Intergroup preference, not dehumanization, explains social biases in emotion attribution

Florence E. Enock*, Steven P. Tipper, Harriet Over

Department of Psychology, University of York, York YO10 5DD, United Kingdom

ARTICLE INFO

Keywords:
Dehumanization
Infrahumanization
Intergroup bias
Prejudice

ABSTRACT

Psychological models can only help improve intergroup relations if they accurately characterise the mechanisms underlying social biases. The claim that outgroups suffer dehumanization is near ubiquitous in the social sciences. We challenge the most prominent psychological model of dehumanization - infrahumanization theory - which holds outgroup members are subtly dehumanized by being denied human emotions. We examine the theory across seven intergroup contexts in thirteen pre-registered and highly powered experiments (N = 1690). We find outgroup members are not denied uniquely human emotions relative to ingroup members. Rather, they are ascribed prosocial emotions to a lesser extent but antisocial emotions to a greater extent. Apparent evidence for infrahumanization is better explained by ingroup preference, outgroup derogation and stereotyping. Infrahumanization theory may obscure more than it reveals about intergroup bias.

1. Introduction

The claim that outgroup members are perceived as ‘less than human’ has been extremely influential in social psychology, social neuroscience, philosophy and sociology. It has entered into public rhetoric as well, regularly being discussed in the media. Blatant forms of dehumanization are thought to reveal themselves in propaganda and other forms of hate speech in which outgroup members are described as less than human creatures, for example as similar to rats, parasites or vermin (Haslam, 2006; Smith, 2011). Blatant dehumanization has been linked to extreme intergroup harm such as genocide, torture and police brutality towards African Americans (Goff, Eberhardt, Williams, & Jackson, 2008; Goff, Jackson, Di Leone, Culotta, & DiTomasso, 2014; Smith, 2011; Tirrell, 2012).

Subtler forms of dehumanization, in which outgroups are considered somewhat less human, are hypothesised to be widespread and are typically studied in lab-based settings (Harris & Fiske, 2006; Haslam, 2006; Leyens et al., 2000, 2001). In subtle forms of dehumanization, outgroup members are thought to possess uniquely human qualities to a lesser extent than do the ingroup. Three psychological models of subtle dehumanization have been particularly prominent. According to Harris and Fiske (2006), to the extent outgroups are dehumanized, they are thought to possess mental states to a lesser extent than do the ingroup. According to the dual model, outgroups are thought to possess uniquely human character traits to a lesser extent than do the ingroup (Haslam, 2006). According to infrahumanization theory, perhaps the most prominent of the three models, outgroups are thought to possess uniquely human emotions to a lesser extent than do the ingroup (Leyens et al., 2000, 2001). These subtle forms of dehumanization have been linked to negative outcomes including reduced prosocial behaviour towards outgroups (Cuddy, Rock, & Norton, 2007; Vaes, Paladino, Castelli, Leyens, & Giovanazzi, 2003). In a world of social division, with frequently occurring cases of discrimination based on religion, ethnicity and gender, to name only a few, understanding the extent and consequences of dehumanization is crucial.

In recent years, several theoretical critiques of research on dehumanization have emerged (Appiah, 2008; Bloom, 2017; Lang, 2010, 2020; Manne, 2016, 2018; Over, 2020, 2021; Rai, Valdesolo, & Graham, 2017; Smith, 2014, 2020). These critiques suggest that perceiving outgroups as ‘less than human’ might be less common than it first appears. Considering blatant dehumanization, as evidenced by historical examples of propaganda, Manne (2016) and Bloom (2017) have both pointed out that victims said to be ‘dehumanized’ are often described with terms that only really make sense when applied to humans, albeit negative and antisocial ones. For example, in Nazi propaganda, Jewish people were frequently described as ruthless, corrupt, treacherous and criminally minded, terms out of place when used to describe an animal or a machine. Relatedly, the hypothesised causal connection between
dehumanization and intergroup harm has been questioned. Several theorists suggest that being perceived as having certain human qualities, such as being corrupt, spiteful or deceptive, may actually increase people’s risk of harm (Appiah, 2008; Bloom, 2017; Lang, 2010, 2020; Manne, 2016, 2018; Over, 2020, 2021). Empirical research showing that morally-motivated intergroup harm is not linked to dehumanization lends support to these critiques (Rai et al., 2017).

Turning to lab-based research, Over (2020, 2021) argues that what appears to be evidence for dehumanization, as operationalised by the dual model (Haslam, 2006) and infrahumanization theory (Leyens et al., 2000, 2001), may be better explained in terms of intergroup preference effects (a general tendency to prefer the ingroup to the outgroup). According to the dual model of dehumanization, outgroup members tend to be attributed uniquely human character traits to a lesser extent than are the ingroup (Haslam, 2006). However, to date, the overwhelming majority of the traits included in empirical research are socially desirable, for example, warmth, rationality, civility and refinement. Over (2020, 2021) hypothesises that while outgroup members may be thought to possess some uniquely human qualities to a lesser extent, for example, civility, refinement and rationality, there may be other uniquely human qualities that are more strongly attributed to the outgroup than the ingroup. For example, antisocial human characteristics such as jealousy, arrogance and bitterness seem to only make sense in the context of humans but are unlikely to be attributed more strongly to ingroup than outgroup members.

Recent experimental work from Enock, Flavell, Tipper, and Over (2021) supports Over’s critique, presenting an empirical challenge to the dual model of dehumanization (Haslam, 2006). Enock et al. (2021) first established that people tend to associate undesirable characteristics as well as desirable ones with humans, confirming an omission from the dual model. Subsequently, seven experiments tested the predictions of the dual model directly against a social preference account in three distinct intergroup contexts - political opponents, immigrants and criminals. Results showed no evidence for dehumanization when undesirable as well as desirable human traits were included in the stimuli. Rather, in line with the social preference account, desirable traits were ascribed more strongly to ingroup members than outgroup members and undesirable traits more strongly to outgroup members than ingroup members, irrespective of perceived humanness (Enock et al., 2021).

Perhaps the most prominent psychological model of dehumanization is infrahumanization theory (Leyens et al., 2000, 2001). This theory is distinct from the dual model in that it proposes outgroup members are denied uniquely human emotions rather than character traits. The model is founded on the widespread notion that there is a distinction between secondary emotions (such as pride and guilt) and primary emotions (such as happiness and anger). The former are considered unique to humans, the latter shared with other animals (Demoulin et al., 2004; Ekman, 1992; Leyens et al., 2000). Seminal work has found that when choosing emotions to best describe different groups, people preferentially ascribe uniquely human emotions more strongly to ingroup members (Leyens et al., 2001). For example, across a range of social contexts, participants ascribed uniquely human emotions such as hope, compassion, pride, melancholy, disappointment and remorse, to ingroup members to a greater extent than to outgroup members (Banton, West, & Kinney, 2020; Cortes, Demoulin, Rodriguez, Rodriguez, & Leyens, 2005; Leyens et al., 2001; Paladino et al., 2002; Prati, Crisp, Meleady, & Rubin, 2016).

Infrahumanization research has proliferated in recent years (Leyens, 2009; Vaes, Leyens, Paola Paladino, & Pires Miranda, 2012). Effects have been reported across explicit and implicit measures (Boccato, Cortes, Demoulin, & Leyens, 2007; Paladino et al., 2002), and a multitude of intergroup contexts, including regional, religious and racial identities (Banton et al., 2020; Rodríguez-Pérez, Delgado-Rodríguez, Betancor-Rodríguez, Leyens, & Vaes, 2011), university affiliations (Vaes et al., 2003) and minimal groups (Demoulin et al., 2009; Simon & Gutsell, 2020). The importance of the model is highlighted by the use of infrahumanization as an outcome measure in interventions to improve intergroup relations (Brown, Eller, Leeds, & Stace, 2007; Capozza, Trifiletti, Vezzali, & Favara, 2013; Prati et al., 2016; Tam et al., 2007).

Of key importance to infrahumanization theory is the claim that subtle dehumanization is distinct from intergroup preference because participants ascribe both positive (e.g., hope, admiration) and negative (e.g., guilt, remorse) uniquely human emotions more strongly to the ingroup than the outgroup (Leyens et al., 2000, 2001). That the effect is observed for negative human emotions is claimed to be crucial for separating infrahumanization from intergroup preference. According to infrahumanization theory, the process of perceiving others as lacking negative human emotions shows a subtle form of dehumanization that is separate from derogation because here, ‘humanness’ is different from ‘good’ (Castano & Giner-Sorolla, 2006; Haslam & Loughnan, 2014; Haslam & Stratemeyer, 2016). As Leyens et al. (2000, p.189) note in their original hypotheses:

**People should more easily associate their ingroup than an outgroup with secondary emotions. This preferential association should be true-independent of the valence of the secondary emotions. Indeed, it is the category of secondary emotions as such that is considered typically human. No qualification is made for positive or negative secondary emotions.**

Explaining this further, Leyens et al. (2001, p. 398) argue:

> If the attribution of secondary emotions to the ingroup reflected a mere-positivity effect, it would lose its interest and originality… would people also select more negative secondary emotions for their ingroup than for the outgroup? A positive answer… would get rid of a simple positivity bias in the case of secondary emotions.

In their original studies, Leyens et al. (2001) included valence (positive/negative) as a factor in their analysis. They found that more secondary emotions were attributed to the ingroup than the outgroup and that this effect was not qualified by the valence of the emotions (see Leyens et al., 2001, p. 402, Fig. 2). Subsequent work from Cortes et al. (2005) demonstrated similar effects, showing more positive (e.g., contentment, delight) and negative (e.g., melancholy, resignation) uniquely human emotions to be ascribed to the ingroup than the outgroup, again not qualified by emotion valence (see Cortes et al., 2005, p. 247, Fig. 1). Convergent evidence is provided by research that treats valence and humanness as continuous factors (Castano & Giner-Sorolla, 2006) and by research that has measured the attribution of positive (e.g., amazement, compassion) and negative (e.g., despair, guilt) emotions in separate experiments (Paladino et al., 2002).

Owing to the inclusion of negative as well as positive emotions in infrahumanization research, the theory thus appears to be immune to Over’s (2020, 2021) critique. However, we reconsider infrahumanization theory through a framework that understands emotions as social as well as individual experiences (Parkinson, 1996). Some emotions are by definition positive to experience but are somewhat unkind to others (e.g., schadenfreude), while some may be negative to experience but are not inherently unkind to others (e.g., disappointment). We suggest that although emotions such as guilt and remorse are negative to experience, they are not necessarily antisocial in character. Rather, they indicate appropriate responses to moral wrong-doing and thus people who display them are viewed positively (Searns & Parrott, 2012). Understanding emotion as a social phenomenon raises the conceptual distinction between valence (positive or negative to experience) and sociality (prosocial or antisocial as viewed by others).

To our knowledge, no work has yet measured whether previously reported infrahumanization effects are independent of emotion sociality. Terms frequently included as exemplars of negative uniquely human emotions in infrahumanization work such as melancholy, guilt and remorse, may be negative to experience, but are not obviously antisocial. This omission makes it impossible to determine whether infrahumanization really is separable from ingroup preference and thus...
whether it holds unique explanatory value in intergroup relations.

Twenty years since infrahumanization theory was proposed, we revisit and test its founding claims in thirteen pre-registered, highly powered experiments. In our first six experiments (Study 1) we show that previously reported infrahumanization effects broadly replicate across multiple intergroup contexts. In six subsequent experiments (Study 2), we remove the confound in previous research by introducing emotions that differ in sociality rather than valence. In line with the social preference account, we show that apparent evidence for infrahumanization is better explained by ingroup preference and stereotyping. In a final experiment (Study 3), we provide further evidence for the social preference account by replicating the pattern of results observed in Study 2 in a minimal group design.

In Studies 1 and 2, we use the same six intergroup contexts. The precise social conditions necessary for infrahumanization have not been clearly established within the field and it has been noted that it may not always occur (e.g., Castano & Giner-Sorolla, 2006; Demoulin et al., 2009). However, a comprehensive review of prior empirical evidence suggests that outgroups are particularly likely to be infrahumanized if they threaten one's worldview, are disliked, and belong to a social category that one would not want to belong to (Leyens, 2009). Initial infrahumanization studies included students from the Canary Islands versus those from mainland Spain as the intergroup context (Leyens et al., 2000, 2001). The researchers noted general hostility between these groups, with each seeing the other as a ‘disliked’ outgroup, suggesting ‘outgroup derogation’ was likely (Leyens et al., 2001, pp. 396–399). Follow-up work included Spanish or Belgian ingroup members and North African individuals as outgroup members, at the time a ‘very stigmatized minority and low-status group in Belgium and in the Canary Islands’ (Paladino et al., 2002, p. 113). Most infrahumanization studies focus on social contexts that are similarly grounded in antag

Fig. 1. The average points at which outgroups and ingroups were marked on the blatant dehumanization scale across Studies 1, 2 and 3. All outgroups were significantly dehumanized relative to the corresponding ingroup (all ps < 0.001 in Studies 1 & 2, p = .002 in Study 3).

We chose our groups to maximise our chances of replicating infrahumanization effects if they occur (Leyens, 2009). The first outgroup we chose was Muslims (Christian ingroup) (Expts. 1a&2a). Dehumanization of religious outgroups, including of Muslims by Christians, has been widely reported (Banton et al., 2020; Kteily, Hodson, & Bruneau, 2016; Viki, Osgood, & Phillips, 2013) and discrimination against Muslims is a pressing social problem in many Western societies (Calfano, 2018; Hewstone & Schmid, 2014). The remaining outgroups were criminals (Expts. 1b&2b), child molesters (Expts. 1c&2c), anti-vaxxers (Expts. 1d&2d), people who do not adhere to social distancing regulations during the Covid-19 pandemic (‘non-social distancers’) (Expts. 1e&2e), and climate change deniers (Expts. 1f&2f). Prior work reports dehumanization of criminals and sex offenders (Bastian, Denson, & Haslam, 2013; Viki, Fullerton, Raggett, Tait, & Wiltshire, 2012). We introduced the three additional outgroup contexts (anti-vaxxers, non-social distancers and climate change deniers) based on current pertinence. At the time of data collection in early April 2020, the UK had been in full ‘lockdown’ for just over one week and tension between individuals who did and did not adhere to the guidelines was developing (Proser, Judge, Bolderdijk, Blackwood, & Kurz, 2020). Similarly, social division between those who are pro- and anti-vaccination has been particularly salient during the COVID-19 pandemic (Johnson et al., 2020). Rather than seeking to be exhaustive, the intergroup contexts we chose for Studies 1 and 2 illustrate the conceptual distinction between infrahumanization theory and our alternative social preference account. In Study 3, we replicated our results in a minimal group design. This allowed us to further demonstrate the generalisability of our results in a social context free from prior stereotypes and intergroup antagonism.

2. Data collection and availability

All experiments reported in this manuscript took place online and were created and administered using Qualtrics (https://www.qualtrics.com). Participants were recruited through Prolific (https://www.prolific.co) and a different sample was included for each experiment reported. Informed consent was obtained at the start of each session according to approved ethical procedures. Participants were compensated at an approximate rate of £7.50 per hour. All studies were pre-registered and the data is available open access. Links to pre-registration documents and raw data for each study can be found at: https://osf.io/rzb5n/

3. Pretest

One concern with prior work is that there are inconsistencies in how emotions are categorised; whether or not items are considered uniquely human changes between studies. For example, ‘enjoyment’ is considered uniquely human whereas ‘joy’ is not (Leyens et al., 2001; Paladino et al., 2002). ‘Happiness’ is sometimes considered uniquely human (Paladino et al., 2002) but sometimes not (Cortes et al., 2005). These problems may arise because previous studies have not rigorously pretested the emotion terms they used as stimuli as well as from translational discrepancies across studies conducted in different languages. In order to formally determine which emotions (in English) tend to be considered uniquely human and which tend to be considered shared with other
animals, we conducted a pretest in which we asked participants to rate fifty-four common emotion terms on three scales: human uniqueness, valence of experience and sociality.

3.1. Pretest methods

3.1.1. Participants
Sixty participants completed the ratings (22 female, 37 male, 1 ‘other’), aged between 18 and 54 (Mean age = 26.8, SD = 7.98). All participants were fluent in English. Eight people failed one or more attention checks and their data was excluded and replaced.

3.1.2. Scales
We chose fifty-four common emotion terms and asked participants to rate them on Humanness (the extent to which it is believed each emotion is experienced by humans compared to other species), Valence (the extent to which it is believed each emotion is positive or negative to experience) and Sociality (the extent to which it is believed each emotion is prosocial or antisocial), using three separate sliding scales. The full list of emotion terms, along with additional information about scale presentation, is in supplementary information.

3.1.3. Procedure
Participants were informed that the study would examine the ways in which people understand emotional terms and that they would be asked to rate emotion words on the three separate scales. Once informed consent was obtained, brief demographic and screening questions were asked. Then, participants were taken through the three question blocks. Participants were debriefed and redirected back to Prolific for payment. The session took approximately twelve minutes.

3.2. Pretest results and discussion

Our pretest confirmed that emotions differ both in valence and sociality. We present the mean ratings for each emotion on Humanness, Valence and Sociality in supplementary information, Table S1. The ‘basic’ (or ‘primary’) emotions such as fear, sadness, happiness and surprise featured among the emotions most thought of as shared with other species (Ekman, 1992). In line with infrahumanization theory, we largely replicated prior work from Demoulin et al. (2004), who also reported emotions such as nostalgia and optimism to be most uniquely human, and emotions such as fear and surprise to be least uniquely human. Importantly, however, none of the terms commonly included as negative secondary emotions in previous research, such as guilt, remorse, resignation and melancholy (Leyens et al., 2001; Paladino et al., 2002) were perceived to be antisocial, even though they were considered negative to experience.

Overall, there was a general positive association between mean scores across participants for valence and sociality. This suggests that, across a broad range of emotion terms, emotions that make us feel positive are also viewed as prosocial and emotions that make us feel negative may be viewed as more antisocial. However, for the specific negative emotions commonly included in prior infrahumanization work, participants’ scores on valence and sociality scales were not strongly (if at all) associated. For example, correlations were $r(58) = 0.131, p = .318$ for regret, $r(58) = 0.187, p = .153$ for melancholy, $r(58) = 0.262, p = .035$ for disillusion, and $r(58) = 0.060, p = .651$ for remorse. This shows that the kinds of negative emotions that infrahumanization researchers have included in previous research (e.g., regret, melancholy, disillusion, remorse - Banton et al., 2020; Leyens et al., 2001; Paladino et al., 2002; Vaes et al., 2003) may be negative to experience but are not antisocial and so are not best placed to determine whether infrahumanization is separate from intergroup preference. This provides strong grounds for re-examining the nature of intergroup bias in emotion attribution.

4. Study 1: Replicating previous research

In our first six experiments we sought to replicate previous research. Participants rated how strongly they believed ingroup and outgroup members to experience sixteen emotions. Four emotions were unique to humans and positive (nostalgia, optimism, humility, hope), four were unique to humans and negative, (disillusion, regret, melancholy, remorse) four were shared with other animals and positive, (happiness, tenderness, surprise, love) and four were shared with other animals and negative (fear, loneliness, sadness, nervousness).

4.1. Study 1 methods

4.1.1. Participants
A power analysis using MorePower 6.0.4 found a minimum N of 126 to be necessary to detect interactions with a medium effect size (partial eta squared 0.06) with an alpha of 0.05 and power of 0.8. 130 different participants completed the ratings in each experiment. Participants were only eligible if they were 18 or over, fluent in English and had not taken part in any of the other experiments reported. Data collection for each experiment took place completely separately. We excluded and replaced any participants that failed one or more of the attention checks.

In Experiment 1a (Muslim outgroup), participants could only take part if they identified as Christian. Seven people failed one or more attention check. Of the final sample, 95 participants were female and 35 were male, aged from 18 to 68 (Mean age = 35.3, SD = 14.07).

In Experiment 1b (criminal outgroup), participants could only take part if they had not served previous prison sentences. Five people failed one or more attention check. Of the final sample, 58 participants were female, 70 were male, 1 was nonbinary and 1 indicated ‘prefer not to say’. Ages ranged from 18 to 59 (Mean age = 27.1, SD = 7.94).

In Experiment 1c (child molester outgroup), four people failed one or more attention check. Of the final sample, 84 participants were female, 46 were male, and ages ranged from 18 to 57 (Mean age = 28.5, SD = 10.66).

In Experiment 1d (anti-vaxxer outgroup), participants could only take part if they were pro vaccination. Four people failed one or more attention check and three additional people were excluded and replaced because they indicated they were anti vaccination. Of the final sample, 53 participants were female, 75 were male and 2 were non-binary/ agender, aged from 18 to 60 (Mean age = 27.0, SD = 8.92).

In Experiment 1e (non-social distancer outgroup), participants could only take part if they were living in the UK and reported that they were following current social distancing regulations. Three people were excluded because they indicated that they were not adhering to social distancing regulations and their data was excluded and replaced. Of the final sample, 95 participants were female and 35 were male, aged from 18 to 68 (Mean age = 35.3, SD = 14.07).

In Experiment 1f (climate change deniers as outgroup), participants could only take part if they believed in climate change. Five people failed one or more attention check. Of the final sample, 52 participants were female, 78 were male and ages ranged from 18 to 62 (Mean age = 26.6, SD = 9.22).

4.1.2. Stimuli development
We chose emotions from our pretest data (supplementary information, Table S1) that best fit the four emotion categories of interest: unique to humans and positive, unique to humans and negative, shared with other animals and positive and shared with other animals and negative. Table 1 shows the list of emotion words included in the final stimulus sets. From the most and least uniquely human terms, we chose four rated as highly positive and four rated as highly negative. In developing the items for our emotion categories, we ensured that humanness ratings were closely matched between the positive and negative conditions for each level of humanness so that we could accurately separate valence effects from ones of humanness. See supplementary
Table 1
Emotion terms included for each condition in all experiments.

|                | Study 1                          | Study 2                          |
|----------------|----------------------------------|----------------------------------|
| Positive       | Nostalgia                        | Nostalgia                        |
| Negative       | Disillusion                      | Arrogance                        |
| Prosocial      | Humility                         | Contempt                         |
| Antisocial     | Sadness                          | Scared                           |
| love           | Love                             | Nervousness                      |
| tenderness     | Tenderness                       | Sadness                          |
| surprise       | Surprise                         | Tenderness                       |
| happiness      | Happiness                        | Loneliness                       |
| Optimism       | Optimism                         | Tenderness                       |
| Humility       | Humility                         | Irritation                       |
| hope           | Hope                             | Love                             |
| Infrahumans    | Child                            | Anger                            |
| All animals    | Infrahumanized                   | Universals                       |

For further details on stimuli development.

4.1.3. Scales

Participants indicated on unmarked sliders how strongly they thought the ingroup and outgroup in each experiment experienced the sixteen emotions from Not at all (0) to Very strongly (100), with the midpoint Somewhat (50). For example, in Experiment 1b, the outgroup block began ‘In the following questions, please consider the group: Individuals with criminal convictions’. Then, participants would respond to each item, such as ‘How strongly do you think a typical criminal feels nostalgia’. Ingroup and outgroup items were presented in two separate blocks shown on sequential screens, the order of which was counterbalanced across participants. The sixteen emotion items within each block were randomised and one attention check per block was also included approximately halfway through, such as ‘Please indicate Somewhat’.

Participants also completed the blatant dehumanization scale (Kteily, Bruneau, Waytz, & Cotterill, 2015) (Fig. 1) and a simple preference measure for both groups. In the blatant dehumanization scale, participants saw the ‘ascent of man’ image and were asked to indicate on an unmarked slider how evolved they considered the average member of each group to be, with 0 corresponding to the very bottom and 100 to the most human at the very top. In the attitude scale, participants were asked to indicate how they felt about each group using an unmarked sliding scale from Extremely Negative (0) to Extremely Positive (100). For all scales, half of the participants responded to ingroup items first and half to outgroup items first.

We included the group preference and blatant dehumanization measures to check that our chosen groups were the kinds that we should expect to see infrahumanized should the process occur. Prior work shows that infrahumanization measures correlate positively with blatant dehumanization scores (Kteily et al., 2015). Thus, though they are not claimed to measure the same construct, they have been shown to reliably co-occur. We included the attitude measure as confirmation that the outgroups were social categories that participants ‘would not like or want to belong to’ (Leyens, 2009), also increasing chances of detecting infrahumanization if it occurs.

4.1.4. Procedure

Participants were informed that the study was designed to help us understand the ways in which people ascribe emotions to different groups of individuals and stated the particular groups of interest for each experiment. Participants were instructed that they would be asked to rate sixteen emotion words on two scales, one for each social category, and then complete two scales asking about attitudes to each group. Once informed consent was obtained, brief demographic and screening (if relevant) questions were asked. Then, participants were taken through the two experimental blocks. Following this, participants completed the group preference and then the blatant dehumanization scales. Lastly, participants were debriefed and redirected back to Prolific for payment. On average, the sessions took under ten minutes.

4.1.5. Design and data analysis

In line with our pre-registered analysis plan, we conducted 2 (Group: ingroup/outgroup) x 2 (Valence: positive/negative) x 2 (Humanness: unique to humans / shared with other animals) within subjects ANOVAs to test for infrahumanization in intergroup emotion attributions. Scores for each emotion category were obtained by calculating the mean of the four emotion terms within the category for each participant. For example, a participant’s score for uniquely human positive emotion ascriptions towards the ingroup would be the mean of their ratings on Nostalgia, Optimism, Humility and Hope within the ingroup block. More detail about the design is available in supplementary information.

In this design, infrahumanization would be observed in an interaction whereby uniquely human emotions are more strongly ascribed to the ingroup, independent of valence (Leyens et al., 2000). This should not be the case for emotions shared with other animals, for which previous work found the reverse or no difference (Leyens et al., 2001). For example, in original experiments demonstrating infrahumanization, Leyens et al. (2001) showed that more positive and negative uniquely human emotions were attributed to the ingroup than the outgroup and that this was not qualified by an interaction with valence (see Leyens et al., 2001, p. 402, Fig. 2).

Though previous studies do not find interactions between intergroup emotion attributions and valence, there tend to be main effects of valence such that ascriptions of positive emotions are generally higher than negative ones. These results are not central to the predictions of infrahumanization theory nor the social preference account.

In following up significant interactions, we report only comparisons between ingroup and outgroup ratings for each condition, in line with testing the main hypotheses. We measured differences in ratings for ingroup and outgroup on the attitude and ‘blatant dehumanization’ scales using paired-samples t-tests. All tests were two-sided and met the assumptions necessary for our statistical approaches.

4.2. Study 1 results

4.2.1. Blatant dehumanization and attitude scores

In every experiment, the outgroup was rated as significantly less human than the ingroup on the blatant dehumanization scale (all $p < 0.001$). Additionally, participants reported feeling significantly more negative towards the outgroup than the ingroup on the attitude scale (all $p < 0.001$). Fig. 1 shows the points at which outgroups and ingroups were marked on the blatant dehumanization scale. The extent to which outgroups were ‘blatantly dehumanized’ varied greatly across our intergroup contexts. The average point at which Muslims and ‘criminals’ were marked fell between the most ‘evolved’ looking human silhouette and the more caveman-like silhouette next to it on the scale. ‘Child molesters’ and ‘non-social distancers’ were rated much further down on the ascent scale, nearer to the midway point between the ape-like and modern human-like depictions. Fig. S1 (supplementary information) shows mean results for each ingroup and outgroup on the dehumanization and attitude measures.

4.2.2. Intergroup emotion ascriptions

4.2.2.1. Experiment 1a. For ratings towards Muslims (outgroup) and Christians (ingroup), there were significant main effects of group, F(1, 129) = 32.54, $p < .001$, $\eta^2 = 0.201$, and of valence, F(1, 129) = 101.80, $p < .001$, $\eta^2 = 0.441$, with ratings higher overall for ingroup than outgroup and for positive than negative emotions. There was no significant main effect of humanness, F(1, 129) = 0.76, $p = .384$, $\eta^2 = 0.006$. There were significant interactions between group and humanness, F(1, 129) = 11.77, $p = .001$, $\eta^2 = 0.084$, group and valence, F(1, 129) = 17.16, $p < .001$, $\eta^2 = 0.117$, and humanness and valence, F(1, 129) = 32.54, $p < .001$, $\eta^2 = 0.201$. Pairwise comparisons showed that overall,
ratings were higher for ingroup than outgroup both for uniquely human emotions (p < .001) and for emotions shared with other animals (p = .001). Ratings were overall higher for ingroup than outgroup for positive emotions (p < .001), but there were no differences between groups for negative emotions (p = .463).

All effects were qualified in a significant three-way interaction, F(1, 129) = 11.37, p = .001, \( \eta^2_p = 0.081 \). Planned comparisons showed that ratings were higher for ingroup than outgroup for positive uniquely human terms (p < .001), negative uniquely human terms (p = .006), and positive terms shared with other animals (p < .001). However, there was no difference between ingroup and outgroup for negative terms shared with other animals (p = .104).

4.2.2.2. Experiment 1b. For ratings towards Criminals (outgroup) and ‘individuals with no criminal history’ (ingroup), there were significant main effects of group, F(1, 129) = 31.84, p < .001, \( \eta^2_p = 0.198 \), valence, F(1, 129) = 4.64, p = .033, \( \eta^2_p = 0.035 \), and humanness, F(1, 129) = 35.86, p < .001, \( \eta^2_p = 0.218 \). Ratings were higher overall for ingroup than outgroup, for negative than positive emotions, and for emotions shared with other animals than for uniquely human emotions.

There were significant interactions between group and humanness, F(1, 129) = 7.62, p = .007, \( \eta^2_p = 0.056 \), group and valence, F(1, 129) = 167.70, p < .001, \( \eta^2_p = 0.565 \), and humanness and valence, F(1, 129) = 4.16, p = .043, \( \eta^2_p = 0.031 \). Pairwise comparisons showed that overall, ratings were higher for ingroup than outgroup both for uniquely human emotions and for emotions shared with other animals (ps < .001). Ratings were overall higher for ingroup than outgroup for positive emotions (p < .001), and higher for outgroup than ingroup for negative emotions (p = .006).

All effects were qualified in a significant three-way interaction, F(1, 129) = 4.56, p = .035, \( \eta^2_p = 0.034 \). Planned comparisons showed that ratings were higher for ingroup than outgroup for positive terms, both uniquely human and shared with other animals (ps < .001), but higher for outgroup than ingroup on negative terms, both uniquely human (p = .007) and shared with other animals (p = .020).

4.2.2.3. Experiment 1c. For ratings towards ‘child molesters’ (outgroup) and ‘individuals with no criminal history’ (ingroup), there were significant main effects of group, F(1, 129) = 154.31, p < .001, \( \eta^2_p = 0.545 \), and of humanness, F(1, 129) = 83.51, p < .001, \( \eta^2_p = 0.393 \), but not of valence, F(1, 129) = 2.97, p = .087, \( \eta^2_p = 0.023 \). Ratings were higher overall for ingroup than outgroup and for emotions shared with other animals than for uniquely human emotions.

There were significant interactions between group and humanness, F(1, 129) = 17.70, p < .001, \( \eta^2_p = 0.121 \), group and valence, F(1, 129) = 119.32, p < .001, \( \eta^2_p = 0.481 \), and humanness and valence, F(1, 129) = 28.98, p < .001, \( \eta^2_p = 0.189 \). Pairwise comparisons showed that overall, ratings were higher for ingroup than outgroup both for uniquely human emotions and for emotions shared with other animals (ps < .001). Ratings were also overall higher for ingroup than outgroup for positive
emotions (p < .001), and for negative emotions (p = .001).

All effects were qualified in a significant three-way interaction, F(1, 129) = 14.10, p < .001, η²p = 0.099. Planned comparisons showed that ratings were higher for ingroup than outgroup for positively uniquely human terms (p < .001), negative uniquely human terms (p < .001), and positive terms shared with other animals (p < .001). However, there was no difference between ingroup and outgroup for negative terms shared with other animals (p = .287).

4.2.2.4. Experiment 1d. For ratings towards ‘anti-vaxxers’ (outgroup) and ‘pro-vaxxers’ (ingroup), there were significant main effects of group, F(1, 129) = 40.42, p < .001, η²p = 0.239, and of valence, F(1, 129) = 69.59, p < .001, η²p = 0.350, but not of group, F(1, 129) = 1.02, p = .315, η²p = 0.008. Ratings were higher overall emotions shared with other animals than for uniquely human emotions and for positive than negative emotions.

There were significant interactions between Group and Valence, F(1, 129) = 88.99, p < .001, η²p = 0.408, Group and Humaneness, F(1, 129) = 11.49, p = .001, η²p = 0.082, and Valence and Humaneness, F(1, 129) = 8.41, p = .004, η²p = 0.061. Pairwise comparisons showed that overall, ratings were higher for ingroup than outgroup for uniquely human emotions (p = .017) but not for emotions shared with other animals (p < .358). Ratings were overall higher for ingroup than outgroup for positive emotions (p < .001) but higher for ingroup than outgroup for negative emotions (p = .001). The three-way interaction was not significant, F(1, 129) = 0.31, p = .580, η²p = 0.002.

4.2.2.5. Experiment 1e. For ratings towards ‘non-social distancers’ (outgroup) and ‘social distancers’ (ingroup), there were significant main effects of group, F(1, 129) = 239.50, p < .001, η²p = 0.650, and of humaneness, F(1, 129) = 60.13, p < .001, η²p = 0.318, but not of valence, F(1, 129) = 1.75, p = .188, η²p = 0.013. Ratings were higher overall for ingroup than outgroup and for emotions shared with other animals than for uniquely human emotions.

There were significant interactions between Group and Humaneness, F(1, 129) = 38.46, p < .001, η²p = 0.230, group and valence, F(1, 129) = 16.08, p < .001, η²p = 0.111, and humaneness and valence, F(1, 129) = 113.12, p < .001, η²p = 0.467. Pairwise comparisons showed that overall, ratings were higher for ingroup than outgroup both for uniquely human emotions and for emotions shared with other animals (ps < .001). Ratings were also overall higher for ingroup than outgroup for positive emotions (p < .001), and for negative emotions (p = .001).

All effects were qualified in a significant three-way interaction, F(1, 129) = 149.18, p < .001, η²p = 0.536. Planned comparisons showed that ratings were higher for ingroup than outgroup for all four emotion conditions - positive uniquely human terms, negative uniquely human terms, positive terms shared with other animals and negative terms shared with other animals (all ps < .001).

4.2.2.6. Experiment 1f. For ‘climate change deniers’ (outgroup) and ‘climate change believers’ (ingroup), there were significant main effects of group, F(1, 129) = 171.51, p < .001, η²p = 0.571, and of humaneness, F(1, 129) = 27.79, p < .001, η²p = 0.177, but not of valence, F(1, 129) = 0.85, p = .395, η²p = 0.007. Ratings were higher overall for ingroup than outgroup and for emotions shared with other animals than for uniquely human emotions.

There were significant interactions between group and humaneness, F(1, 129) = 3.92, p = .05, η²p = 0.029, group and valence, F(1, 129) = 38.99, p < .001, η²p = 0.232, and humaneness and valence, F(1, 129) = 17.02, p < .001, η²p = 0.117. Pairwise comparisons showed that overall, ratings were higher for ingroup than outgroup both for uniquely human emotions and for emotions shared with other animals (ps < .001). Ratings were also overall higher for ingroup than outgroup for positive emotions and for negative emotions (ps = .001). The three-way interaction was not significant, F(1, 129) = 0.13, p = .724, η²p = 0.001.

Mean scores (M) and standard errors of the mean (SE) for each of the conditions across all experiments in Study 1 are shown in supplementary information, Table S2. Fig. 2 shows results from Study 1.

4.3. Study 1 discussion

Results partially replicated the predictions of infrahumanization theory - some outgroups were rated overall as experiencing both positive and negative uniquely human emotions to a lesser extent than the ingroup. However, these outgroups were also rated as experiencing emotions shared with other animals to a lesser extent than the ingroup. In each experiment, the outgroup was rated as ‘less human’ on the blatant dehumanization scale, confirming that these were the types of intergroup contexts in which we ought to see infrahumanization effects should the process occur. We next examined whether controlling for the sociality of emotional terms explained apparent evidence for infrahumanization.

5. Study 2: Testing the social preference account

In our next six experiments, we test whether what appears to be infrahumanization may be explained by ingroup preference and stereotyping. Rather than comparing intergroup ascriptions of emotions that varied by how positive or negative they are to experience, we compared ascriptions of emotions than varied by how prosocial or antisocial they are in character. Participants rated the same six groups on four types of emotional experience: unique to humans and prosocial (nostalgia, optimism, humility, hope), unique to humans and antisocial (arrogance, schadenfreude, contempt, scorn), shared with other animals and prosocial (happiness, tenderness, love, friendliness) and shared with other animals and antisocial (hostility, irritation, anger, disgust).

This design pits the predictions of infrahumanization against a social preference account. Infrahumanization would be observed in an interaction between Group and Humaneness such that uniquely human emotions will be more strongly ascribed to ingroup than outgroup, both for prosocial and antisocial emotions (i.e., Leyens et al., 2001, p. 402, Fig. 2). In contrast, we hypothesised that prosocial emotions will typically be assigned more strongly to the ingroup and antisocial ones to the outgroup, regardless of humaneness (Fig. 3 shows both predictions).

5.1. Study 2 methods

5.1.1. Participants

Based on the same power analysis as reported for Study 1, 130 different participants completed the ratings in each experiment. The same eligibility criteria were applied as for Study 1.

For Experiment 2a (Muslim outgroup), participants could only take part if they identified as Christian. Six people failed one or more attention check. Of the final sample, 86 participants were female, 42 male and 2 were non-binary/agency, aged from 18 to 71 (Mean age = 33.6, SD = 11.79).

For Experiment 2b (criminal outgroup), participants could only take part if they had not served previous prison sentences. One person failed one or more attention check. Of the final sample, 62 participants were female, 68 were male and ages ranged from 18 to 65 (Mean age = 26.9, SD = 8.65).

For Experiment 2c (child molester group), three people failed one or more attention check. Of the final sample, 87 were female, 42 male and 1 non-binary, with an age range of 18 to 61 (Mean age = 31.6, SD = 10.14).

In Experiment 2d (anti-vaxxer outgroup), participants could only take part if they were pro vaccination. Seven people failed one or more attention check and four additional people were excluded and replaced because they indicated that they were anti vaccination. Of the final sample, 50 were female and 80 were male, aged from 18 to 51 (Mean age = 25.9, SD = 8.01).
In Experiment 2e (non-social distancers outgroup), participants could only take part if they were living in the UK and following social distancing regulations. Data was excluded and replaced for three participants who failed one or more attention check and one additional participant who indicated that they were not adhering to social distancing regulations. Of the final sample, 86 were female, 42 were male and 2 were non-binary/agender, aged from 18 to 71 (Mean age = 33.6, SD = 11.79).

For Experiment 2f (climate change deniers outgroup), participants could only take part if they believed in climate change. Two people failed one or more attention check. Of the final sample, 53 were female, 76 male and 1 non-binary, with an age range of 18 to 60 (Mean age = 27.3, SD = 8.36).

5.1.2. Stimuli development

We chose emotions from our pretest data (supplementary information, Table S1) that best fit the four emotion categories of interest: unique to humans and prosocial, unique to humans and antisocial, shared with other animals and prosocial and shared with other animals and antisocial. From the most and least uniquely human terms, we chose four rated as highly prosocial and four rated as highly antisocial, this time ignoring valence ratings. Table 1 shows the full list of emotion words. We chose the emotions such that humanness ratings were closely matched between the prosocial and antisocial conditions at each level of humanness. This was so that dimensions of Sociality and Humanness were orthogonal, allowing us to accurately separate effects of each. See supplementary information for further details on stimuli development.

5.1.3. Scales

We employed the same six intergroup contexts as for Study 1. Apart from including different emotion items, the emotion attribution scales were identical as to those described for Study 1. Participants again completed the blatant dehumanization scale (Kteily et al., 2015) (Fig. 1) and the group preference scale for the ingroup and outgroup in each experiment.

5.1.4. Procedure, design and data analysis

The procedure was the same as outlined for Study 1. The design and data analysis were almost identical as described for Study 1 though with the Sociality (prosocial/antisocial) variable instead of the Valence (positive/negative) variable.

5.2. Study 2 results

5.2.1. Blatant dehumanization and attitude scores

The outgroup was always rated as significantly less human than the ingroup on the blatant dehumanization scale (all ps < 0.001). Additionally, participants reported feeling significantly more negative towards the outgroup than the ingroup on the attitude scale (all ps < 0.001) (Fig. 1). Fig. S1 (supplementary information) shows mean results for each ingroup and outgroup on the dehumanization and attitude measures.

5.2.2. Intergroup emotion ascription ratings

5.2.2.1. Experiment 2a. For ratings towards Muslims (outgroup) and Christians (ingroup), there were main effects of humanness, $F(1, 129) = 28.75, p < .001, \eta^2 = 0.182$, and of sociality, $F(1, 129) = 147.39, p < .001, \eta^2 = 0.533$, but not of group, $F(1, 129) = 0.42, p = .517, \eta^2 = 0.003$. Ratings were higher overall for emotions shared with other animals than for uniquely human emotions, and for prosocial than antisocial emotions.

Fig. 3. Evidence for social preferences but not infrahumanization. Contrary to infrahumanization theory, ratings were higher for the ingroup than the outgroup for prosocial emotions but higher for the outgroup than the ingroup for antisocial emotions across all group contexts. Note that while main effects of valence have been reported in some prior work (see Leyens et al., 2001, p. 402, Fig. 2), we do not plot these in our predictions, as they are not relevant for distinguishing between the theories. Error bars represent standard errors.
There was a significant interaction between group and sociality, $F(1, 129) = 39.45, p < .001, \eta^2_p = 0.234$, but not between group and humanness, $F(1, 129) = 0.75, p = .389, \eta^2_p = 0.006$, nor between humanness and sociality, $F(1, 129) = 1.74, p = .190, \eta^2_p = 0.013$. Pairwise comparisons showed that overall, ratings were higher for ingroup than outgroup for prosocial emotions, but higher for outgroup than ingroup for antisocial emotions ($ps < .001$).

All effects were qualified in a significant three-way interaction $F(1, 129) = 3.97, p = .048, \eta^2_p = 0.030$. Planned comparisons showed that ratings were higher for ingroup than outgroup on prosocial terms, both uniquely human and shared with other animals, and for ingroup than outgroup.

There was a significant interaction between group and sociality, $F(1, 129) = 234.42, p < .001, \eta^2_p = 0.645$, and between group and humanness, $F(1, 129) = 21.82, p < .001, \eta^2_p = 0.145$. Pairwise comparisons showed that overall, ratings were higher for ingroup than outgroup for prosocial emotions, but higher for outgroup than ingroup for antisocial emotions ($ps < .001$). Though the interaction between group and humanness was significant, this did not reflect the infrahumanization prediction. Ratings were overall higher for ingroup than outgroup both for uniquely human emotions and for emotions shared with other animals ($ps < .001$). Importantly, ratings were higher for ingroup than outgroup both for uniquely human and shared with other animals ($ps < .001$) and for emotions shared with other animals ($p = .004$). The three-way interaction was not significant, $F(1, 129) = 0.28, p = .600, \eta^2_p = 0.002$.

There was a significant interaction between group and sociality, $F(1, 129) = 213.36, p < .001, \eta^2_p = 0.623$, and group, $F(1, 129) = 59.99, p < .001, \eta^2_p = 0.306$, and between group and humanness, $F(1, 129) = 56.59, p < .001, \eta^2_p = 0.305$. Pairwise comparisons showed that overall, ratings were higher for ingroup than outgroup for prosocial emotions, but higher for outgroup than ingroup for antisocial emotions ($ps < .001$). Ratings were also higher for outgroup than ingroup for uniquely human emotions, but higher for ingroup than outgroup on emotions shared with other animals ($ps < .001$).

All effects were qualified in a significant three-way interaction $F(1, 129) = 37.90, p < .001, \eta^2_p = 0.227$. Planned comparisons showed that ratings were higher for ingroup than outgroup on prosocial terms, both uniquely human and shared with other animals and higher for outgroup than ingroup on antisocial terms both uniquely human and shared with other animals ($ps < .001$).

There was a significant interaction between group and sociality, $F(1, 129) = 102.37, p < .001, \eta^2_p = 0.442$, and group, $F(1, 129) = 6.32, p = .013, \eta^2_p = 0.047$, and a marginal effect of sociality, $F(1, 129) = 3.85, p = .052, \eta^2_p = 0.029$. Ratings were higher overall for emotions shared with other animals than for uniquely human emotions, for ingroup than outgroup, and for antisocial than prosocial emotions. Ratings were higher also for ingroup than outgroup for uniquely human emotions, but higher for outgroup than ingroup on emotions shared with other animals ($ps < .001$).

All effects were qualified in a significant three-way interaction $F(1, 129) = 14.04, p < .001, \eta^2_p = 0.098$. Planned comparisons showed that ratings were higher for ingroup than outgroup on prosocial terms, both uniquely human ($p < .001$) and shared with other animals ($p < .001$), and higher for outgroup than ingroup uniquely human antisocial terms ($p < .001$). However, there was no difference between ingroup and outgroup on antisocial terms shared with other animals ($p = .200$).

Mean scores (M) and standard errors of the mean (SE) for each of the conditions across all experiments in Study 2 are shown in supplementary information, Table S3. Fig. 3 shows results for Study 2.
5.3. Study 2 discussion

Contrary to infrahumanization theory, outgroups were not denied uniquely human emotions relative to ingroups. All outgroups were thought to experience prosocial emotions to a lesser extent than ingroups, both for uniquely human emotions and for those shared with other animals. However, all outgroups were also thought to experience uniquely human antisocial emotions to a greater extent than ingroup members. Muslims, criminals, child molesters, anti-vaxxers and non-social distance were rated as experiencing antisocial emotions shared with other animals to a greater extent than the ingroup. However, there was no difference between climate change deniers and the ingroup for this condition. This may be because in this context, it is reasonable to suppose believers in climate change experience substantial levels of emotions such as anger and irritation. This highlights the importance of social context and stereotyping as well as ingroup preferences in explaining emotion attribution. Study 3 employed a minimal groups design in order to measure similar effects in the absence of learned stereotypes and historical negative feeling.

6. Study 3: Testing the social preference account with minimal groups

In our final experiment, we aim to replicate findings from Study 2 within a minimal groups design. Though we chose our six outgroup exemplars for Studies 1 and 2 to maximise our chances of detecting infrahumanization should it occur (based on past empirical work and also following suggestions from Leyens, 2009), it remains possible that learned stereotypes and intergroup antagonism may have weighted responses towards reflecting social preferences as opposed to subtle dehumanization. By using a minimal groups design, we were able to ensure we tested between the two competing hypotheses in the absence of these additional factors. Prior work has reported infrahumanization effects in minimal groups (Demoulin et al., 2009; Simon & Gutsell, 2020), meaning this approach was methodologically appropriate for comparing the two theories.

Participants were first allocated to novel groups using a dot estimation task (Diehl, 1990; Ratner & Amadio, 2013; Tajfel, 1970). Then, participants rated the novel ingroup and novel outgroup on emotional experiences as described for Study 2. The study design and predictions were identical as for Study 2. Infrahumanization would be observed in an interaction between group and humanness such that uniquely human emotions are more strongly ascribed to ingroup than outgroup, both for prosocial and antisocial emotions. However, we again hypothesised that prosocial human emotions would typically be attributed more strongly to the ingroup and antisocial ones to the outgroup.

6.1. Study 3 methods

6.1.1. Participants

Based on the same power analysis as reported for Studies 1 and 2, 130 different participants completed the ratings in each experiment. Participants were eligible to take part if they were over 18, fluent in English, and had not taken part in any of the other experiments reported presently. Nine people failed one or more attention check and their data was excluded and replaced. Of the final sample, 56 participants were female and 74 were male, aged from 18 to 57 (Mean age = 26.8, SD = 9.10).

6.1.2. Minimal group paradigm

Participants were assigned to novel groups based on a classic dot estimation task (e.g., Diehl, 1990; Ratner & Amadio, 2013; Tajfel, 1970). Participants were told that the purpose of the study was to help us understand the ways in which people ascribe emotions to different groups of individuals and that they would first perform a simple numerical estimation task to identify which group (out of two) they belonged to. Instead of the common categories of ‘over-estimators’ and ‘under-estimators’, we used the terms ‘spatial-estimators’ and ‘object-estimators’. This was because the emotion attribution task relied on a form of estimation (of emotional experience) and we wanted to ensure the group labels did not interact with later emotion judgments. Before the task, participants were told that people can be categorised as taking an object-centred approach to estimation (‘object-estimators’) or a spatial-centred approach to estimation (‘spatial-estimators’) and that individual tendencies for the two styles are equally distributed in the population.

In the dot estimation task, participants saw eleven images of random patterns of dots each on the screen for 1 s. After each image, participants had to enter the number of dots they believed they had seen before the next image appeared. The task and stimuli were based on an Inquisit script from Millisecond (https://www.millisecond.com) and adapted for presentation on Qualtrics.

Following the task, half of the participants were told they had been classified as a spatial-estimator and the other half were told they had been classified as an object-estimator. This procedure met the key criteria for a minimal group paradigm (Tajfel, 1970; Tajfel, Billig, Bundy, Flament, 1971).

6.1.3. Stimuli, scales, procedure, design and data analysis

The emotion stimuli, scale presentation, procedure, design and planned data analysis were all exactly the same as outlined for Study 2.

6.2. Study 3 results

6.2.1. Blatant dehumanization and attitude scores

The novel ingroup was rated as significantly less human than the novel outgroup on the blatant dehumanization scale (p = .002) (Fig. 1). However the novel outgroup was still rated closest to the silhouette reminiscent of a modern human on the scale, and was not ‘blatantly dehumanized’ to the extent that outgroups in Studies 1 and 2 were. Participants reported feeling significantly more negative towards the novel outgroup than the novel ingroup on the attitude scale (p < .001). Fig. S1 (supplementary information) shows mean results for the ingroup and outgroup on the blatant dehumanization and group preference measures.

6.2.2. Intergroup emotion ascription ratings

For ratings towards novel ingroup and outgroup members (minimal group design), there were main effects of group, F(1, 129) = 7.58, p = .007, ηp² = 0.055, of humanness, F(1, 129) = 32.93, p < .001, ηp² = 0.203, and of sociality, F(1, 129) = 99.74, p < .001, ηp² = 0.436. Ratings were higher overall for ingroup than outgroup, for emotions shared with other animals than for uniquely human emotions, and for prosocial than antisocial emotions.

There was a significant interaction between group and sociality, F(1, 129) = 22.45, p < .001, ηp² = 0.148, but not between group and humanness, F(1, 129) = 1.84, p = .177, ηp² = 0.014, nor between humanness and sociality, F(1, 129) = 0.15, p = .704, ηp² = 0.001. Pairwise comparisons showed that overall, ratings were higher for ingroup than outgroup for prosocial emotions (p < .001), but higher for outgroup than ingroup for antisocial emotions (p = .007).

All effects were qualified in a significant three-way interaction F(1, 129) = 13.25, p < .001, ηp² = 0.093. Planned comparisons showed that ratings were higher for ingroup than outgroup on prosocial terms, both uniquely human and shared with other animals (ps < .001). Ratings were higher for outgroup than ingroup on uniquely human antisocial terms (p < .001), but for antisocial terms shared with other animals there was no difference between ingroup and outgroup (p = .637).

Mean scores (M) and standard errors of the mean (SE) for each of the conditions in Study 3 are shown in, supplementary information, Table S4. Fig. 4 shows results for Study 3.
6.3. Study 3 discussion

Contrary to the predictions of infrahumanization theory, novel outgroup members were not denied uniquely human emotions relative to novel ingroup members. There was no significant interaction between group and humanness but there was an interaction between group and sociality. Outgroup members were thought to experience prosocial uniquely human emotions to a lesser extent than ingroup members, but antisocial uniquely human emotions to a greater extent than ingroup members. This finding shows that even in a novel group context free from learned stereotypes and antagonism, our social preference account better explains intergroup biases in emotion attribution than infrahumanization theory.

7. General discussion

We found no convincing evidence for infrahumanization. In our first set of studies we broadly replicated previously reported effects (Study 1) showing our paradigm was well placed to detect infrahumanization if it occurs. Our subsequent results suggest that, in the seven intergroup contexts we employed, what appeared to be evidence for infrahumanization can be better explained by social preference (Studies 2 and 3). When emotion terms varied on sociality rather than on valence, people did not ‘subtly dehumanize’ the outgroups we included by denying them uniquely human emotions. Rather, they attributed prosocial emotions more strongly to ingroup members and antisocial emotions more strongly to outgroup members, regardless of humanness. This accords with recent critiques of the social psychological literature on dehumanization more generally (Bloom, 2017; Lang, 2010, 2020; Manne, 2016, 2018; Over, 2020, 2021; Smith, 2014, 2016) and offers an important conceptual development to our understanding of intergroup bias in emotion judgments.

In practical terms, more accurately characterising the ways in which the emotions of different groups are perceived has important implications for real world settings such as criminal justice, in which certain defendants, for example those perceived as belonging to a religious outgroup, might be unfairly viewed as possessing lower levels of remorse but also higher levels of contempt as a result of their group membership. One of the main reasons why infrahumanization theory has been influential in intergroup relations research is because it has been causally linked to negative behavioural consequences. For example, previous research has suggested that infrahumanizing outgroups reduces prosocial behaviour towards them (Cuddy et al., 2007; Vaes et al., 2003; Vaes, Paladin, & Leyens, 2002). In light of the present findings, future research would benefit from revisiting previously-reported links between biases in emotion attribution and prosocial and antisocial behaviours.

Our results dovetail with recent empirical work that challenges the predictions made by Haslam’s (2006) dual model of dehumanization (Enock et al., 2021). This research showed that when undesirable human-specific characteristics (such as ‘corrupt’ and ‘selfish’) are included in overall measures of humanness, there is no evidence for either animalistic or mechanistic dehumanization of outgroups as characterised by the dual model. Rather, desirable human qualities are more strongly attributed to ingroup members and undesirable human qualities to outgroup members. The present work extends these findings by further demonstrating the importance of considering sociality confounds when measuring psychological processes of ‘dehumanization’, this time through another highly prominent framework within the field.

During the review process, it was put to us that because dimensions of valence and sociality correlate highly in our pretest, the two constructs are ‘indistinguishable’, thus rendering our critique obsolete. We believe this represents a misunderstanding. Height and weight are strongly positively correlated, yet they are distinct constructs. Similarly,
even though emotions that are generally perceived as prosocial may also perceived as positive to experience, and emotions that are generally perceived as antisocial may also be perceived as negative to experience, the two constructs are clearly conceptually distinct. While sadness is negative to experience, it is not inherently antisocial in character. Schadenfreude on the other hand is, by definition, positive to experience but antisocial in character. Many research findings converge on the view that while ‘regret’ or ‘remorse’ are negative to feel they are not unkind in character (Stearns & Parrott, 2012, see also Parkinson, 1996; Vaish & Hepach, 2020; van Kleef et al., 2016). Our argument is that the kinds of negative emotions that proponents of infrahumanization theory have included in previous research, such as disillusion, regret, melancholy and remorse (e.g., Banton et al., 2020; Leyens et al., 2001; Paladino et al., 2002; Vaes et al., 2003) are negative to experience but are not antisocial and so are not best placed to determine whether infrahumanization is separate from intergroup preference.

While our results offer an important and novel empirical critique of prior work on infrahumanization, we acknowledge that we only tested seven intergroup contexts. Our social contexts varied in animosity, with some containing high prior animosity (criminals and child molesters) and some containing little or no prior animosity (minimal groups). Despite this, without testing many more groups (such as those based on nationality, race or gender), it remains a possibility that infrahumanization could sometimes occur even when sociality of emotion is controlled. However, participants explicitly dehumanized all seven outgroups relative to the ingroup on the blatant dehumanization scale, suggesting we would likely observe infrahumanization if it occurs (Kteily et al., 2015). Further, the groups we chose exemplify the criteria for infrahumanization proposed in prior work (Leyens, 2009).

Our results demonstrate both ingroup favouritism (assigning greater prosocial feeling to the ingroup) and outgroup derogation (assigning greater antisocial feeling to the outgroup) (Brewer, 1999; Hewstone et al., 2002). However, we also note that group specific stereotypes and particular social contexts are likely to play an important role in these processes (Fiske et al., 2002). For example, it is likely that group status may affect the specific emotions that are ascribed to group members. Emotions such as ‘contempt’ and ‘schadenfreude’ are included as exemplars of antisocial uniquely human emotions in the present work, but it may not be the case that outgroup members perceived as ‘lower status’ such as homeless people would be attributed these antisocial emotions to a greater extent than the ingroup because these particular emotions imply a position of relatively high status. There may be other uniquely human yet antisocial emotions that a ‘lower status’ outgroup may be more likely to be perceived as experiencing, such as bitterness or envy. Similarly, it is possible that groups such as ‘immigrants’ could be perceived as feeling high levels of optimism or nostalgia by nature of their situation. Our goal in this research was not to explore the many nuances of emotion attribution in intergroup contexts but rather more modest in scope, we aimed to show that apparent evidence for infrahumanization may obscure more than it reveals about intergroup bias.

**Author contributions**

FE, HO and ST conceptualised and designed the experiments. FE collected and analysed the data. FE and HO wrote the original draft. All authors edited and reviewed the manuscript.

**Declaration of Competing Interest**

The authors declare no competing interests.

**Acknowledgments**

This research was supported by the European Research Council under the European Union’s Horizon 2020 Programme, grant number ERC-STG- 755719 awarded to HO.

**Appendix A. Supplementary data**

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2021.104865.

**References**

Appiah, K. A. (2008). Experiments in ethics. Harvard University Press.
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.
Bastian, B., Denmon, T. F., & Haslam, N. (2013). The roles of dehumanization and moral outrage in retributive justice. PLoS One, 8(4).
Bloom, P. (2017). The root of all cruelty? The New Yorker. https://www.newyorker.com/magazine/2017/11/27/the-root-of-all-cruelty.
Boccato, G., Cortes, B. P., Demoulin, S., & Leyens, J.-P. (2007). The automaticity of infrahumanization. European Journal of Social Psychology, 37(5), 987–999.
Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? Journal of Social Issues, 55(3), 429-444.
Brown, R., Eller, A., Leeds, S., & Stace, K. (2007). Intergroup contact and intergroup attitudes: A longitudinal study. European Journal of Social Psychology, 37(4), 692-705.
Calfano, B. (2018). Muslims, identity, and American politics. Routledge.

Nevertheless, the field would benefit from careful empirical research rigorously controlling for emotion sociality in more implicit contexts. Additionally, we acknowledge that, following the majority of prior work on infrahumanization, we conceptualised humanity and valence/sociality dichotomously and as such tested our predictions with relatively few exemplars from each category. Though this approach is standard in the field, future work may benefit from testing similar hypotheses but treating humanity and sociality as continuous predictors (see Castano & Giner-Sorolla, 2006).

These possibilities do not detract from the central importance of our critique. More generally, our results illustrate the importance of considering the sociality of emotion terms employed as stimuli, a methodological advancement that will be crucial to incorporate in any future studies of emotion attribution in other intergroup contexts. To accurately test for ‘infrahumanisation’, future research must consider the central role of emotion sociality as separate from emotional experience. Prior work has reported infrahumanization to be extremely widespread in society and prevalent across a multitude of intergroup divides (Banton et al., 2020; Cortes et al., 2005; Cuddy et al., 2007; Gaunt, 2009; Leyens et al., 2000, 2001; Rodríguez-Pérez et al., 2011; Simon & Gutsell, 2020; Vaes et al., 2002, 2003). Rigorous measurement, tighter experimental control and more careful consideration of social context may change some or all of the conclusions from previous research.

If psychological research is to effectively inform intervention to improve intergroup relations, it is essential it accurately characterises the underlying mechanisms of intergroup bias. Our findings suggest the construct of infrahumanization may obscure more than it reveals about intergroup bias.
