UNIVERSITY OF LEEDS

This is a repository copy of *Are upside-down faces perceived as “less human”?*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/205173/

Version: Accepted Version

Article:
Eggleston, A. orcid.org/0000-0003-4123-3225, Cook, R. and Over, H. (Cover date: 2023) Are upside-down faces perceived as “less human”? Journal of Experimental Psychology: Human Perception and Performance, 49 (12). pp. 1503-1517. ISSN 0096-1523

https://doi.org/10.1037/xhp0001167

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Are upside-down faces perceived as 'less human'?

Adam Eggleston[1,2*], Richard Cook[1,3], and Harriet Over[1]

[1]Department of Psychology,
University of York, York, U.K.

[2]School of Education, Language and Psychology,
York St John University, York, U.K.

[3]School of Psychology,
University of Leeds, Leeds, U.K.

*Correspondence:
a.eggleston@yorksj.ac.uk
School of Education, Language and Psychology,
York St John University,
Lord Mayor's Walk,
York, North Yorkshire
YO31 7EX, U.K.

Abstract

According to Perceptual Dehumanization Theory (PDT), faces are only perceived as "truly human" when processed in a configural fashion. Consistent with this theory, previous research indicates that when faces are inverted, a manipulation hypothesized to disrupt configural processing, the individuals depicted are attributed fewer uniquely human qualities. In a seminal paper, Hugenberg and colleagues (2016) reported that faces appeared less creative, less thoughtful, less empathetic, and possessed less 'humanness' when inverted. Across four highly powered and pre-registered experiments, we demonstrate that inversion does not influence the attribution of uniquely human traits specifically. Rather, in line with research on face processing, inversion impedes face encoding more generally, causing trait attributions to tend towards the mean. Positively valanced faces (i.e., those judged to be trustworthy when presented upright) are perceived to be less creative, considerate, thoughtful and empathetic when inverted. Conversely, negatively valanced faces (i.e., those judged to be untrustworthy when presented upright) are judged to be more creative, considerate, thoughtful, and empathetic when inverted. Furthermore, we show that the effect of inversion on judgments of "humanness" reflects a general phenomenon that can be replicated with other (non-face) stimulus categories that also possess a canonical orientation. These findings suggest that a key line of evidence for PDT is considerably less convincing than it first appears.

Keywords: Configural face processing; Dehumanization; Intergroup harm; Perceptual Dehumanization; Facial trustworthiness.

Public Significance Statement

Perceptual Dehumanization Theory states that i) faces are only perceived as truly human when they are processed configurally, and ii) upside-down faces are perceived as less human than upright faces because they recruit less configural processing. We interrogate these claims across four pre-registered and highly powered studies. We show that presenting faces upside-down influences perceived facial distinctiveness but not the attribution of humanity. We conclude that a key line of evidence for Perceptual Dehumanization Theory is considerably less convincing than it first appears.

Are upside-down faces perceived as 'less human'?

Humans are capable of extraordinary cruelty towards one another (Smith, 2020). A particularly influential claim within the social sciences is that a psychological process of dehumanization is one important cause of harm (Harris & Fiske, 2011; Haslam, 2006; Smith, 2016). According to social psychological theories of dehumanization, some individuals and groups are perceived to be less human than others. As a result, it is argued that they are more likely to be treated with indifference, neglect, and even open hostility (Harris & Fiske, 2011; Haslam, 2006; Haslam & Loughnan, 2014; Smith, 2020). Within this broad field of research, one prominent theory hypothesises that dehumanization sometimes takes the form of a 'bottom-up' perceptual process (Cassidy et al., 2017; Deska & Hugenberg, 2017; Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al., 2016; Wilson et al., 2018; Young et al., 2019).

Perceptual Dehumanization Theory (PDT) proposes that human faces typically engage specialised visual processing different from that recruited by objects. That is, human faces are thought to engage 'configural' processing in which individual facial features are integrated into a single coherent representation (Deska & Hugenberg, 2017; Fincher et al., 2017).[1] To the extent that individuals are perceptually dehumanized, their faces are assumed to recruit piecemeal visual processing, which – according to the authors of PDT – is the same kind of visual processing engaged by objects and the faces of non-human animals (Deska & Hugenberg, 2017; Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al., 2016). Perceptual dehumanization is thought to prevent individuals from being perceived in their full humanity and, consequently, places them at increased risk of harm (Deska & Hugenberg, 2017; Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg

[1] Note that whereas some proponents of PDT refer to "configural" face processing (Hugenberg et al. (2016), others refer to "configural and holistic processing" (Fincher & Tetlock, 2016). Whether the distinction is intentional or significant remains unclear. See section 6.2 for further discussion

et al., 2016). The logic at the heart of PDT is summarised by Hugenberg et al. (2016, p.168) who maintain that:

> "because human faces are processed configurally, in a manner distinct from other objects, we argue that configural processing is strongly associated with humanity and may therefore serve as a cue for humanity".

Further articulating the logic underlying PDT, Fincher and Tetlock (2017, pp. 288-289) claim:

> "A humanizing mode of perception begins when the perceiver engages mechanisms of visual processing that evolved to recognize human faces (…). In this mode, the perceiver processes the face configurally—that is, as a gestalt—recognizing not just a nose and a mouth and eyes but a person's face (…). This configural mode employs brain regions dedicated to face detection, which enable us to individuate faces better than other kinds of stimuli. However, we do not always see people in their full humanity—we sometimes engage in a dehumanizing mode of perception. This dehumanizing mode of perception begins when the perceiver focuses upon specific features such as lips or eyebrows rather than taking in the face as a whole (…). This is the same piece-by-piece mode of processing that we use to distinguish objects, such as when you recognize your coat in a closet".

The majority of research testing PDT has sought to manipulate the extent to which facial images are processed configurally and measure the effect of such manipulations on the attribution of humanity to the individuals depicted (Cassidy et al., 2017; Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al., 2016; Wilson et al., 2018). Although several paradigms have been adopted, by far the most common approach used in PDT research is to invert faces and measure the impact on perceptions of humanness (Cassidy et al., 2022; Civile et al., 2019; Deska & Hugenberg, 2017; Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al., 2016; Wilson et al., 2018). Proponents of PDT maintain that

inverted faces are attributed less humanity because orientation inversion selectively impairs configural processing (Deska & Hugenberg, 2017; Fincher et al., 2017).

The attribution of humanness has been operationalised in a number of different ways. In some research, it is measured directly, for example by asking participants 'how human' individuals appear (Cassidy et al., 2022; Hugenberg et al., 2016). In other research, humanity is inferred from the attribution of uniquely human character traits. According to a particularly prominent theoretical account of dehumanization from social psychology – the dual model (Haslam, 2006) – humans are thought to be distinguished from animals by virtue of possessing character traits such as rationality, civility and refinement. Similarly, humans are thought to be distinguished from objects by virtue of possessing character traits such as emotional warmth, depth and individuality. When individuals or groups are dehumanized, they are thought to be attributed uniquely human traits to a lesser extent and attributed traits humans share with other animals and objects such as passivity, superficiality, and rigidity, to a greater extent (Haslam, 2006).

In a seminal paper investigating PDT, Hugenberg et al. (2016) found that inverted faces were thought to possess uniquely human qualities of empathy, thoughtfulness, creativity, and consideration to a lesser extent than were upright faces. Inverted faces were also thought to be less 'humanlike' (Hugenberg et al., 2016). These findings have subsequently been replicated and extended by other researchers (e.g., Cassidy et al., 2022). This work appears to accord with the view that faces which fail to fully engage typical face processing are attributed fewer human qualities. If true, these results have important implications for our understanding of both face perception and intergroup harm.

Recently, however, an alternative interpretation of these results has been presented as part of a broader critique of PDT (Over & Cook, 2022). Over and Cook (2022) argue that what appears to be evidence for perceptual dehumanization in this paradigm may be better explained by the observation that orientation inversion impedes observers' ability to derive accurate perceptual descriptions of faces (McKone & Yovel, 2009; Rossion, 2008; Yin, 1969). When observing inverted faces, individuals are less able to detect and encode

stimulus variation: expressions seem more neutral, and face shapes appear less distinct (i.e., closer to the population average). Consistent with this view, inversion has been shown to adversely affect a wide range of perceptual decisions about faces that are unrelated to theoretical models of dehumanization including judgements about identity (Yin, 1969), age (Murphy & Cook, 2017), gender (Murphy et al., 2020), similarity (Biotti et al., 2019), expression (McKelvie, 1995; Prkachin, 2003), attractiveness (Bäuml, 1994; Cook & Duchaine, 2011), and adiposity (Thompson & Wilson, 2012). The fact that inverted facial percepts are impoverished might plausibly explain why individuals who appear thoughtful, considerate, creative and empathetic when their faces are viewed upright appear to exhibit these attributes to a lesser extent when their faces are viewed upside-down (Over & Cook, 2022).

1.1 Present work

In this study, we sought to compare these rival accounts. In our first experiment, we seek to replicate the findings of Hugenberg et al. (2016). In doing so, we test whether our method is sensitive to perceptual dehumanization effects if they occur. In Experiments 2 – 4, we pit the predictions of PDT against the alternative interpretation. We do so in three ways.

In Experiment 2, we compare the attribution of traits to faces that vary in valence (judged to be trustworthy vs. judged to be untrustworthy). We reasoned that the previously reported effects of inversion on judgments of uniquely human traits could be the product of Hugenberg et al. (2016) using a combination of socially desirable character traits (e.g., empathy, creativity) and inadvertently choosing faces that happened to elicit positive social evaluation when viewed upright; i.e., faces that were high on apparent facial trustworthiness. According to PDT, orientation manipulations should exert a similar effect on all faces. That is, all faces should appear to possess uniquely human character traits such as empathy, thoughtfulness, creativity, and consideration to a lesser extent when shown upside-down because configural processing is always impeded by inversion. In contrast, Over and Cook (2022) predict that the effect of face inversion on social evaluation will vary depending on the

particular faces used. According to the alternative interpretation, inversion does not impact judgments of humanity specifically, but rather makes faces appear less distinct because observers are less able to detect and encode distinguishing features. If correct, character attributions will tend towards the average when faces are shown upside-down. Sometimes this process will lead to more favourable social evaluations.

Facial trustworthiness is thought to be a global measure of facial valence: Unfamiliar faces that are spontaneously judged trustworthy by observers, are rated more positively on a host of dimensions (e.g., warmth, agreeableness, kindness, approachability) than are unfamiliar faces judged to be untrustworthy (e.g., Oosterhof & Todorov, 2008). Over and Cook (2022) predict that faces that appear trustworthy when viewed upright will be rated as less empathetic, thoughtful, creative, and considerate when viewed inverted. Conversely, faces that appear untrustworthy when viewed upright will be rated more empathetic, thoughtful, creative, and considerate when viewed inverted (Over & Cook, 2022).

In Experiment 3, we examine the sorts of trait judgement affected by orientation inversion. If the effect of inversion on the attribution of traits reflects perceptual dehumanization, then the trait judgments affected should be limited to, or especially strong for, those that distinguish humans from non-human entities such as animals and objects (Haslam, 2006). However, if these effects are attributable to impoverished perceptual description, inversion should disrupt the attribution of a wide range of character traits including those that are perceived to be shared with animals (e.g., being trusting or calm).

Hugenberg et al. (2016) report that human faces appear less “humanlike” when inverted – an effect that we replicate in our first two experiments. According to Hugenberg et al. (2016), this is a particularly convincing measure of perceptual dehumanization because of its “high face validity”. In Experiment 4, we address whether this finding can be explained by a more general phenomenon whereby exemplars of stimulus categories with a canonical orientation are judged less representative of their category when shown in the non-canonical orientation. To this end, we test whether sheep and cars are also judged to be less ‘sheeplike’ and ‘carlike’ when shown upside-down.

2. Experiment 1

In this experiment, we sought to confirm that the findings of Hugenberg et al. (2016) replicate. Following Hugenberg et al. (2016), we investigated whether faces are thought to possess the uniquely human attributes of creativity, consideration, empathy, thoughtfulness, and humanity to a greater extent when presented upright than when presented inverted.

2.1 Methods

2.1.1. Transparency and Openness

All four experiments were pre-registered. The pre-registration details, demographic questionnaire, and data supporting all of the analyses described are available here: https://osf.io/wtv4h/?view_only=406c25ae3b144b3c86dad1bb93ef6c41. Unedited face images were taken from the Chicago Face Database (Ma et al., 2015). The specific IDs of the face stimuli used and examples of all non-human stimuli can also be found online via the OSF link above.

2.1.2. Participants

Sample size was decided in advance and pre-registered. An *a priori* power analysis based on pilot data ($N = 60$), and conducted in G*Power 3.1.9.7, indicated that a minimum N of 84 was necessary to achieve power of .9 ($\alpha = .05$) with a small-medium effect size ($d = 0.36$).

Ninety participants were recruited in 2022 via www.prolific.co. One participant was excluded and replaced having failed more than 50% of the attention checks. Of the 90 participants included in the final analysis ($M_{age} = 39.36$, $SD_{age} = 13.38$), 55 identified their preferred gender pronouns as she, her, hers; 30 as he, him, his; 3 as they, them, theirs, and 2 preferred not to say. All participants described English as their first language and reported current residence within the UK. Based on the options given in our demographic questionnaire, 81 participants identified as White British, 1 as White and Black Caribbean, 1

as Caribbean, 1 as White and Black African, 2 as Indian, 3 as Pakistani, and one preferred not to say. All participants gave written informed consent and received a small honorarium of £3.43 for approximately 25 minutes participation. For all studies, the procedures were approved by the University of York Ethics Committee and were performed in accordance with the Committee's guidelines and the Declaration of Helsinki.

2.1.3. Materials

Following Hugenberg et al. (2016), participants viewed 40 White male faces selected from the Chicago Face Database (Ma et al., 2015). In line with Hugenberg et al. (2016), faces were edited to exclude clothing and hair cues and converted to greyscale. The cropped faces were standardized to an aspect ratio of 380 × 560 pixels. All faces used were previously rated on a scale of trustworthiness (scale range 1-7; M = 3.28, Min = 2.56, Max = 3.92; see Ma et al., 2015). The 40 faces were divided into two sets of 20. Half of the participants judged the first set upright and the second set inverted, while half judged the first set inverted, and the second set upright (see Supplementary Materials).

2.1.4. Design

All participants viewed images of faces upright and inverted in a within-subjects design. Orientation was counterbalanced across two conditions such that all faces were presented equally often upright and inverted. Following the same procedure as Hugenberg et al. (2016), participants were asked to indicate how thoughtful, empathetic, considerate, creative, and humanlike each face appeared, on scales from 0 (Not at all) to 100 (Extremely). Consistent with Hugenberg et al. (2016), the key dependent variable was the average of all trait attributions.

2.1.5. Procedure

The experiment was administered online via Gorilla (Anwyl-Irvine et al., 2020). The task had to be completed using a personal computer or laptop – it would not run on a tablet or mobile

device. These restrictions were implemented by selecting the appropriate the presentation options in Gorilla Experiment Builder, using the default device parameters.

At the start of the experiment participants were instructed that people often show accuracy in personality ratings of others at zero acquaintance. This was done to be consistent with previous work (Hugenberg et al., 2016), though accuracy in personality ratings is a claim disputed within the impression formation literature (e.g., Cook et al., 2022; Cook & Over, 2021; Efferson & Vogt, 2013; Todorov & Porter, 2014). All participants then completed two practice trials to get used to the speed of stimulus presentation. During the practice trials, participants viewed two faces (one inverted, one upright) not used in the main experiment and rated how gentle they appeared. This was followed by the main task whereby participants viewed each of the 40 White male faces (20 upright, 20 inverted) one at a time, in a random order. Each trial began with a fixation cross (250 ms). A face stimulus was then displayed at the centre of the screen (750 ms), followed immediately by a mask constructed of high-contrast greyscale ovals (250 ms). Finally, a rating screen appeared which indicated the to-be-rated trait (see Figure 1a). Participants made a self-paced rating via mouse click. This was repeated for each face and trait combination, resulting in 200 experimental trials (20 faces × 2 orientations × 5 traits). After 100 trials, participants were offered a one-minute break to help them maintain attention. Four attention checks were also included, one after each set of 50 test trials. Participants saw one of the faces from the practice trials shown upright or inverted and simply had to respond with the correct orientation. Participants who failed to respond correctly on at least 3 of the 4 attention checks were excluded and replaced. All participants were fully debriefed about the aims of the study and the reason for using all White male faces.

2.2. Results

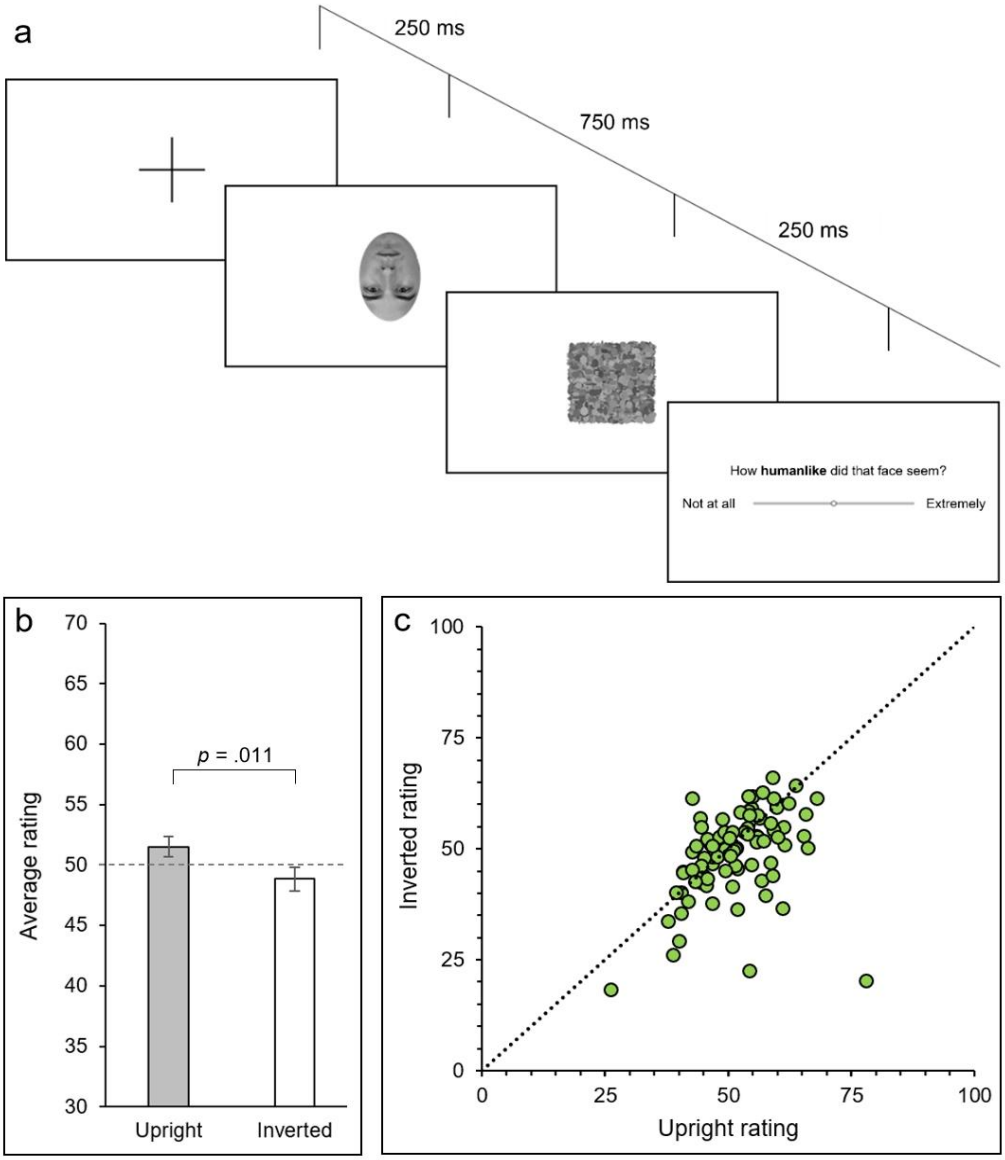
2.2.1. Pre-registered analyses

A paired samples *t*-test revealed that, on average, participants rated upright faces (M = 51.50, SD = 8.10) as possessing uniquely human traits to a significantly greater extent than

inverted faces (M = 48.84, SD = 9.50), $t(89) = 2.59$, $p = .011$, $d = .27$ (Figure 1b, 1c). Although the effect size is relatively modest, these findings are in line with the previously reported findings of Hugenberg et al. (2016).

Figure 1

Overview of the method and results for Experiment 1



Note. (a) Schematic illustration of a trial sequence from Experiment 1. (b) Average trait ratings for faces presented upright and inverted in Experiment 1. Error bars denote ±1 standard error. (c) Participants' upright trait ratings plotted against their inverted trait ratings. Where points fall to the right of the diagonal, participants have awarded higher ratings in the upright condition. Where points fall to the left of the diagonal, participants have awarded higher ratings in the inverted condition.

2.2.2. Exploratory analyses

2.2.2.1. Investigating the consistency of the effect

In addition to our pre-registered analysis, we conducted an exploratory 2 (Orientation: Upright, Inverted) × 5 (Attribute: thoughtful, empathetic, considerate, creative, humanlike) within-subjects ANOVA (see Figure S1 in the Supplementary Material). We did this to test whether the effect of inversion was similar for all five traits. The data violated assumptions of sphericity so the ANOVA results described have undergone Greenhouse-Geisser correction.

A significant Orientation × Attribute interaction was found, $F(2.87, 255.34) = 11.32$, $p < .001$, $\eta^2_p = .11$. Bonferroni corrected pairwise comparisons revealed that upright faces were judged significantly more humanlike ($M = 76.30$, $SD = 17.40$) than inverted faces ($M = 70.08$, $SD = 22.06$), $t(89) = 3.98$, $p < .001$, $d = .42$. Similarly, upright faces were judged more thoughtful ($M = 49.47$, $SD = 10.43$) than inverted faces ($M = 45.33$, $SD = 9.78$), $t(89) = 3.25$, $p = .002$, $d = .34$. Comparisons between upright and inverted ratings on the remaining traits were in the same direction but failed to reach statistical significance (all $ps > .174$). These results suggest that the influence of inversion may be particularly pronounced for ratings of ‘humanlike’ and 'thoughtful’ compared to other attributes.

2.2.2.2. Exploring an alternative interpretation

Hugenberg et al. (2016) provide little information about the stimuli used in their study or the selection criteria. However, Over and Cook (2022) reasoned that the apparent effects of inversion on judgments of uniquely human traits could be the product of Hugenberg et al. (2016) using a combination of socially desirable character traits (e.g., empathy, creativity) and inadvertently choosing faces that happened to elicit positive social evaluation when viewed upright; i.e., faces that were high on facial trustworthiness. According to this alternative view, faces are perceived as less distinct when inverted, hence any ratings thereof will tend towards the mean. Over and Cook (2022) hypothesized that faces judged trustworthy when viewed upright will appear to possess socially desirable attributes to a

lesser extent when inverted. However, faces judged untrustworthy when viewed upright will appear to possess socially desirable attributes to a greater extent when inverted.

As a first step towards exploring this possibility, we performed a correlational analysis on our data to assess whether there was a relationship between the effect of orientation on trait attribution and facial trustworthiness. For each trait-face combination, we calculated ΔOrientation: the average trait rating awarded by participants to each face when shown upright minus the average trait rating awarded to the same face when shown inverted. Because the trait constructs employed are all socially desirable, a positive ΔOrientation score indicates that the face secured more favourable evaluation when shown upright, while a negative ΔOrientation score indicates that the face secured more favourable evaluation when shown inverted. We then correlated the ΔOrientation scores with the facial trustworthiness of each face as previously reported by Ma et al. (2015). For all five attributes, facial trustworthiness was positively associated with the ΔOrientation scores (Figure 2; Table 1). Further information on these analyses is provided in the Supplementary Materials (Tables S1 & S2).

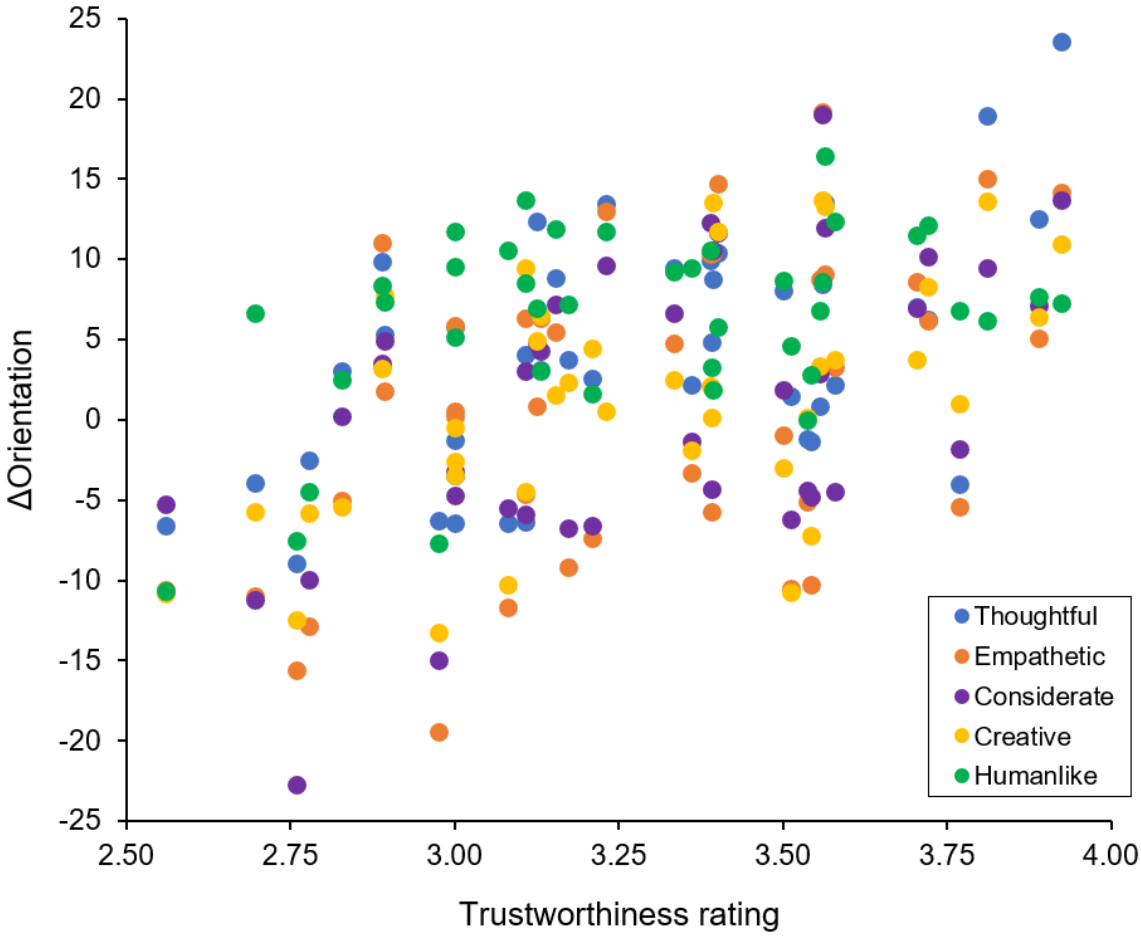
Table 1

Correlations between ΔOrientation scores and facial trustworthiness for the five attributes

	ΔOrientation		r ($CI_{95\%}$)	p-value
	Mean (SD)	Range		
Humanlike	6.21 (5.91)	-10.67 : 16.44	0.42 (0.13 : 0.65)	.007
Thoughtful	4.14 (7.44)	-8.93 : 23.58	0.54 (0.28 : 0.73)	< .001
Empathetic	0.82 (9.64)	-19.44 : 19.20	0.48 (0.20 : 0.69)	.002
Considerate	0.86 (8.71)	-22.69 : 19.04	0.55 (0.29 : 0.73)	<.001
Creative	1.29 (7.52)	-10.67 : 16.44	0.54 (0.28 : 0.73)	<.001

Figure 2

Scatterplot illustrating the correlations between ΔOrientation scores and facial trustworthiness for each of the five attributes measured in Experiment 1.



This analysis provides initial evidence for the alternative account. Next, we investigate this idea in a pre-registered experiment that compared attributions of socially desirable human attributes to faces that varied systematically in apparent trustworthiness.

3. Experiment 2

In Experiment 2, we pit the predictions of PDT against those of the alternative account by comparing the attribution of human traits to faces that vary in apparent trustworthiness. PDT predicts that all faces will appear less creative, considerate, empathetic, thoughtful, and humanlike, when inverted because configural face processing, and thus the perception of humanness, is always impeded by inversion (Cassidy et al., 2022; Hugenberg et al., 2016).

According to the alternative view, inversion does not lead to impaired attribution of uniquely human traits. Rather, inverted faces appear less distinctive than upright faces because observers are less able to detect and encode distinguishing features (Over & Cook, 2022). Consequently, all trait ratings will tend towards the average when face stimuli are shown inverted. Thus, like Hugenberg et al. (2016) we predict that faces that secure positive evaluation when shown upright (trustworthy faces) will appear less creative, considerate, empathetic, thoughtful, and humanlike when inverted. Unlike Hugenberg et al. (2016), however, we predict that faces that secure negative evaluation when shown upright (untrustworthy faces) will appear more creative, considerate, empathetic, thoughtful, and humanlike when inverted.

3.1. Methods

3.1.1. Participants

Sample size was decided in advance and pre-registered. An *a priori* power analysis conducted using MorePower 6.0.4 found that a minimum *N* of 126 was necessary to detect interactions with a medium effect size (η^2_p = .06) with a power of .8 (α = .05). Note, the results of a pilot study (*N* = 40) suggested a large effect. For the purposes of the power calculation, however, we were more conservative (adopting the benchmark medium effect of η^2_p = .06), as effect size estimates obtained with small samples can be unreliable.

A sample of 130 participants was recruited in 2022 via Prolific and tested online via Gorilla (Anwyl-Irvine et al., 2020). Eleven were excluded and replaced following our pre-registered exclusion criteria. Of the 130 participants included in the analysis (M_{age} = 39.31, SD_{age} = 14.60), 93 identified their preferred gender pronouns as she, her, hers; 31 as he, him, his; 5 as they, them, theirs, and 1 preferred not to say. All participants described English as their first language and reported current residence within the UK. Based on the options given in our demographic questionnaire,113 participants identified as White British, 2 as White Irish, 1 as White and Black Caribbean, 2 as White and Asian, 1 as Indian, 1 as

Pakistani, 1 as Bangladeshi, 3 as Black African, 1 as Black Caribbean, 1 as Asian Filipino, 1 as Southeast Asian, 1 as White Slavic, and 2 preferred not to say. All participants gave written informed consent and received a small honorarium of £3.34 for approximately 25 minutes participation.

3.1.2. Materials

A selection of 40 White male faces were again selected from the Chicago Face Database (Ma et al., 2015). We elected to use White male faces to ensure that the results of the second experiment were directly comparable to those of Experiment 1. Unlike Experiment 1, however, the face stimuli used in Experiment 2 were purposely selected because they had been judged relatively trustworthy (20 faces: $M = 3.67$, $SD = 0.13$) or untrustworthy (20 faces: $M = 2.76$, $SD = 0.19$). A *t*-test confirmed that these two sets of faces differed significantly in terms of their trustworthiness ratings, $t(38) = 17.86$, $p < .001$. The trustworthy and untrustworthy faces chosen were among the most and least trustworthy White male faces in the database. Further information on the trustworthiness ratings is available in the Supplementary Materials (Figure S2).

3.1.3. Design & Procedure

Face Type (trustworthy, untrustworthy) and Orientation (upright, inverted) were manipulated within-subjects. Stimulus order was randomized. The procedure was identical to Experiment 1. Participants rated faces on the same five attributes (thoughtful, empathetic, considerate, creative, humanlike). Once again, two practice trials preceded the main experiment. Participants who failed 50% or more of the attention checks were excluded.

3.2. Results

3.2.1. Pre-registered analyses

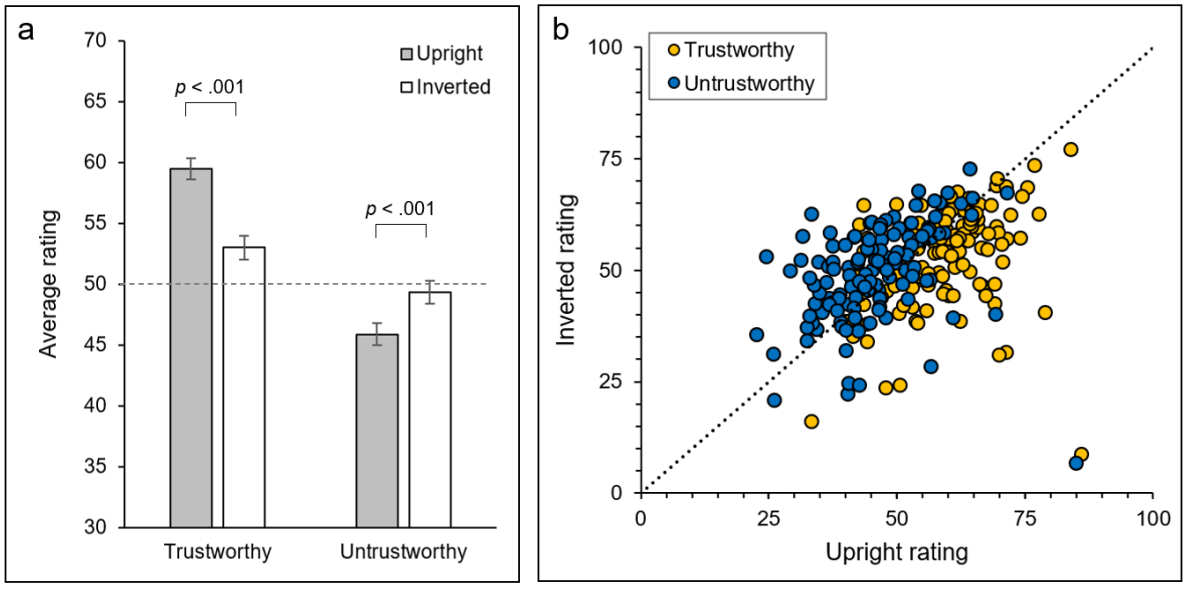
A 2 (Orientation: upright, inverted) × 2 (Face Type: trustworthy, untrustworthy) within-subjects ANOVA was conducted (Figure 3). There was no significant main effect of

Orientation (inverted: $M = 51.17$, $SD = 10.98$; upright: $M = 52.67$, $SD = 11.92$), $F(1, 129) = 2.29$, $p = .133$, $\eta^2_p = .02$. A significant main effect of Face Type was found. As expected, the trustworthy faces were rated higher on all five trait dimensions ($M = 56.23$, $SD = 10.81$) than were the untrustworthy faces ($M = 47.61$, $SD = 10.48$), $F(1, 129) = 405.84$, $p < .001$, $\eta^2_p = .76$.

In line with our predictions and contrary to those of PDT, a highly significant Orientation × Face Type interaction was found, $F(1, 129) = 245.55$, $p < .001$, $\eta^2_p = 0.66$. Bonferroni corrected pairwise comparisons revealed that trustworthy faces were awarded lower average ratings when inverted ($M = 52.99$, $SD = 10.98$) than when upright ($M = 59.46$, $SD = 9.63$), $t(129) = 6.21$, $p < .001$, $d = .54$. Conversely, untrustworthy faces were awarded higher average ratings when inverted ($M = 49.34$, $SD = 10.72$) than when upright ($M = 45.88$, $SD = 9.98$), $t(129) = 3.33$, $p = .001$, $d = .29$.

Figure 3

The results of Experiment 2



Note. (a) Average trait ratings for faces presented upright and inverted in Experiment 2. Error bars denote ±1 standard error. (b) Participants' upright trait ratings plotted against their inverted trait ratings. Where points fall to the right of the diagonal, participants have awarded higher ratings in the upright condition. Where points fall to the left of the diagonal, participants have awarded higher ratings in the inverted condition.

3.2.2. Exploratory analyses

We ran an exploratory 2 (Orientation: upright, inverted) × 2 (Face Type: trustworthy, untrustworthy) × 5 (Attribute: humanlike, thoughtful, empathetic, considerate, creative) within-subjects ANOVA to assess whether our predicted pattern of results held for all five judgments considered separately (see Figure S3 in the Supplementary Material). A significant Orientation × Face Type × Attribute interaction was found, $F(4, 516) = 25.01$, $p < .001$, $\eta^2_p = .16$. For the traits creative, thoughtful, considerate and empathetic, the pattern was similar and followed the predictions of Over and Cook (2022). Trustworthy faces were seen to possess these attributes to a lesser extent when inverted (all *p*s < .001), while untrustworthy faces were seen to possess these attributes to a greater extent when inverted (all *p*s < .003). Interestingly, the pattern for one judgment – how humanlike the faces seemed – differed and followed the predictions of PDT. Trustworthy and untrustworthy faces were both seen as less humanlike when inverted (*ps* < .001). We revisit this finding in Experiment 4.

4. Experiment 3

In our third experiment, we sought to replicate and extend the findings of Experiment 2. We investigated whether orientation disproportionately influences attributions of uniquely human traits, or whether it influences the attribution of all traits, including those shared with animals. To address this question, we asked participants to rate faces judged trustworthy and untrustworthy on trait terms that, according to previous research, are perceived to be unique to humans: open-minded, civilised, sophisticated, and knowledgeable (Enock et al., 2021; Haslam, 2006). We also asked participants to rate these faces on another set of trait terms that, according to previous research, are perceived to be shared with other animals: trusting, curious, genuine, and calm (Enock et al., 2021; Haslam, 2006).

PDT asserts that inversion influences the attribution of humanness. Consequently, inversion ought to disproportionately influence the attribution of uniquely human traits (Cassidy et al., 2022; Deska et al., 2017; Deska et al., 2018; Hugenberg et al., 2016; Wilson

et al., 2018). On the other hand, Over and Cook (2022) propose that inversion does not specifically influence perceptions of humanness. Rather, inversion leads to an impoverished perceptual representation that disrupts a host of judgements about faces. This alternative account predicts that inversion will influence the attribution of traits shared with other species in a similar way to the attribution of uniquely human traits (Over & Cook, 2022).

4.1. Methods

4.1.1. Participants

As we had no predictions for a three-way interaction, but did predict a two-way interaction, we maintained the same sample size based on our power analysis from Experiment 2. Once again, a pre-registered sample of 130 participants were recruited in 2022 via Prolific and tested online via Gorilla (Anwyl-Irvine et al., 2020). No participants were replaced or excluded. Of the 130 participants (M_{age} = 41.84, SD_{age} = 14.47), 99 identified their preferred gender pronouns as she, her, hers; 20 as he, him, his; 4 as they, them, theirs, and 7 preferred not to say. All participants described English as their first language and 128 reported current residence within the UK. One participant resided in Canada and another in Ireland. Based on the options given in our standardised demographic questionnaire,117 of the participants identified as White British, 2 as White Irish, 1 as White and Asian, 1 as Asian British, 1 as Black British, 1 as Indian, 1 as Chinese, 2 as Caribbean, 1 as Arab, and 3 preferred not to say. All participants gave written informed consent and received a small honorarium of £3.75 for approximately 30 minutes participation.

4.1.2. Materials

The same 40 White male faces used in Experiment 2 were used in Experiment 3 to ensure that our results were directly comparable.

4.1.3. Design & Procedure

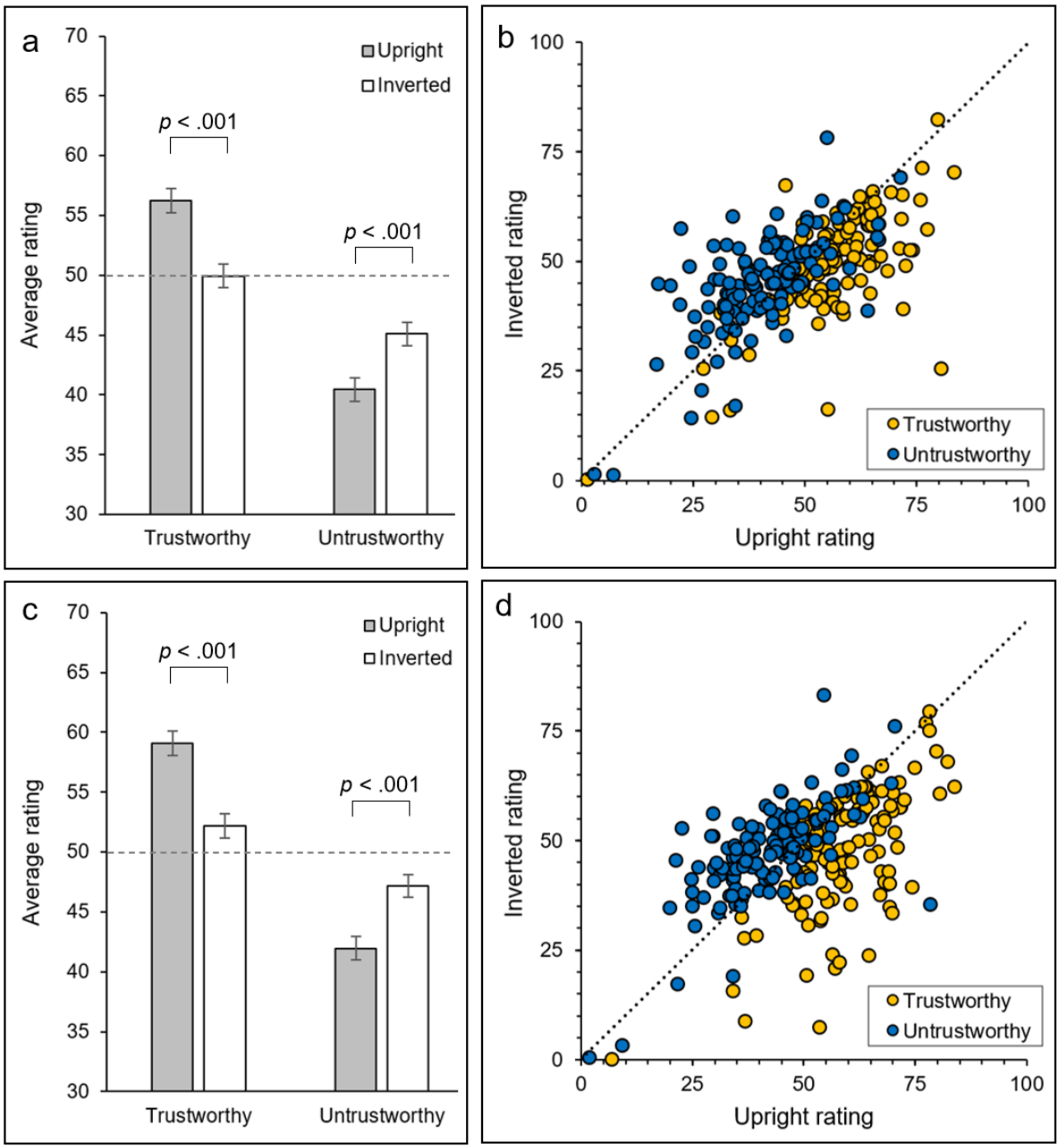
All faces were rated on 8 socially desirable character traits. According to previous findings (Enock et al., 2021), people typically perceive four of these traits to be unique to humans (open-minded, civilised, sophisticated, knowledgeable) and four to be shared with other species (trusting, curious, genuine, calm). The procedure was identical to Experiment 2 with the exception that the faces were rated on the 8 traits described above. In total there were 320 experimental trials (20 faces × 2 orientations × 8 traits). Breaks were offered after 100 trials and 200 trials. Two additional attention checks were added resulting in 6 overall. Participants who failed 50% or more of the attention checks were excluded. Once again, two practice trials preceded the main experiment.

4.2. Results

A 2 (Orientation: upright, inverted) × 2 (Trait Type: unique, shared) × 2 (Face Type, trustworthy, untrustworthy) within-subjects ANOVA was conducted (Figure 4; Supplementary Figure S4). As expected, there was a significant main effect of Face Type, $F(1, 129) = 476.96$, $p < .001$, $\eta^2_p = .79$, indicating that trustworthy faces ($M = 54.37$, $SD = 11.79$) were awarded higher ratings than untrustworthy faces ($M = 43.67$, $SD = 11.48$). There was also a significant main effect of Trait Type with faces on average rated higher on shared traits ($M = 50.11$, $SD = 12.80$) than on uniquely human traits ($M = 47.94$, $SD = 12.73$), $F(1, 129) = 36.09$, $p < .001$, $\eta^2_p = .22$. There was no significant main effect of Orientation (Inverted, $M = 48.60$, $SD = 11.48$; Upright, $M = 49.45$, $SD = 14.01$), $F(1, 129) = 1.50$, $p = .222$, $\eta^2_p = .01$. Contrary to the predictions of PDT, and in line with the results of Experiment 2, there was a significant Orientation × Face Type interaction, $F(1, 129) = 321.55$, $p < .001$, $\eta^2_p = 0.71$. Bonferroni corrected pairwise comparisons revealed that trustworthy faces received lower ratings when they were inverted ($M = 51.06$, $SD = 11.20$) than when shown upright ($M = 57.69$, $SD = 10.81$), $t(129) = 8.64$, $p < .001$, $d = .76$. Conversely, untrustworthy faces received higher ratings when they were inverted ($M = 46.14$, $SD = 10.68$) than when shown upright ($M = 41.21$, $SD = 11.07$), $t(129) = 6.50$, $p < .001$, $d = .57$. A significant Trait Type ×

Figure 4

The results of Experiment 3



Note. (a) Average uniquely human trait ratings for faces presented upright and inverted. Error bars denote ±1 standard error. (b) Participants' upright ratings plotted against their inverted ratings for uniquely human traits. For each participant, the mean ratings awarded to trustworthy faces (yellow) and untrustworthy faces (blue) are plotted separately. Where points fall to the right of the diagonal, participants have awarded higher ratings in the upright condition. Where points fall to the left of the diagonal, participants have awarded higher ratings in the inverted condition. (c) Average shared trait ratings for faces presented upright and inverted. (d) Participants' upright ratings plotted against their inverted ratings for shared traits.

Face Type interaction was also found, $F(1, 129) = 5.95$, $p = .016$, $\eta^2_p = .044$, whereby the effect of trustworthiness was slightly greater for shared traits than for uniquely human traits. Importantly, there was no Orientation × Trait Type interaction, $F(1, 129) = 0.01$, $p = .921$, $\eta^2_p < .001$. In other words, we found no evidence that orientation disproportionately affects the attribution of uniquely human attributes, relative to those shared with other species.

We observed highly significant Orientation × Face Type interactions when uniquely human traits [$F(1,129) = 224.21$, $p < .001$, $\eta^2p = .64$] and shared traits $F(1,129) = 308.60$, $p < .001$, $\eta^2p = .71$] were considered separately. Trustworthy faces received lower ratings when they were inverted than when they were upright regardless of whether the attributes being rated were uniquely human or shared (both *p*s < .001). Similarly, untrustworthy faces received higher ratings when they were inverted than when they were upright regardless of whether the attributes being rated were uniquely human or shared (both *p*s < .001).

Nevertheless, the Trait Type × Face Type × Orientation interaction tended towards significance, $F(1, 129) = 3.86$, $p = .052$, $\eta^2_p = .029$. This non-significant trend was such that the Face Type × Orientation interaction tended to be stronger when judging shared traits than when judging uniquely human traits. Importantly, this trend is in the opposite direction as would be predicted by PDT; i.e., according to PDT, any interaction with Orientation should manifest more strongly when judging uniquely human traits.

5. Experiment 4

In Experiment 2, we found that most trait judgments followed the pattern of results predicted by Over and Cook (2022). Faces judged trustworthy when viewed upright were rated as less empathetic, creative, thoughtful, and considerate when inverted. Faces judged untrustworthy when viewed upright were rated as more empathetic, creative, thoughtful, and considerate when inverted. Ratings of how humanlike the faces were followed a different pattern, however, seemingly in line with the predictions of PDT. Replicating and extending the findings of Hugenberg et al. (2016) and Cassidy et al. (2022), both trustworthy and untrustworthy faces were viewed as less humanlike when inverted. Interestingly this is also

the item that Hugenberg et al. (2016) highlight as the most convincing measure of dehumanization. However, we suggest that what appears to be evidence of perceptual dehumanization may be the product of a more general phenomenon whereby exemplars of stimulus categories with a canonical orientation are judged less representative of their category when shown in the non-canonical orientation (Over & Cook, 2022).

To test this alternative interpretation, we ran a final experiment in which we asked participants to judge 'how humanlike' faces appeared when presented upright and inverted. In addition, we also presented images of sheep and cars upright and inverted and asked participants how 'sheeplike' and 'carlike' they appeared. If the effects of inversion on humanlike judgements are the product of a more general tendency to report that objects with a canonical orientation look less prototypical when shown upside-down, we reasoned it should be possible to replicate this effect with other non-face stimuli (e.g., sheep and cars).

5.1. Methods

5.1.1. Participants

Sample size was decided in advance and pre-registered. Using pilot data ($N = 40$), we conducted a simulation-based power analysis using the Superpower's Power Shiny App (Lakens & Caldwell, 2021). Results suggested that a minimum N of 90 would be required to detect a 2 × 3 interaction ($\eta^2_p = .06$) in a repeated measures ANOVA with a power of .8 ($\alpha = .05$).

Ninety participants were recruited in 2022 via Prolific and tested online via Gorilla (Anwyl-Irvine et al., 2020). In accordance with our pre-registered exclusion criteria, 7 participants were replaced having failed the attention check. Of the 90 participants included in the analysis ($M_{age} = 40.94$, $SD_{age} = 14.29$), 58 identified their preferred gender pronouns as she, her, hers; 28 as he, him, his; and 4 preferred not to say. All participants described English as their first language and 84 reported current residence within the UK. One participant resided in Canada, 1 in Ireland, 1 in France, 1 in Spain, 1 in Italy, and 1 in Israel.

Based on the options given in our demographic questionnaire, 84 participants identified as White British, 1 as Asian British, 1 as Caribbean, 1 as White and Black Caribbean, 1 as White and Black African, 1 as Black British, and 1 as White European. All participants gave written informed consent and received a small honorarium of £2.50 for approximately 20 minutes participation. [2]

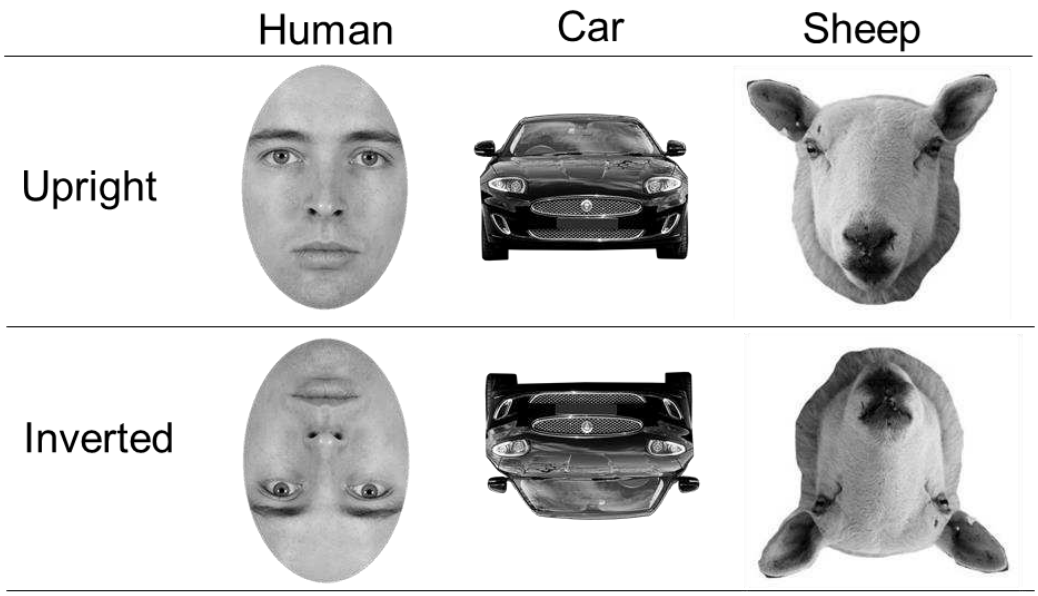
5.1.2. Materials

The stimuli used in Experiment 4 consisted of 40 images of White male faces, 40 images of sheep faces, and 40 images of cars viewed from the front (Figure 5). The face stimuli were the same 40 images employed in Experiment 1. As in Experiment 1, the facial images were presented in greyscale with an aspect ratio of 380 × 560 pixels. The car and sheep images were sourced from various websites. The exact aspect ratio of the car and sheep varied slightly across exemplars. The car stimuli were approximately 465 × 380 pixels while the sheep stimuli were approximately 465 × 330 pixels. The car and sheep images were also converted to greyscale to ensure consistency with the facial images. Each set of 40 images was divided into two sets of 20 images for the purposes of counterbalancing (see Supplementary Materials).

[2] Across the four experiments described, we employed three different power analysis programs to estimate the required sample-size (G-Power, MorePower, & Shiny App). The use of these different applications is stated explicitly in the respective pre-registration documents for each experiment. In the past, we have routinely used G-Power for all our power calculations. Over the course of this project, however, we became aware of a particular limitation of G-Power (version 3.1.9.7) when calculating power for repeated-measures designs with multiple within-subjects factors such as those employed in Experiments 2-4. Given that we often employ this kind of design, we have been exploring various alternatives. While we recognise that the use of different applications is slightly inelegant, we hope that readers will understand that authentic pre-registered science occasionally contains some honest wrinkles.

Figure 5

Examples of the human, car and sheep stimuli used in Experiment 4



5.1.3. Design

Orientation (upright, inverted) and Target Type (faces, sheep, cars) were manipulated within-participants. Orientation was counterbalanced across participants such that each stimulus appeared equally often upright and inverted. Stimulus order was randomized.

5.1.4. Procedure

Trials began with a fixation cross (250 ms), followed by a stimulus image at the centre of the screen (750 ms), and then a mask constructed of high-contrast ovals (250 ms). Finally, participants viewed one of three rating scales; the human scale (“How humanlike did that image seem?” rated from 0 [not at all] to 100 [extremely]), the sheep scale (“How sheeplike did that image seem?” rated from 0 [not at all] to 100 [extremely]), or the car scale (“How carlike did that image seem?” rated from 0 [not at all] to 100 [extremely]). They then made a self-paced rating via mouse click. In total, there were 120 experimental trials (20 images × 2

orientations × 3 categories). After 60 trials, participants were offered a break. The break was followed immediately by an attention check. Participants were asked how long they took for their break. Instructions indicated that this was an attention check and to select 20 seconds from the dropdown list. Participants were excluded if they failed this attention check. At the end of the experiment participants were thanked and fully debriefed. Two practice trials preceded the main experiment.

5.2. Results

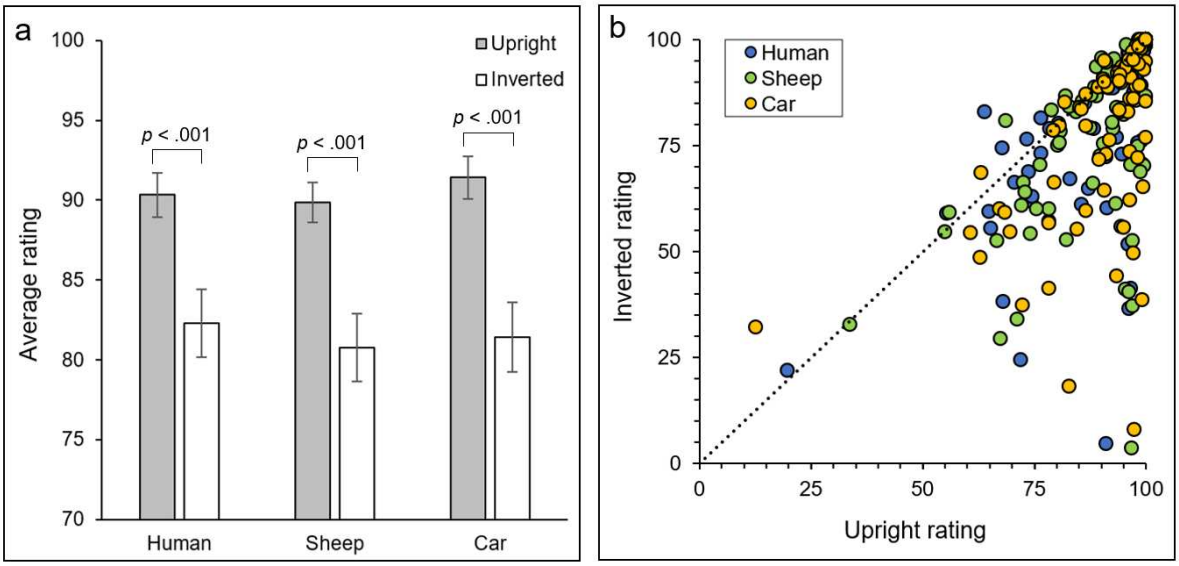
In accordance with our pre-registered analysis plan, a 2 (Orientation: Upright, Inverted) × 3 (Target Type: Human, Sheep, Car) within-subjects ANOVA was conducted (Figure 6). Results for Target Type and Orientation × Target Type were shown to violate assumptions of sphericity and have been subjected to Greenhouse-Geisser correction, accordingly.

In line with our prediction, a significant main effect of Orientation was found, $F(1, 89) = 28.68$, $p < .001$, $\eta^2_p = .24$. Bonferroni corrected pairwise comparisons revealed that upright exemplars were rated as more prototypical than were inverted exemplars for all three stimulus categories (all ps $< .001$). There was no significant main effect of Target Type, $F(1.81, 161.13) = 1.40$, $p = .250$, $\eta^2_p = .02$. However, there was a significant Orientation × Target Type interaction, $F(1.63, 144.22) = 3.48$, $p = .043$, $\eta^2_p = .04$, whereby the effect of Orientation was greater for cars than for human faces.

It has previously been reported that inverted faces seem less humanlike than upright faces (Hugenberg et al., 2016). We were able to replicate this effect in Experiment 2 with both trustworthy and untrustworthy faces. Rather than being evidence for a psychological process of perceptual dehumanization, however, the results of Experiment 4 indicate that this effect reflects a simpler, more general phenomenon: Exemplars of stimulus categories with a canonical orientation (e.g., sheep, cars) are judged less representative of their category when shown in the non-canonical orientation.

Figure 6

The results of Experiment 4



Note. (a) Average prototypicality ratings for upright and inverted images in Experiment 4. Error bars denote ±1 standard error. (b) Participants' upright ratings plotted against their inverted trait ratings. Where points fall to the right of the diagonal, participants have awarded higher prototypicality ratings in the upright condition.

6. General Discussion

According to PDT, dehumanization sometimes takes the form of a bottom-up perceptual process. Whereas upright faces are typically thought to engage configural processing (Deska & Hugenberg, 2017; Fincher et al., 2017), dehumanized faces are thought to recruit piecemeal visual processing. According to the authors of PDT, this piecemeal visual processing is the same kind of analysis engaged by inanimate objects and non-human animals (Deska & Hugenberg, 2017; Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al., 2016). The engagement of configural face processing, it is argued, is crucial for seeing others in their full humanity and protects against discrimination and other forms of harm (Cassidy et al., 2017; Civile et al., 2019; Hugenberg et al., 2016).

If PDT is correct, then manipulations that modulate the engagement of configural processing will affect the attribution of humanity (Fincher & Tetlock, 2016; Hugenberg et al.,

2016). By far the most common approach to testing this claim is to invert faces and measure the impact on perceptions of humanness (Cassidy et al., 2022; Civile et al., 2019; Deska et al., 2017; Fincher & Tetlock, 2016; Hugenberg et al., 2016). In a seminal paper on this topic, Hugenberg et al. (2016) found that faces appeared to possess certain uniquely human character traits (e.g., thoughtful, empathetic, considerate, creative) to a lesser extent when inverted.

In the experiments described here, we tested an alternative explanation for these influential results (Over & Cook, 2022). Apparent effects of inversion on attributions of humanity may be better explained by the observation that inversion disrupts the ability of observers to derive accurate perceptual descriptions of faces (McKone & Yovel, 2009; Rossion, 2008; Yin, 1969). Because observers are less able to encode stimulus variation when faces are inverted, face shapes appear average and nondescript. Consequently, one might expect all trait ratings to tend towards the average when facial stimuli are inverted. Hugenberg et al. (2016) do not report which faces from the Chicago Face Database they included as stimuli. It is possible, however, that the faces used tended to secure positive social evaluation when viewed upright (i.e., on average, they were relatively high on facial trustworthiness). When rating trustworthy faces, regression to the mean induced by inversion would tend to suppress ratings for positive, socially desirable attributes; i.e., facial cues that afford positive evaluation when viewed upright may pass undetected by observers when viewed inverted.

In Experiment 2, we compared the attribution of traits to faces that varied systematically in apparent trustworthiness. Orientation inversion would be expected to disrupt configural face processing irrespective of facial trustworthiness. As such, PDT predicts that inversion should impede the attribution of humanity to trustworthy and untrustworthy faces alike (Fincher & Tetlock, 2016; Hugenberg et al., 2016). The alternative account of Over and Cook (2022), however, predicts that the influence of inversion will vary depending on the valance of the to-be-judged faces. In line with this prediction, and contrary to the predictions of PDT, we found that: i) Faces that appear trustworthy when upright are

perceived to have fewer socially desirable uniquely human attributes when inverted. ii) Faces that appear untrustworthy when upright are perceived to have more socially desirable human attributes when inverted.

In Experiment 3, we further compared the explanatory power of PDT and the alternative account proposed by Over and Cook (2022) by measuring how inversion influences the attribution of socially desirable traits that are unique to humans and those that are shared with non-human animals. We chose this manipulation because, according to PDT, inversion disrupts the attribution of humanness (Cassidy et al., 2022; Civile et al., 2019; Deska et al., 2017; Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al., 2016; Wilson et al., 2018). If this is true, then inversion should disproportionately affect ratings of uniquely human traits (Cassidy et al., 2022; Deska et al., 2017; Deska et al., 2018; Hugenberg et al., 2016; Wilson et al., 2018). According to the alternative account, however, both kinds of attribution depend on a perceptual description that is rendered imprecise and nondescript by orientation inversion (McKone & Yovel, 2009; Rossion, 2008). As such, the alternative account predicts that inversion should influence the attribution of traits shared with other species in a similar way to the attribution of uniquely human traits. In line with the prediction of Over and Cook (2022), and replicating the results of Experiment 2, we found that faces that appear trustworthy when viewed upright were attributed fewer socially desirable traits when inverted, regardless of whether or not these traits are perceived to be unique to humans. Conversely, faces that appear untrustworthy when upright were attributed more socially desirable traits when inverted, regardless of whether or not these traits are perceived to be unique to humans.

In our final experiment, we investigated how orientation inversion affects judgements of prototypicality. In Experiment 2, when asked how ‘humanlike’ stimuli appeared, both trustworthy and untrustworthy faces were judged more humanlike when upright than when inverted. This finding is consistent with the predictions of PDT and replicates previous findings described by Hugenberg et al. (2016) and Cassidy et al. (2022). However, we reasoned this effect might not be specific to human faces but rather represent a more

general phenomenon that applies to any stimulus class with a canonical orientation. In such cases, any exemplar may be judged less prototypical of its category when shown upside-down. Consistent with this view, we found that images of cars and sheep faces were also judged less carlike and sheeplike when inverted.

In these studies, we chose to focus exclusively on the attribution of socially desirable traits. We did this because prominent accounts of dehumanization maintain that socially desirable attributes such as open-mindedness, empathy, and sophistication are central to the concept 'human' and thus most appropriate for measuring dehumanization (Haslam, 2006; Kteily & Landry, 2022; Leyens et al., 2001). However, recent critiques have pointed out that negative attributes such as jealousy, spite, and cunning are also perceived as unique to humans (Enock et al., 2021). In future research, it would be interesting to further test the alternative viewpoint by measuring how inversion influences the attribution of undesirable character traits to faces. The alternative view, developed by Over and Cook (2022) and tested here, predicts that faces which appear trustworthy when upright will be attributed undesirable human qualities to a greater extent when inverted. Faces which appear untrustworthy when upright will be attributed undesirable human qualities to a lesser extent when inverted.

While we have focused on the effects of inversion on the attribution of human character traits, support for PDT has been drawn from a number of different measures including prosocial and antisocial behavioural intentions (Fincher & Tetlock, 2016). PDT predicts that, because inverted faces do not recruit configural processing, participants will be more likely to harm individuals when they view their faces inverted. Consistent with this view, Fincher and Tetlock (2016) found that fictitious criminals received harsher sentences from participants when their faces were shown inverted than when their faces were shown upright. In contrast, Over and Cook (2022) predict that any effects of inversion on punishment will be moderated by the valance of the faces. To the extent that participants are more willing to punish individuals who appear untrustworthy, faces that appear untrustworthy when presented upright may be treated more leniently when inverted.

6.1 Other lines of evidence for PDT

The effects of orientation inversion represent a key line of evidence for PDT. We acknowledge, however, that evidence for the theory has been drawn from others sources as well. In light of our findings, it would be valuable for future research to revisit these other lines of evidence. For example, Fincher and Tetlock (2016) used another common measure of configural processing to measure perceptual dehumanization: the composite face effect. Specifically, the authors found that the faces of dehumanized individuals (e.g., violent criminals) failed to induce the composite face effect. However, a number of methodological and analytical choices complicate interpretation of these results. Crucial variables were not counterbalanced in this study and some unusual analytical choices were made. For example, the authors calculate *d′* sensitivity indices from 32 trials, having employed a 2 × 2 × 2 factorial design – only 4 trials per cell (Over & Cook, 2022). It would be interesting to revisit these findings in light of this critique and the current results.

Another line of evidence for PDT comes from studies utilising image filtering (Fincher & Tetlock, 2016). In their second experiment, Fincher and Tetlock (2016) report that participants endorsed more lenient sentences for criminals depicted in filtered images than in unmodified full-spectrum images. Upright facial images that have had their high-spatial frequency content removed are thought to engage more configural processing (Goffaux & Rossion, 2006). Thus, Fincher and Tetlock (2016) argue that observers experience more empathy for the filtered images than they do for unmodified images and consequently endorse more lenient sentences. However, a plausible alternative explanation of these results is that the spatial filtering gave the faces a smooth and blemish-free appearance rendering the individuals depicted more attractive (Jaeger et al., 2018). It is well-established that attractive faces tend to be positively evaluated on multiple dimensions – the “what is beautiful is good” stereotype (Dion et al., 1972). If this effect is found to be robust, it would be valuable for future research to investigate why participants endorse more lenient

sentencing for the filtered images, and the extent to which this effect is attributable to perceptual dehumanization.

Finally, proponents of PDT often draw support from studies of the so-called 'other-race effect'. There has been much interest in the observation that observers of one ethnicity (e.g., White British) sometimes struggle to individuate the faces of individuals of other ethnicities (e.g., British Chinese) – the so-called other race effect (Furl et al., 2002; Hugenberg et al., 2010; Natu & O'Toole, 2013; Sangrigoli et al., 2005). Some studies have reported that observers who struggle to individuate the faces of a particular ethnicity, show signs of diminished configural processing when viewing similar faces (Hancock & Rhodes, 2008; Michel et al., 2006; Tanaka et al., 2004). Proponents of PDT point out that poor treatment of ethnic outgroup members (consistent with 'dehumanization') together with diminished configural face processing of ethnic out-groups accords well with the predictions of PDT. It is possible that members of ethnic outgroups are dehumanized – in part – because members of the ethnic ingroup fail to process their faces configurally (Hugenberg et al., 2016) or vice versa (Fincher et al., 2017).

Once again, however, this line of evidence is not compelling. First, many authors have failed to observe diminished configural processing of other-race faces (e.g., Mondloch et al., 2010; Wong et al., 2021). Second, if these effects are robust, there is a plausible alternative interpretation, whereby a third factor – a lack of inter-group contact – is responsible for both the prejudice and discrimination experienced by the ethnic outgroup, and diminished configural processing (Over & Cook, 2022). There is considerable evidence that observers' ability to process a stimulus configurally is affected by their perceptual (individuation) experience with that stimulus category (Bukach et al., 2006; Richler, Wong, et al., 2011). If observers do exhibit diminished configural processing of other-race faces, it may reflect the fact that their 'diet of faces' (the range of faces encountered in their daily lives) is relatively limited (Furl et al., 2002; Sangrigoli et al., 2005). Thus, it is possible to explain the poor treatment of ethnic outgroup members together with diminished configural face processing of ethnic outgroups, without hypothesising a causal link between the two.

6.2 Some notes on configural (holistic) face processing

PDT was proposed by authors working at the interface of visual perception and social psychology. Regrettably, some of the ideas drawn from face perception research are presented as incontrovertible fact, or without appropriate nuance within the resulting literature (Over & Cook, 2022). In this closing section, we wish to highlight some relevant issues surrounding configural (or 'holistic') processing that remain controversial in the contemporary face processing literature.

First, proponents of PDT consistently refer to "configural" processing without justification or explanation. Within the face processing literature, however, a distinction is often drawn between "configural" and "holistic" face processing (Piepers & Robbins, 2013). The term "configural processing" is typically used to refer to the representation of the 'second-order' spatial relationships between features; for example, the distance between the eyes, or the between, the eyes, nose, and mouth (e.g., Leder et al., 2001; Maurer et al., 2002). Increasingly, authors use the term "holistic processing" to describe a form of representation in which the features and their spatial relationships are represented as a non-decomposable whole (Farah et al., 1998; McKone et al., 2007). Holistic processing is thought to improve perceptual decisions about individual features (e.g., eyes, nose, mouth) as well as their spatial relationships (Hayward et al., 2016; Tanaka & Farah, 1993; Yovel & Kanwisher, 2008).

Second, there is growing scepticism within the face perception research community that configural (holistic) processing is a single unitary construct, as implied by proponents of PDT. For many years, it was widely assumed that various behavioural effects – including the face inversion effect, the composite face effect, the part-whole effect, and the Thatcher illusion – all index the same phenomenon: either configural or holistic processing (e.g., Farah et al., 1998; Maurer et al., 2002). However, there is growing evidence that these effects do not reflect the operation of a single perceptual mechanism (Chua et al., 2015;

Fitousi, 2016; Psalta et al., 2014; Rezlescu et al., 2017). For example, individual observers' susceptibility to three classic markers of "configural/holistic processing – the face inversion effect, the composite face effect, and the part-whole effect – show little or no correlation (Rezlescu et al., 2017). Thus, configural (holistic) face processing might describe a collection of related perceptual, predictive, and attentional phenomena rather than a single unitary construct. It would be valuable for proponents of PDT to specify their definition of configural processing more precisely in order to enable further critical engagement with their central ideas.

Third, within the PDT literature it is often claimed that inverted faces fail to engage configural (holistic) processing. However, several findings call this assertion into question. For example, evidence from the composite face paradigm suggests that inverted faces do recruit configural (holistic) processing, albeit less strongly (Susilo et al., 2013) or less efficiently (Richler, Mack, et al., 2011) than upright faces. It is also worth noting that both upright and inverted faces produce aperture effects (Murphy & Cook, 2017; Murphy et al., 2020). Where stimuli are processed configurally (holistically), perceptual decisions should be impaired when observers are made to inspect exemplars through a dynamic viewing window that prevents them from seeing all stimulus regions simultaneously. Perceptual decisions that depend on a serial piecemeal analysis should show little or no aperture decrement – piecemeal evidence can be accumulated through the aperture. The fact that inverted faces produce aperture effects is suggestive of some form of configural (holistic) processing (Murphy & Cook, 2017; Murphy et al., 2020). More broadly, several further findings also call into question the view that upright and inverted faces are processed in a qualitatively different manner (e.g., Meinhardt et al., 2019; Sekuler et al., 2004). It is thus unclear whether orientation manipulations offer the neat way to manipulate configural processing that is claimed by proponents of PDT.

Finally, proponents of PDT stress that configural (holistic) processing is selectively engaged by human faces (Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al.,

2016; Young et al., 2019). This is a key assumption of PDT: it is the near-perfect contingency between the presence of a human and the engagement of configural (holistic) processing that is thought to allow the latter to eventually 'signal humanity' (e.g., Hugenberg et al., 2016). However, several behavioural effects attributed to configural (holistic) processing are produced by non-face stimuli (Bukach et al., 2006; Over & Cook, 2022; Richler, Wong, et al., 2011). In particular, a range of non-face stimuli produce composite effects, thought to be a key index of configural (holistic) processing, including fingerprints (Vogelsang et al., 2017), Chinese characters (Hsiao & Cottrell, 2009), words (Wong et al., 2011), and synthetic objects-of-expertise (Chua et al., 2015; Wong et al., 2009). These findings represent a substantial challenge for PDT because it seems extremely unlikely that observers attribute humanity to fingerprints, Chinese characters or synthetic objects-of-expertise.

6.3 Conclusion

Taken together, the results described here suggest that, rather than influencing attributions of humanness specifically, inversion disrupts the ability of observers to derive accurate perceptual descriptions of faces (McKone & Yovel, 2009; Rossion, 2008). Relative to upright facial percepts, the description of local features (e.g., Murphy et al., 2020) and inter-feature spatial relationships (e.g., Leder et al., 2001) may be impoverished in inverted facial percepts. Because inverted faces appear nondescript, a host of trait and character attributions tend towards the average when faces are shown upside-down. In light of these results, a key line of evidence for PDT is considerably less convincing than it first appears. This works accords with broader critiques of the social psychological literature on dehumanization and suggests the need to revisit some of the central claims in this field (Bloom, 2017; Manne, 2016; Over, 2021).

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388–407.

Bäuml, K. H. (1994). Upright versus upside-down faces: How interface attractiveness varies with orientation. *Perception & Psychophysics*, *56*(2), 163-172.

Biotti, F., Gray, K. L. H., & Cook, R. (2019). Is developmental prosopagnosia best characterised as an apperceptive or mnemonic condition? *Neuropsychologia*, *124*, 285–298.

Bloom, P. (2017). The root of all cruelty? *The New Yorker*. https://www.newyorker.com/magazine/2017/11/27/the-root-of-all-cruelty

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: the power of an expertise framework. *Trends in Cognitive Sciences*, *10*(4), 159-166.

Cassidy, B. S., Krendl, A. C., Stanko, K. A., Rydell, R. J., Young, S. G., & Hugenberg, K. (2017). Configural face processing impacts race disparities in humanization and trust. *Journal of Experimental Social Psychology*, *73*, 111–124.

Cassidy, B. S., Wiley, R. W., Sim, M., & Hugenberg, K. (2022). Decoding complex emotions and humanization show related face processing effects. *Emotion*, *22*(2), 362–373.

Chua, K.-W., Richler, J. J., & Gauthier, I. (2015). Holistic processing from learned attention to parts. *Journal of Experimental Psychology: General*, *144*(4), 723–729.

Civile, C., Colvin, E., Siddiqui, H., & Obhi, S. (2019). Labelling faces as 'Autistic' reduces the inversion effect. *Autism*, *23*(6), 1596-1600.

Cook, R., & Duchaine, B. (2011). A look at how we look at others: Orientation inversion and photographic negation disrupt the perception of human bodies. *Visual Cognition*, *19*(4), 445-468.

Cook, R., Eggleston, A., & Over, H. (2022). The cultural learning account of first impressions. *Trends in Cognitive Sciences*, *26*(8), 656-668.

Cook, R., & Over, H. (2021). Why is the literature on first impressions so focused on White faces? *Royal Society Open Science*, *8*(9), e211146.

Deska, J. C., Almaraz, S. M., & Hugenberg, K. (2017). Of mannequins and men: ascriptions of mind in faces are bounded by perceptual and processing similarities to human faces. *Social Psychological and Personality Science*, *8*(2), 183–190.

Deska, J. C., & Hugenberg, K. (2017). The face-mind link: Why we see minds behind faces, and how others' minds change how we see their face. *Social and Personality Psychology Compass*, *11*(12), e12361.

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.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285-290.

Efferson, C., & Vogt, S. (2013). Viewing men's faces does not lead to accurate predictions of trustworthiness. *Scientific Reports*, *3*(1), e1047.

Enock, F. E., Flavell, J. C., Tipper, S. P., & Over, H. (2021). No convincing evidence outgroups are denied uniquely human characteristics: Distinguishing intergroup preference from trait-based dehumanization. *Cognition*, *212*, 104682.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is" special" about face perception? *Psychological Review*, *105*(3), 482-498.

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

Fincher, K. M., Tetlock, P. E., & Morris, M. W. (2017). Interfacing with faces: Perceptual humanization and dehumanization. *Current Directions in Psychological Science*, *26*, 288-293.

Fitousi, D. (2016). Comparing the role of selective and divided attention in the composite face effect: Insights from Attention Operating Characteristic (AOC) plots and cross-contingency correlations. *Cognition*, *148*, 34-46.

Furl, N., Phillips, P. J., & O'Toole, A. J. (2002). Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis. *Cognitive Science*, *26*(6), 797-815.

Goffaux, V., & Rossion, B. (2006). Faces are" spatial"--holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 1023-1039.

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology*, *99*(1), 45-56.

Harris, L. T., & Fiske, S. T. (2011). Dehumanized perception: A psychological means to facilitate atrocities, torture, and genocide? *Journal of Psychology*, *219*, 175-181.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, *10*, 252–264.

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology*, *65*, 399–423.

Hayward, W. G., Crookes, K., Chu, M. H., Favelle, S. K., & Rhodes, G. (2016). Holistic processing of face configurations and components. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(10), 1482-1489.

Hsiao, J. H., & Cottrell, G. W. (2009). Not all visual expertise is holistic, but it may be leftist: The case of Chinese character recognition. *Psychological Science*, *20*(4), 455-463.

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review*, *117*, 1168-1187.

Hugenberg, K., Young, S. G., Rydell, R. J., Almaraz, S., Stanko, K. A., See, P. E., & Wilson, J. P. (2016). The face of humanity: Configural face processing influences ascriptions of humanness. *Social Psychological & Personality Science*, *7*, 167–175.

Jaeger, B., Wagemans, F. M., Evans, A. M., & van Beest, I. (2018). Effects of facial skin smoothness and blemishes on trait impressions. *Perception*, *47*(6), 608-625.

Kteily, N. S., & Landry, A. P. (2022). Dehumanization: trends, insights, and challenges. *Trends in Cognitive Sciences*, *26*(3), 222-240.

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, *4*(1), e2515245920951503.

Leder, H., Candrian, G., Huber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. *Perception*, *30*(1), 73-83.

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*, 395–411.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122-1135.

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

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255-260.

McKelvie, S. J. (1995). Emotional expression in upside-down faces: Evidence for configurational and componential processing. *British Journal of Social Psychology*, *34*(3), 325-334.

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, *11*(1), 8-15.

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review*, *16*(5), 778-797.

Meinhardt, G., Meinhardt-Injac, B., & Persike, M. (2019). Orientation-invariance of individual differences in three face processing tasks. *Royal Society Open Science*, *6*(1), e181350.

Michel, C., Rossion, B., Han, J., Chung, C. S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science*, *17*(7), 608-615.

Mondloch, C. J., Elms, N., Maurer, D., Rhodes, G., Hayward, W. G., Tanaka, J. W., & Zhou, G. (2010). Processes underlying the cross-race effect: An investigation of holistic, featural, and relational processing of own-race versus other-race faces. *Perception*, *39*(8), 1065-1085.

Murphy, J., & Cook, R. (2017). Revealing the mechanisms of human face perception using dynamic apertures. *Cognition*, *169*, 25-35.

Murphy, J., Gray, K. L., & Cook, R. (2020). Inverted faces benefit from whole-face processing. *Cognition*, *194*, 104105.

Natu, V., & O'Toole, A. J. (2013). Neural perspectives on the other-race effect. *Visual Cognition*, *21*(9-10), 1081-1095.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *5*, 11087-11092.

Over, H. (2021). Falsifying the dehumanization hypothesis. *Perspectives on Psychological Science*, *16*(1), 33-38.

Over, H., & Cook, R. (2022). Perceptual dehumanization theory: A critique. *Psychological Review*.

Piepers, D. W., & Robbins, R. A. (2013). A review and clarification of the terms "holistic," "configural," and "relational" in the face perception literature. *Frontiers in Psychology*, *3*(559), 1-11.

Prkachin, G. C. (2003). The effects of orientation on detection and identification of facial expressions of emotion. *British Journal of Psychology*, *94*(1), 45-62.

Psalta, L., Young, A. W., Thompson, P., & Andrews, T. J. (2014). Orientation-sensitivity to facial features explains the Thatcher illusion. *Journal of Vision*, *14*(12), e9.

Rezlescu, C., Susilo, T., Wilmer, J. B., & Caramazza, A. (2017). The inversion, part-whole, and composite effects reflect distinct perceptual mechanisms with varied relationships to face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(12), 1961-1973.

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research*, *51*(3), 333-342.

Richler, J. J., Wong, Y. K., & Gauthier, I. (2011). Perceptual expertise as a shift from strategic interference to automatic holistic processing. *Current Directions in Psychological Science*, *20*(2), 129-134. https://www.ncbi.nlm.nih.gov/pubmed/21643512

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica*, *128*(2), 274-289.

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, *16*(6), 440-444.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, *14*(5), 391-396.

Smith, D. L. (2016). Paradoxes of Dehumanization. *Social Theory and Practice*, *42*(2), 416–443.

Smith, D. L. (2020). *On inhumanity: Dehumanization and how to resist it*. Oxford University Press.

Susilo, T., Rezlescu, C., & Duchaine, B. (2013). The composite effect for inverted faces is reliable at large sample sizes and requires the basic face configuration. *Journal of Vision*, *13*(13), e14.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *46*(2), 225-245.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition*, *93*(1), B1-B9.

Thompson, P., & Wilson, J. (2012). Why do most faces look thinner upside down? *i-Perception*, *3*(10), 765-774.

Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different facial images of the same person. *Psychological Science*, *25*(7), 1404-1417.

Vogelsang, M. D., Palmeri, T. J., & Busey, T. A. (2017). Holistic processing of fingerprints by expert forensic examiners. *Cognitive Research: Principles and Implications*, *2*, e15.

Wilson, J. P., Young, S. G., Rule, N. O., & Hugenberg, K. (2018). Configural processing and social judgments: Face inversion particularly disrupts inferences of human-relevant traits. *Journal of Experimental Social Psychology*, *74*(1), 1-7.

Wong, A. C. N., Bukach, C. M., Yuen, C., Yang, L., Leung, S., & Greenspon, E. (2011). Holistic processing of words modulated by reading experience. *PloS One*, *6*(6), e20753.

Wong, A. C. N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for facelike expertise with objects: Becoming a Ziggerin expert—but which type? *Psychological Science*, *20*(9), 1108-1117.

Wong, H. K., Estudillo, A. J., Stephen, I. D., & Keeble, D. R. (2021). The other-race effect and holistic processing across racial groups. *Scientific Reports*, *11*(1), 1-15.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141-145.

Young, S. G., Tracy, R. E., Wilson, J. P., Rydell, R. J., & Hugenberg, K. (2019). The temporal dynamics of the link between configural face processing and dehumanization. *Journal of Experimental Social Psychology*, *85*, 103883.

Yovel, G., & Kanwisher, N. (2008). The representations of spacing and part-based information are associated for upright faces but dissociated for objects: Evidence from individual differences. *Psychonomic Bulletin & Review*, *15*(5), 933-939.