275–298.

Enock, F. E., Flavell, J. C., Tipper, S. P., & Over, H. (2021a). No convincing evidence outgroups are denied uniquely human characteristics: Distinguishing intergroup preference from trait-based dehumanization. *Cognition*, *212*, 1–73. https://doi.org/10.1016/j.cognition.2021.104682

Enock, F. E., & Over, H. (2022). Reduced helping intentions are better explained by the attribution of antisocial emotions than by 'infrahumanization.' *Scientific Reports*, *12*(1), 1–14. https://doi.org/10.1038/s41598-022-10460-0

Enock, F. E., & Over, H. (2023a). Animalistic slurs increase harm by changing perceptions of social desirability. *Royal Society Open Science*, *10*(7).

https://doi.org/10.1098/RSOS.230203

Enock, F. E., & Over, H. (2023b). Exploring the 'Paradox of Dehumanization': References to Human Specific Traits and Mental States are Frequent in Dehumanizing Propaganda. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.4514034

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

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Esses, V. M., Veenvliet, S., Hodson, G., & Mihic, L. (2008). Justice, morality, and the dehumanization of refugees. *Social Justice Research*, *21*(1), 4–25. https://doi.org/10.1007/S11211-007-0058-4

European Commission Against Racism and Intolerance. (2020). *ECRI GENERAL POLICY RECOMMENDATION NO.13: On combatting antigypsyism and discrimination against Roma*. https://rm.coe.int/ecri-general-policy-recommendation-no-13-on-combating-anti-gypsyism-an/16808b5aee

Farr, R. M. (1996). *The roots of modern social psychology, 1872–1954: Vol. xvii*. Blackwell publishing. https://psycnet.apa.org/record/1997-08228-000

Fincher, K. M., & Tetlock, P. E. (2016). Perceptual dehumanization of faces is activated by norm violations and facilitates norm enforcement. *Journal of Experimental Psychology: General*, *145*(2), 131–146. https://doi.org/10.1037/XGE0000132

Firinci, Y. (2018). The Possible Interconnectedness of Religious Populism, Stigmatization, Dehumanization and Hate Crimes and an Evaluation on Muslims in Russia. *Journal of Media and Religion Studies*, *1*(2), 255–274. https://dergipark.org.tr/en/pub/mediad/issue/41966/477028

Fisher, M. L. (2012). Understanding the darker side of human nature. *Journal of Social, Evolutionary, and Cultural Psychology*, *6*(4). https://doi.org/10.1037/h0099233

Fisher, R. A. (1992). Statistical Methods for Research Workers. In S. Kotz (Ed.), *Breakthroughs in Statistics*. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-4380-9_6

Fiske, S. T. (2018). Stereotype Content: Warmth and Competence Endure. *Current Directions in Psychological Science*, *27*(2), 67–73. https://doi.org/10.1177/0963721417738825

Fiske, S. T., Cuddy, A. J. C., 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*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fredrickson, B. L., & Roberts, T.-A. (2016). Objectification Theory: Toward Understanding Women's Lived Experiences and Mental Health Risks: *Psychology of Women Quarterly*, *21*(2), 173–206. https://doi.org/10.1111/J.1471-6402.1997.TB00108.X

Freund, R. J., Wilson, W. J., & Mohr, D. L. (2010). Statistical Methods, Third Edition. *Statistical Methods, Third Edition*, 1–796. https://doi.org/10.1016/C2009-0-20216-9

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 358, Issue 1431). https://doi.org/10.1098/rstb.2002.1218

Gaines, S. O. (2012). Stereotyping, prejudice and discrimination revisited: from William James to W. E. B. Du Bois. In J. Dixon & M. Levine (Eds.), *Beyond Prejudice: Extending the Social Psychology of Conflict, Inequality and Social Change* (pp. 105–119). Cambridge University Press.

Gaines, S. O., & Reed, E. S. (1995). Prejudice: From Allport to DuBois. *American Psychologist*, *50*(2), 96–103. https://doi.org/10.1037/0003-066X.50.2.96

Gaunt, R. (2009). Superordinate categorization as a moderator of mutual infrahumanization. *Group Processes and Intergroup Relations*, *12*(6), 731–746. https://doi.org/10.1177/1368430209343297

Gaunt, R. (2013). Ambivalent sexism and the attribution of emotions to men and women.

Revue Internationale de Psychologie Sociale, *26*(2), 29–54. https://doi.org/10.1016/S0022-1031(03)00031-3

Ghani, M. T. (2020). *Inside the Ugliness of the Fast Fashion Industry*. Maverick Youth. https://medium.com/maverickyouth/inside-the-ugliness-of-the-fast-fashion-industry-ac40f6a24e01

Gilhooly, K. J., Fioratou, E., Anthony, S. H., & Wynn, V. (2007). Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology*, *98*(4), 611–625. https://doi.org/10.1111/J.2044-8295.2007.TB00467.X

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*(2), 292–306. https://doi.org/10.1037/0022-3514.94.2.292

Gómez-Martínez, C., & de la Villa Moral-Jiménez, M. (2018). Dehumanization and Islamophobia: attitudes towards the Syrian refugee crisis / Deshumanización e islamofobia: actitudes ante la crisis de los refugiados sirios. *Revista de Psicologia Social*, *33*(2), 215–239. https://doi.org/10.1080/02134748.2018.1435218

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of Mind Perception. *Science*, *315*(5812), 619. https://doi.org/10.1126/SCIENCE.1134475

Gray, K., & Wegner, D. M. (2009). Moral Typecasting: Divergent Perceptions of Moral Agents and Moral Patients. *Journal of Personality and Social Psychology*, *96*(3), 505–520. https://doi.org/10.1037/A0013748

Guilford, J. P. (1967). *Nature of Human Intelligence*. McGraw-Hill. https://psycnet.apa.org/record/1967-35015-000

Hamilton, D. L. (1981). Cognitive Processes in Stereotyping and Intergroup Behavior. In *Cognitive Processes in Stereotyping and Intergroup Behavior* (1st ed.). Psychology Press. https://doi.org/10.4324/9781315668758

Harding, J., Proshansky, H., Kutner, B., & Chein, I. (1969). Prejudice and ethnic relations. In

G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 5, pp. 1–76). Addison-Wesley. http://scholar.google.com/scholar_lookup?title=Prejudice and ethnic relations&author=J. Harding&author=H. Proshansky&author=B. Kutner&author=I. Chein&pages=1-76&publication_year=1969

Harris, L. T., & Fiske, S. T. (2006a). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, *17*(10), 847–853. https://doi.org/10.1111/J.1467-9280.2006.01793.X

Harris, L. T., & Fiske, S. T. (2006b). Dehumanizing the Lowest of the Low. *Psychological Science*, *17*(10), 847–853. https://doi.org/10.1111/j.1467-9280.2006.01793.x

Harris, L. T., & Fiske, S. T. (2011). Dehumanized perception a psychological means to facilitate atrocities, torture, and genocide? *Zeitschrift Fur Psychologie / Journal of Psychology*, *219*(3), 175–181. https://doi.org/10.1027/2151-2604/A000065

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, *10*(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Haslam, N. (2019). The Many Roles of Dehumanization in Genocide. In L. S. Newman (Ed.), *Confronting humanity at its worst: Social psychological perspectives on genocide* (pp. 119–138). Oxford University Press. https://doi.org/10.1093/OSO/9780190685942.003.0005

Haslam, N. (2021). The social psychology of dehumanization. In M. Kronfeldner (Ed.), *The Routledge Handbook of Dehumanization* (Online edn). Routeledge. https://doi.org/10.4324/9780429492464

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*(6), 937–950. https://doi.org/10.1037/0022-3514.89.6.937

Haslam, N., Bastian, B., & Bissett, M. (2004). Essentialist beliefs about personality and their implications. *Personality and Social Psychology Bulletin*, *30*(12), 1661–1673. https://doi.org/10.1177/0146167204271182

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review*

of Psychology, *65*, 399–423. https://doi.org/10.1146/annurev-psych-010213-115045

Haslam, N., & Loughnan, S. (2016). How Dehumanization Promotes Harm. In G. Miller, Arthur (Ed.), *The Social Psychology of Good and Evil* (2nd ed., pp. 140–158). Guilford Publications.

Haslam, N., & Stratemeyer, M. (2016). Recent research on dehumanization. *Current Opinion in Psychology*, *11*, 25–29.

Hatzfeld, J. (2003). *Machete Season: The Killers of Rwanda Speak.* (English ed). Farrar, Straus and Giroux.

Heng, Y. T., Wagner, D. T., Barnes, C. M., & Guarana, C. L. (2018). Archival research: Expanding the methodological toolkit in social psychology. *Journal of Experimental Social Psychology*, *78*, 14–22. https://doi.org/10.1016/J.JESP.2018.04.012

Hewstone, M., & Ward, C. (1985). Ethnocentrism and Causal Attribution in Southeast Asia. *Journal of Personality and Social Psychology*, *48*(3). https://doi.org/10.1037/0022-3514.48.3.614

Hodson, G., & Costello, K. (2007). Interpersonal disgust, ideological orientations, and dehumanization as predictors of intergroup attitudes. *Psychological Science*, *18*(8), 691–698. https://doi.org/10.1111/j.1467-9280.2007.01962.x

Hodwitz, O., & Massingale, K. (2021). Rhetoric and hate crimes: examining the public response to the Trump narrative. *Behavioral Sciences of Terrorism and Political Aggression*, 1–18. https://doi.org/10.1080/19434472.2021.1936121

Hornsey, M. J. (2008). Social Identity Theory and Self-categorization Theory: A Historical Review. *Social and Personality Psychology Compass*, *2*(1), 204–222. https://doi.org/10.1111/J.1751-9004.2007.00066.X

Kahneman, D. (2012). *A proposal to deal with questions about priming effects. An open letter to the scientific community*. http://www.nature.com/polopoly_fs/7.6716.1349271308!/suppinfoFile/KahnemanLetter.pdf

Karantzas, G. C., Simpson, J. A., & Pizzirani, B. (2022). The loss of humanness in close

relationships: An interpersonal model of dehumanization. *Current Opinion in Psychology*, *46*, 101317. https://doi.org/10.1016/J.COPSYC.2022.101317

Kasper, A., Frébert, N., & Testé, B. (2022). Caught COVID-19? Covidiot! *Social Psychology*, *53*(2), 84–95. https://doi.org/10.1027/1864-9335/A000478

Kelman, H. G. (1973). Violence without Moral Restraint: Reflections on the Dehumanization of Victims and Victimizers. *Journal of Social Issues*, *29*(4), 25–61. https://doi.org/10.1111/J.1540-4560.1973.TB00102.X

Kersbergen, I., & Robinson, E. (2019). Blatant Dehumanization of People with Obesity. *Obesity*, *27*(6), 1005–1012. https://doi.org/10.1002/OBY.22460

Kozak, M. N., Marsh, A. A., & Wegner, D. M. (2006). What do i think you're doing? Action identification and mind attribution. *Journal of Personality and Social Psychology*, *90*(4). https://doi.org/10.1037/0022-3514.90.4.543

Krauss, R. M., & Glucksberg, S. (1977). Social and nonsocial speech. *Scientific American*, *236*, 100–105. https://www.jstor.org/stable/24953896

Kronfeldner, M. (2021). The Routledge handbook of dehumanization. In M. Kronfeldner (Ed.), *The Routledge Handbook of Dehumanization*. Routeledge. https://doi.org/10.4324/9780429492464

Kteily, N., & 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. https://doi.org/10.1177/0963721417708230

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. https://doi.org/10.1037/PSPP0000048

Kteily, N., & Landry, A. P. (2022). Dehumanization: trends, insights, and challenges. *Trends in Cognitive Sciences*, *26*(3), 222–240. https://doi.org/10.1016/J.TICS.2021.12.003

Kteily, N., Psychology, E. B.-P. and S., & 2017, U. (2017). Backlash: The politics and real-world consequences of minority group dehumanization. *Journals.Sagepub.Com*, *43*(1), 87. https://doi.org/10.1177/0146167216675334

Kunst, J. R., Kteily, N., & Thomsen, L. (2017). “You Little Creep”: Evidence of Blatant Dehumanization of Short Groups: *Social Psychological and Personality Science*, *10*(2), 160–171. https://doi.org/10.1177/1948550617740613

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(26). https://doi.org/10.1073/pnas.251541498

Landry, A. P., Orr, R. I., & Mere, K. (2022). Dehumanization and mass violence: A study of mental state language in Nazi propaganda (1927–1945). *PLOS ONE*, *17*(11), e0274957. https://doi.org/10.1371/JOURNAL.PONE.0274957

Lang, J. (2010). Questioning Dehumanization: Intersubjective Dimensions of Violence in the Nazi Concentration and Death Camps. *Holocaust and Genocide Studies*, *24*(2), 225–246. https://doi.org/10.1093/HGS/DCQ026

Lang, J. (2020). The limited importance of dehumanization in collective violence. *Current Opinion in Psychology*, *35*, 17–20.

Lazović, V. (2021). Dehumanization Revisited: Media Reports on the Refugee Crisis in British Online Newspapers. *AA: Arbeiten Aus Anglistik Und Amerikanistik, 46*(2), 115–138. https://www.jstor.org/stable/27097917

Leidner, B., Castano, E., & Ginges, J. (2013). Dehumanization, Retributive and Restorative Justice, and Aggressive Versus Diplomatic Intergroup Conflict Resolution Strategies. *Personality and Social Psychology Bulletin*, *39*(2), 181–192. https://doi.org/10.1177/0146167212472208

Leyens, J. P., Cortes, B., Demoulin, S., Dovidio, J. F., Fiske, S. T., Gaunt, R., Paladino, M. P., Rodriguez-Perez, A., Rodriguez-Torres, R., & Vaes, J. (2003). Emotional prejudice, essentialism, and nationalism: The 2002 Tajfel Lecture. *European Journal of Social Psychology*, *33*(6), 703–717. https://doi.org/10.1002/EJSP.170

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*(1), 139–172. https://doi.org/10.1111/j.1751-2409.2007.00006.x

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. https://doi.org/10.1207/S15327957PSPR0402_06

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. https://doi.org/10.1002/EJSP.50

Lindén, M., Björklund, F., & Bäckström, M. (2016). What makes authoritarian and socially dominant people more positive to using torture in the war on terrorism? *Personality and Individual Differences*, *91*, 98–101.

Linville, P. W., & Jones, E. E. (1980). Polarized appraisals of out-group members. *Journal of Personality and Social Psychology*, *38*(5), 689–703. https://doi.org/10.1037/0022-3514.38.5.689

Ljujic, V., Vedder, P., Dekker, H., & Van Geel, M. (2012). Romaphobia: A unique phenomenon? *Romani Studies*, *22*(2), 141–152. https://doi.org/10.3828/RS.2012.8

Loughnan, Stephen, & Haslam, N. (2007). Animals and androids: Implicit associations between social categories and nonhumans. *Psychological Science*, *18*(2), 116–121. https://doi.org/10.1111/J.1467-9280.2007.01858.X

Loughnan, Steve, Haslam, N., Murnane, T., Vaes, J., Reynolds, C., & Suitner, C. (2010). Objectification leads to depersonalization: The denial of mind and moral concern to objectified others. *European Journal of Social Psychology*, *40*(5), 709–717. https://doi.org/10.1002/EJSP.755

Loughnan, Steve, Pina, A., Vasquez, E. A., & Puvia, E. (2013). Sexual Objectification Increases Rape Victim Blame and Decreases Perceived Suffering: *Psychology of Women Quarterly*, *37*(4), 455–461. https://doi.org/10.1177/0361684313485718

Luft, A. (2015). Toward a Dynamic Theory of Action at the Micro Level of Genocide: Killing, Desistance, and Saving in 1994 Rwanda. *Sociological Theory*, *33*(2), 148–172.

https://doi.org/10.1177/0735275115587721

Luft, A. (2019). *Dehumanization and the Normalization of Violence: It's Not What You Think*. Items: Social Science Research Council. https://items.ssrc.org/insights/dehumanization-and-the-normalization-of-violence-its-not-what-you-think/

MacInnis, C. C., & Hodson, G. (2012). Intergroup bias toward "Group X": Evidence of prejudice, dehumanization, avoidance, and discrimination against asexuals: *Group Processes & Intergroup Relations*, *15*(6), 725–743. https://doi.org/10.1177/1368430212442419

Manne, K. (2016). Humanism: A Critique. *Social Theory and Practice*, *42*(2), 389–415. https://doi.org/10.5840/SOCTHEORPRACT201642221

Manne, K. (2018). *Down girl: The logic of misogyny* (Online edn). Oxford University Press. https://doi.org/10.1093/oso/9780190604981.001.0001

Maoz, I., & McCauley, C. (2008). Threat, Dehumanization, and Support for Retaliatory Aggressive Policies in Asymmetric Conflict: *Journal of Conflict Resolution*, *52*(1), 93–116. https://doi.org/10.1177/0022002707308597

Marcu, A., & Chryssochoou, X. (2005). Exclusion of ethnic groups from the realm of humanity: Prejudice against the Gypsies in Britain and in Romania. *Psicologia Politica*, *30*(May 2005), 41–56. https://psycnet.apa.org/record/2005-16373-003

Markowitz, D. M., & Slovic, P. (2020). Social, psychological, and demographic characteristics of dehumanization toward immigrants. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9260–9269. https://doi.org/10.1073/PNAS.1921790117

Markowitz, D. M., & Slovic, P. (2021). Why we dehumanize illegal immigrants: A US mixed-methods study. *PLoS ONE*, *16*(10). https://doi.org/10.1371/JOURNAL.PONE.0257912

Martínez, R., Rodríguez-Bailón, R., & Moya, M. (2012). Are they Animals or Machines? Measuring Dehumanization. *The Spanish Journal of Psychology*, *15*(3), 1110–1122. https://doi.org/10.5209/rev_SJOP.2012.v15.n3.39401

McCall, C., & Blascovich, J. (2009). How, When, and Why to Use Digital Experimental Virtual Environments to Study Social Behavior. *Social and Personality Psychology Compass*, *3*(5), 744–758. https://doi.org/10.1111/J.1751-9004.2009.00195.X

Mendelsohn, J., Tsvetkov, Y., & Jurafsky, D. (2020). A Framework for the Computational Linguistic Analysis of Dehumanization. *Frontiers in Artificial Intelligence*, *3*, 540127. https://doi.org/10.3389/FRAI.2020.00055/BIBTEX

Miranda, M., Gouveia-Pereira, M., & Vaes, J. (2014). When in Rome... Identification and acculturation strategies among minority members moderate the dehumanisation of the majority outgroup. *European Journal of Social Psychology*, *44*(4), 327–336. https://doi.org/10.1002/EJSP.2025

Mixed Migration Centre. (2023). *Climate and mobility case study January 2023: Lagos, Nigeria: Ajegunle*. https://reliefweb.int/report/nigeria/climate-and-mobility-case-study-january-2023-western-sahel-lagos-nigeria-ajegunle

Morehouse, K. N., Maddox, K., & Banaji, M. R. (2023). All human social groups are human, but some are more human than others: A comprehensive investigation of the implicit association of “Human” to US racial/ethnic groups. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(22), e2300995120. https://doi.org/10.1073/PNAS.2300995120

Murray, J. (2020, May 2). Channel 4 Dispatches report accused of “dehumanising” Travellers. *The Guardian*. https://www.theguardian.com/world/2020/may/02/truth-about-traveller-crime-channel-4-dispatches-show-accused-dehumanising

Nariman, H. S., Hadarics, M., Soufizadeh, A. M., & Kende, A. (2020). The mediating role of moral exclusion between authoritarianism and outgroup discrimination. *International Journal of Intercultural Relations*, *74*. https://doi.org/10.1016/j.ijintrel.2019.10.001

Ng, K. (2021). Humanism: A Defence. *Philosophical Topics*, *59*(1), 145–164. https://www.jstor.org/stable/48652165

Nielsen, A. H., Kvande, M. E., & Angel, S. (2023). Humanizing and dehumanizing intensive care: Thematic synthesis (HumanIC). *Journal of Advanced Nursing*, *79*(1), 385–401.

https://doi.org/10.1111/JAN.15477

Nussbaum, M. C. (1995). Objectification. *Philosophy & Public Affairs*, *24*(4), 249–291. https://doi.org/10.1111/J.1088-4963.1995.TB00032.X

Obermann, M. L. (2011). Moral disengagement in self-reported and peer-nominated school bullying. *Aggressive Behavior*, *37*(2), 133–144. https://doi.org/10.1002/AB.20378

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Opeyemi, S. (2020, June). *Lagos state is likely to get hotter and more humid, leading to greater health risks*. The Conversation. https://theconversation.com/lagos-state-is-likely-to-get-hotter-and-more-humid-leading-to-greater-health-risks-140327

Opotow, S. (1990). Moral Exclusion and Injustice: An Introduction. In *Journal of Social Issues* (Vol. 46, Issue 1). https://doi.org/10.1111/j.1540-4560.1990.tb00268.x

Over, H. (2021a). Seven Challenges for the Dehumanization Hypothesis. *Perspectives on Psychological Science*, *16*(1), 3–13. https://doi.org/10.1177/1745691620902133

Over, H. (2021b). Falsifying the Dehumanization Hypothesis. *Perspectives on Psychological Science*, *16*(1), 33–38. https://doi.org/10.1177/1745691620969657

Over, H., & Cook, R. (2022). Perceptual Dehumanization Theory: A Critique. *Psychological Review*. https://doi.org/10.1037/REV0000388

Over, H., & McCall, C. (2018). Becoming us and them: Social learning and intergroup bias. *Social and Personality Psychology Compass*, *12*(4), e12384. https://doi.org/10.1111/SPC3.12384

Pacilli, M. G., Roccato, M., Pagliaro, S., & Russo, S. (2015). From political opponents to enemies? The role of perceived moral distance in the animalistic dehumanization of the political outgroup: *Group Processes & Intergroup Relations*, *19*(3), 360–373. https://doi.org/10.1177/1368430215590490

Paladino, M. P., Leyens, J. P., Rodriguez, R., Rodriguez, A., Gaunt, R., & Demoulin, S. (2002). Differential Association of Uniquely and Non Uniquely Human Emotions with the Ingroup and the Outgroup. *Group Processes & Intergroup Relations*, *5*(2), 105–117.

https://doi.org/10.1177/1368430202005002539

Paladino, M. P., Vaes, J., Castano, E., Demoulin, S., & Leyens, J. P. (2004). Emotional infra-humanization in intergroup relations: The role of national identification in the attribution of primary and secondary emotions to Italians and Germans. *Cahiers de Psychologie Cognitive*, *22*(4–5).

Phillips, B. (2022). “They’re Not True Humans:” Beliefs about Moral Character Drive Denials of Humanity. *Cognitive Science*, *46*(2). https://doi.org/10.1111/COGS.13089

Phillips, B. (2023). Normative dehumanization and the ordinary concept of a true human. *Current Research in Ecological and Social Psychology*, *5*, 100155. https://doi.org/10.1016/J.CRESP.2023.100155

Pizzirani, B., & Karantzas, G. C. (2018). The association between dehumanization and intimate partner abuse: *Journal of Social and Personal Relationships*, *36*(5), 1527–1541. https://doi.org/10.1177/0265407518811673

Posselt, G. (2017). Can Hatred Speak? On the Linguistic Dimensions of Hate Crime. *Linguistik Online*, *82*(3). https://doi.org/10.13092/LO.82.3712

Prati, F., Crisp, R. J., Meleady, R., & Rubini, M. (2016). Humanizing Outgroups Through Multiple Categorization: The Roles of Individuation and Threat. *Personality and Social Psychology Bulletin*, *42*(4), 526–539. https://doi.org/10.1177/0146167216636624

Prati, F., Crisp, R. J., & Rubini, M. (2015). Counter-stereotypes reduce emotional intergroup bias by eliciting surprise in the face of unexpected category combinations. *Journal of Experimental Social Psychology*, *61*, 31–43. https://doi.org/10.1016/j.jesp.2015.06.004.

Prati, F., & Giner-Sorolla, R. (2018). Perceiving mixed valence emotions reduces intergroup dehumanisation. *Cognition & Emotion*, *32*(5), 1018–1031. https://doi.org/10.1080/02699931.2017.1383885

Rai, T. S., & Fiske, A. P. (2011). Moral Psychology Is Relationship Regulation: Moral Motives for Unity, Hierarchy, Equality, and Proportionality. *Psychological Review*, *118*(1), 57–75. https://doi.org/10.1037/A0021867

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. https://doi.org/10.1073/PNAS.1705238114

Reicher, S., Haslam, S. A., & Rath, R. (2008). Making a Virtue of Evil: A Five-Step Social Identity Model of the Development of Collective Hate. *Social and Personality Psychology Compass*, *2*(3), 1313–1344. https://doi.org/10.1111/j.1751-9004.2008.00113.x

Reinecke, M. G., Wilks, M., & Bloom, P. (2021). Developmental changes in perceived moral standing of robots. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*. https://escholarship.org/uc/item/8f32d068

Reis, H. T., & Gosling, S. D. (2010). Social Psychological Methods Outside the Laboratory. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of Social Psychology, Volume 1* (5th ed., pp. 82–114). Wiley.

Reynolds, K. J., Alexander Haslam, S., & Turner, J. C. (2012). Prejudice, social identity and social change: resolving the Allportian problematic. In *Beyond Prejudice*. https://doi.org/10.1017/cbo9781139022736.004

Richards, L., Fernández-Reino, M., & Blinder, S. (2023). *BRIEFING - UK Public Opinion toward Immigration: Overall Attitudes and Level of Concern*. https://migrationobservatory.ox.ac.uk/resources/briefings/uk-public-opinion-toward-immigration-overall-attitudes-and-level-of-concern/

Riva, P., & Andrighetto, L. (2012). “Everybody feels a broken bone, but only we can feel a broken heart”: Group membership influences the perception of targets’ suffering. *European Journal of Social Psychology*, *42*(7), 801–806. https://doi.org/10.1002/EJSP.1918

Roccato, M., Mannarini, T., & Pacilli, M. G. (2018). Ingroup Identification, Outgroup Infrahumanization, and Intention to Mobilize in Land Use Conflicts. *Environment and Behavior*, *50*(5), 512–534. https://doi.org/10.1177/0013916517708396

Rodrigues, D. L., Lopes, D., & Huic, A. (2021). What Drives the Dehumanization of Consensual Non-Monogamous Partners? *Archives of Sexual Behavior*, *50*(4), 1587–

1597. https://doi.org/10.1007/S10508-020-01895-5

Rodríguez-Pérez, A., Delgado-Rodríguez, N., Betancor-Rodríguez, V., Leyens, J.-P., & Vaes, J. (2011). Infra-humanization of outgroups throughout the world. The role of similarity, intergroup friendship, knowledge of the outgroup, and status. *Anales de Psicología*, *27*(3), 679–687.

Rodríguez-Torres, R., Leyens, J. P., Pérez, A. R., Rodriguez, V. B., Quiles Del Castillo, M. N., Demoulin, S., & Cortés, B. (2005). The lay distinction between primary and secondary emotions: A spontaneous categorization? *International Journal of Psychology*, *40*(2), 100–107. https://doi.org/10.1080/00207590444000221

Rodríguez, V. B., Mateo, E. A., Rodríguez-Pérez, A., & Rodríguez, N. D. (2016). Do they feel the same as us? The infrahumanization of individuals with Down syndrome. *Psicothema*, *28*(3), 311–317. https://doi.org/10.7334/PSICOTHEMA2016.10

Rousseau, D. L., Gorman, B., & Baranik, L. E. (2023). Crossing the Line: Disgust, Dehumanization, and Human Rights Violations. *Socius*, *9*. https://doi.org/10.1177/23780231231157686

Rubio-Fernández, P. (2017). The director task: A test of Theory-of-Mind use or selective attention? *Psychonomic Bulletin and Review*, *24*(4), 1121–1128. https://doi.org/10.3758/S13423-016-1190-7

Rudman, L. A., & Mescher, K. (2012). Of Animals and Objects: Men's Implicit Dehumanization of Women and Likelihood of Sexual Aggression. *Personality and Social Psychology Bulletin*, *38*(6), 734–746. https://doi.org/10.1177/0146167212436401

Ruibyte, L., Viduoliene, E., & Balseviciene, B. (2016). Perception of criminals: What stereotypes hold future law enforcement officers? *SOCIETY. INTEGRATION. EDUCATION. Proceedings of the International Scientific Conference*, *1*, 515–523. https://doi.org/10.17770/SIE2016VOL1.1520

Russell, J. A., & Fehr, B. (1994). Fuzzy Concepts in a Fuzzy Hierarchy: Varieties of Anger. *Journal of Personality and Social Psychology*, *67*(2), 186–205. https://doi.org/10.1037/0022-3514.67.2.186

Sakki, I., & Castrén, L. (2022). Dehumanization through humour and conspiracies in online hate towards Chinese people during the COVID-19 pandemic. *British Journal of Social Psychology*, *61*(4), 1418–1438. https://doi.org/10.1111/BJSO.12543

Saminaden, A., Loughnan, S., & Haslam, N. (2010). Afterimages of savages: Implicit associations between primitives, animals and children. *British Journal of Social Psychology*, *49*(1), 91–105. https://doi.org/10.1348/014466609X415293

Sánchez, R. P., & García, C. Y. (2016). Mortality salience, political orientation and minimization of terrorists' secondary emotions. *Psicothema*, *28*(1), 47–52. https://doi.org/10.7334/PSICOTHEMA2015.147

Savage, R. (2013). Modern genocidal dehumanization: a new model. *Patterns of Prejudice*, *47*(2), 139–161. https://doi.org/10.1080/0031322X.2012.754575

Seabrook, R. C., Ward, L. M., & Giaccardi, S. (2019). Less than human? Media use, objectification of women, and men's acceptance of sexual aggression. *Psychology of Violence*, *9*(5), 536–545. https://doi.org/10.1037/VIO0000198

Seager, M., & Barry, J. A. (2019). Cognitive Distortion in Thinking About Gender Issues: Gamma Bias and the Gender Distortion Matrix. *The Palgrave Handbook of Male Psychology and Mental Health*, 87–104. https://doi.org/10.1007/978-3-030-04384-1_5

Seelau, S. M., & Seelau, E. P. (2005). Gender-Role Stereotypes and Perceptions of Heterosexual, Gay and Lesbian Domestic Violence. *Journal of Family Violence*, *20*(363). https://doi.org/10.1007/S10896-005-7798-4

Sidanius, J., & Pratto, F. (1999). *Social dominance : an intergroup theory of social hierarchy and oppression*.

Simon, J. C., & Gutsell, J. N. (2020). Effects of minimal grouping on implicit prejudice, infrahumanization, and neural processing despite orthogonal social categorizations. *Group Processes and Intergroup Relations*, *23*(3), 323–343. https://doi.org/10.1177/1368430219837348

Sin, K. F., Yang, L., & Ye, F. T. F. (2023). Self-dehumanization and other-dehumanization toward students with special educational needs: examining their prevalence,

consequences and identifying solutions—a study protocol. *BMC Psychology*, *11*(1), 1–12. https://doi.org/10.1186/S40359-023-01178-3/TABLES/3

Smith, D. L. (2011). *Less Than Human: Why We Demean, Enslave, and Exterminate Others*. St. Martin's Publishing Group.

Smith, D. L. (2014). Dehumanization, essentialism, and moral psychology. *Philosophy Compass*, *9*(11), 814–824. https://doi.org/10.1111/PHC3.12174

Smith, D. L. (2016). Paradoxes of Dehumanization. *Social Theory and Practice*, *42*(2), 416–443. https://doi.org/10.5840/SOCTHEORPRACT201642222

Smith, D. L. (2020). *On Inhumanity: Dehumanization and How to Resist It*. Oxford University Press. https://doi.org/10.1093/oso/9780190923006.001.0001

Smith, D. L. (2021a). Challenge 8: A Response to Over (2021). *Perspectives on Psychological Science*, *16*(1), 22–23. https://doi.org/10.1177/1745691620917351

Smith, D. L. (2021b). *Making Monsters: The Uncanny Power of Dehumanization*. Harvard University Press. https://doi.org/10.4159/9780674269781

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

Sofsky, W. (1996). *The Order of Terror: The Concentration Camp*. Princeton University Press. https://doi.org/10.1515/9781400822188

Stanton, G. H. (1998). *The 8 Stages of Genocide*. Genocide Watch. http://genocidewatch.net/2013/03/14/the-8-stages-of-genocide/

Staub, E. (1989). *The Roots of Good and Evil: The Origins of Genocide and Other Group Violence*. Cambridge University Press. https://archive.org/details/rootsofevilorigi0000stau/page/n5/mode/2up

Stearns, D. C., & Parrott, W. G. (2012). When feeling bad makes you look good: Guilt, shame, and person perception. *Cognition & Emotion*, *26*(3), 407–430. https://doi.org/10.1080/02699931.2012.675879

Sterna, R., & Zibrek, K. (2021). Psychology in Virtual Reality: Toward a Validated Measure

of Social Presence. *Frontiers in Psychology*, *12*, 705448. https://doi.org/10.3389/FPSYG.2021.705448/BIBTEX

Struch, N., & Schwartz, S. H. (1989). Intergroup Aggression: Its Predictors and Distinctness From In-Group Bias. *Journal of Personality and Social Psychology*, *56*(3), 364–373. https://doi.org/10.1037/0022-3514.56.3.364

Tajfel, H., & Turner, J. (1979). An Integrative Theory of Intergroup Conflict. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33–47). Brooks/Cole.

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgement. *British Journal of Psychology (London, England : 1953)*, *54*. https://doi.org/10.1111/j.2044-8295.1963.tb00865.x

Tam, T., Hewstone, M., Cairns, E., Tausch, N., Maio, G., & Kenworthy, J. (2007). The Impact of Intergroup Emotions on Forgiveness in Northern Ireland. *Group Processes & Intergroup Relations*, *10*(1), 119–135. https://doi.org/10.1177/1368430207071345

Tam, T., Hewstone, M., Kenworthy, J. B., Cairns, E., Marinetti, C., Geddes, L., & Parkinson, B. (2008). Postconflict Reconciliation: Intergroup Forgiveness and Implicit Biases in Northern Ireland. *Journal of Social Issues*, *64*(2), 303–320. https://doi.org/10.1111/J.1540-4560.2008.00563.X

Taylor, P. (2020). System Entrapment: Dehumanization While Help-Seeking for Suicidality in Women Who Have Experienced Intimate Partner Violence. *Qualitative Health Research*, *30*(4), 530–546. https://doi.org/10.1177/1049732319857671

Tileagă, C. (2007). Ideologies of moral exclusion: A critical discursive reframing of depersonalization, delegitimization and dehumanization. *British Journal of Social Psychology*, *46*(4), 717–737. https://doi.org/10.1348/014466607X186894

Tirrell, L. (2012). Genocidal Language Games. In I. Maitra & M. K. McGowan (Eds.), *Speech and Harm: Controversies over Free Speech* (pp. 174–221). Oxford University Press. http://refhub.elsevier.com/S0010-0277(21)00101-3/rf0285

Todosijević, B., & Enyedi, Z. (2002). Authoritarianism vs. Cultural Pressure Anti-Gypsy

Prejudice in Hungary. *Journal of Russian & East European Psychology*, *40*(5), 31–54. https://doi.org/10.2753/RPO1061-0405400531

United Nations. (2018, October 31). *Refugees, Migrants Branded 'Threats', Dehumanized in Campaigns Seeking Political Gain, High Commissioner Tells Third Committee, Appealing for Return to Dignity*. United Nations Meetings Coverage and Press Releases. https://press.un.org/en/2017/gashc4247.doc.htm

United Nations. (2020, December 16). *New UN report warns against 'dehumanizing' migrants looking for work*. UN News: Global Perspective Human Stories. https://news.un.org/en/story/2022/12/1131797

Utych, S. M. (2018). How Dehumanization Influences Attitudes toward Immigrants. *Political Research Quarterly*, *71*(2), 440–452. https://doi.org/10.1177/1065912917744897

Vaes, J. (2023). Dehumanization after all: Distinguishing intergroup evalutation from trait-based dehumanization. *Cognition*, *231*, 105329. https://doi.org/10.1016/J.COGNITION.2022.105329

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*(1), 64–106. https://doi.org/10.1080/10463283.2012.665250

Vaes, J., Paladino, M. P., Castelli, L., Leyens, J. P., & Giovanazzi, A. (2003). On the Behavioral Consequences of Infrahumanization: The Implicit Role of Uniquely Human Emotions in Intergroup Relations. *Journal of Personality and Social Psychology*, *85*(6), 1016–1034. https://doi.org/10.1037/0022-3514.85.6.1016

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

Vaes, J., Paladino, M. P., & Leyens, J. P. (2002). The lost e-mail: Prosocial reactions induced by uniquely human emotions. *British Journal of Social Psychology*, *41*(4), 521–534. https://doi.org/10.1348/014466602321149867

Vala, J., Pereira, C., & Costa-Lopes, R. (2009). Is the attribution of cultural differences to minorities an expression of racial prejudice? *International Journal of Psychology : Journal International de Psychologie*, *44*(1), 20–28. https://doi.org/10.1080/00207590802057837

Van Fleet, D. D., & Atwater, L. (1997). Gender neutral names: Don't be so sure! *Sex Roles*, *37*(1–2), 111–123. https://doi.org/10.1023/A:1025696905342/METRICS

Viki, T. G., & Abrams, D. (2003). Infra-humanization: Ambivalent sexism and the attribution of primary and secondary emotions to women. *Journal of Experimental Social Psychology*, *39*(5), 492–499. https://doi.org/10.1016/S0022-1031(03)00031-3

Viki, T. G., & Calitri, R. (2008). Infrahuman outgroup or suprahuman ingroup: The role of nationalism and patriotism in the infrahumanization of outgroups. *European Journal of Social Psychology*, *38*, 1054–1061. https://doi.org/https://doi.org/10.1002/ejsp.495

Viki, T. G., Fullerton, I., Raggett, H., Tait, F., & Wiltshire, S. (2012). The role of dehumanization in attitudes toward the social exclusion and rehabilitation of sex offenders. *Journal of Applied Social Psychology*, *42*(10), 2349–2367. https://doi.org/10.1111/j.1559-1816.2012.00944.x

Viki, T. G., Osgood, D., & Phillips, S. (2013). Dehumanization and self-reported proclivity to torture prisoners of war. *Journal of Experimental Social Psychology*, *49*(3), 325–328.

Viki, T. G., Winchester, L., Titshall, L., Chisango, T., Pina, A., & Russell, R. (2006). Beyond secondary emotions: The infrahumanization of outgroups using human-related and animal-related words. *Social Cognition*, *24*(6), 753–775. https://doi.org/10.1521/SOCO.2006.24.6.753

Waxman, S. R. (2010). Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. *European Journal of Social Psychology*, *40*(4). https://doi.org/10.1002/ejsp.732

Waytz, A., & Epley, N. (2012). Social connection enables dehumanization. *Journal of Experimental Social Psychology*, *48*(1), 70–76. https://doi.org/10.1016/j.jesp.2011.07.012

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383–388. https://doi.org/10.1016/J.TICS.2010.05.006

Webber, E. (2023). *They got Brexit done. So why are UK Tories still angry about immigration?* Politico. https://www.politico.eu/article/tories-brexit-done-uk-conservatives-on-immigration-rishi-sunak/

West, S. J., & Thomson, N. D. (2023). Identifying the emotions behind apologies for severe transgressions. *Motivation and Emotion*, *47*(2), 257–269. https://doi.org/10.1007/s11031-022-09993-8

Wiener, R. L., Gervais, S. J., Brnjic, E., & Nuss, G. D. (2014). Dehumanization of older people: The evaluation of hostile work environments. *Psychology, Public Policy, and Law*, *20*(4), 384–397. https://doi.org/10.1037/LAW0000013

Wilks, M., Caviola, L., Kahane, G., & Bloom, P. (2021). Children Prioritize Humans Over Animals Less Than Adults Do. *Psychological Science*, *32*(1), 27–38. https://doi.org/10.1177/0956797620960398

Williams, R. M. (1964). *Strangers next door: Ethnic relations in American Communities*. Prentice-Hall.

Wohl, M. J. A., Hornsey, M. J., & Bennett, S. H. (2012). Why group apologies succeed and fail: Intergroup forgiveness and the role of primary and secondary emotions. *Journal of Personality and Social Psychology*, *102*(2). https://doi.org/10.1037/a0024838

Yousefi, Y. (2020). *Environmental and Social Impacts of Fast Fashion*. Sustainable Development Watch. https://sdwatch.eu/2020/02/environmental-and-social-impacts-of-fast-fashion/

Zebel, S., Zimmermann, A., Viki, T. G., & Doosje, B. (2008). Dehumanization and Guilt as Distinct but Related Predictors of Support for Reparation Policies on JSTOR. *Political Psychology*, *29*(2), 193–219. https://www.jstor.org/stable/20447112

Zhang, H., Chan, D. K. S., Teng, F., & Zhang, D. (2015). Sense of interpersonal security and preference for harsh actions against others: The role of dehumanization. *Journal of*

Experimental Social Psychology, *56*, 165–171. https://doi.org/10.1016/J.JESP.2014.09.014

Zick, A., Küpper, B., & Hövermann, A. (2011). *Intolerance, Predjudice and Discrimination. A European Report*. https://doi.org/http://dx.doi.org/10.15496/publikation-5599

Zimbardo, P. G. (2007). *The Lucifer Effect: Understanding How Good People Turn Evil*. Random House.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, *12*(4), 399–413. https://doi.org/10.1037/1082-989X.12.4.399