# Perceptual Dehumanization Theory: A critique

**Harriet Over[1] and Richard Cook[2]**

[1] Department of Psychology, University of York, UK, harriet.over@york.ac.uk, https://www.york.ac.uk/psychology/staff/academicstaff/ho543/

[2] Department of Psychology, Birkbeck, University of London, richard.cook@bbk.ac.uk, https://www.bbk.ac.uk/our-staff/profile/9153462/richard-cook

**In press at Psychological Review**

**Acknowledgments:** We would like to thank Florence Enock, Adam Eggleston and Steve Tipper for valuable comments on an earlier draft.

**Funding statement:** This research was supported by the European Research Council under the European Union's Horizon 2020 Programme, grant numbers ERC-STG-755719 and ERC-2016-StG-715824.

**Abstract**

Central to Perceptual Dehumanization Theory is the claim that full engagement of a putative module for the visual analysis of faces is necessary in order to recognize the humanity or personhood of observed individuals. According to this view, the faces of outgroup members do not engage domain-specific face processing fully or typically and are instead processed in a manner akin to how the brain processes objects. Consequently, outgroup members are attributed less humanity than are ingroup members. To the extent that groups are perceptually dehumanized, they are hypothesised to be vulnerable to harm. In our article, we challenge several of the fundamental assumptions underlying this theory and question the empirical evidence in its favour. We begin by illustrating the extent to which the existence of domain-specific face processing is contested within the vision science literature. Next, we interrogate empirical evidence that appears to support Perceptual Dehumanization Theory and suggest that alternative explanations for prominent findings in the field are more likely. In the closing sections of the paper, we reflect on the broader logic of the theory and highlight some underlying inconsistencies.

**Key words:**

Configural processing; Dehumanization; Discrimination; Face perception; Minimal group paradigm; Perceptual Dehumanization Theory; Other-race effect.

## 1. Introduction

Discrimination is a powerful and destructive force. Members of marginalised groups face systematic disadvantage in social interaction, employment, health, housing, and the law (Over & McCall, 2018; Pascoe & Smart Richman, 2009; Richeson, 2020). At a global level, intergroup bias fuels wars within and between countries and prevents cooperation on pressing issues of common concern (Bliuc et al., 2015; Smith, 2020). Given the serious consequences of discrimination, it is imperative to understand its nature and origins. A particularly influential idea within the social sciences is that a psychological process of dehumanization is one important cause of discrimination (Harris & Fiske, 2011; Haslam, 2006; Smith, 2016). According to dehumanization theories, victims of intergroup harm are perceived as similar to nonhuman entities. As a result, they are rendered more vulnerable to harm (Harris & Fiske, 2006; Haslam, 2006; Haslam & Loughnan, 2014; Smith, 2020).

Within the broad field of dehumanization, one influential line of research hypothesises that dehumanization can take the form of a 'bottom up' perceptual process (Cassidy et al., 2017; Deska & Hugenberg, 2017; Fincher & Tetlock, 2016; Fincher, Tetlock, & Morris, 2017; Hugenberg et al., 2016; Wilson, Young, Rule, & Hugenberg, 2018; S. G. Young, Tracy, Wilson, Rydell, & Hugenberg, 2019). According Perceptual Dehumanization Theory (PDT), human faces recruit specialised visual processing distinct from that engaged by objects. In particular, faces are thought to engage 'configural processing' (sometimes referred to as 'holistic processing') whereby local facial features are integrated into a coherent unified percept (Deska & Hugenberg, 2017; Fincher et al., 2017). To the extent that outgroups are perceptually dehumanized, their faces and bodies are thought to be processed in a manner akin to the way in which the visual system processes objects and non-human animals (Deska & Hugenberg, 2017; Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al., 2016). When groups are perceptually dehumanized, they are thought to be at greater risk of harm (Deska & Hugenberg, 2017; Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al., 2016).

In the present paper, we assess the plausibility of PDT. We argue that paradigms central to the study of perceptual dehumanization contain significant confounds and that what appears to be evidence for perceptual dehumanization may, in fact, be better explained by other processes. We first provide an overview of PDT and delineate strong and weak versions of the theory (Section 2). Next, we consider why some members of the vision science community contest the fundamental premise of PDT - the existence of domain-specific face processing (Section 3). Following this, we examine whether extant evidence supports the two central claims of the theory (Sections 4 and 5). Finally, we comment on the broader logical structure and plausibility of PDT (Section 6).

**2. What is Perceptual Dehumanization Theory?**

PDT seeks to explain how biases in human perception may contribute to intergroup harm. The theory lies at the interface between social psychology and visual perception and draws its assumptions from these two fields. Drawing on the face perception literature (e.g., Haxby, Hoffman, & Gobbini, 2000; Kanwisher, 2000; McKone, Kanwisher, & Duchaine, 2007), PDT asserts that human faces typically recruit specialised visual processing mediated by dedicated neural substrates (Fincher et al., 2017; Hugenberg et al., 2016). In particular, human faces are thought to engage configural processing (Deska & Hugenberg, 2017; Fincher et al., 2017; Hugenberg et al., 2016), whereby different facial regions are processed in parallel and integrated into a unified percept (Farah, Wilson, Drain, & Tanaka, 1998; Maurer, Le Grand, & Mondloch, 2002; Piepers & Robbins, 2013). In contrast, the faces of non-human animals and objects are thought to engage different regions of the visual brain and recruit a serial parts-based analysis (Deska & Hugenberg, 2017; Fincher et al., 2017; Hugenberg et al., 2016).

When faces fail to engage domain-specific face processing typically, they are processed in a similar way to objects and non-human animals, and consequently, dehumanized. For example, rather than being processed configurally (like ingroup faces), the faces of outgroup members are thought to be processed in a serial piecemeal fashion, like objects and animals (Cassidy et al., 2017; Deska & Hugenberg, 2017; Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al., 2016; Wilson et al., 2018). The argument at the heart of PDT is summarized in the following quotes:

> *"Specifically, 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"* (Hugenberg et al., 2016, p168).
>
> *"This domain specificity raises the interesting possibility that encoding the configuration of face parts conveys information beyond identity and emotional states. Indeed, the experience of processing a face configurally may be a bottom-up signal that we are interacting with another person who possesses fundamental capacities to think, feel, and act in human ways (…). If true,*

*interfering with configural face encoding should lead to diminished capacity to recognize that faces belong to fellow humans, rather than objects or non-human animals"* (Young et al., 2019, p1).

Proponents of PDT operationalise the failure to attribute humanity in different ways. Consistent with the Dual Model of Dehumanization (Haslam, 2006), some studies of PDT infer dehumanization from a failure to attribute uniquely human character traits such as consideration and creativity to the outgroup (Hugenberg et al., 2016; Wilson et al., 2018). Consistent with previous work on Infrahumanization (Leyens et al., 2001), other studies of PDT infer dehumanization from a failure to attribute uniquely human emotions such as nostalgia and guilt to the outgroup (Cassidy, Wiley, Sim, & Hugenberg, 2021). Consistent with the Mental State Model (Harris & Fiske, 2006), other work on PDT infers dehumanization from a failure to attribute mental states to the outgroup (Deska & Hugenberg, 2017). Consistent with philosophical work on dehumanization (Smith, 2011, 2020), some PDT authors allude to a failure attribute a human essence (Hugenberg et al., 2016; S. G. Young et al., 2019).

Regardless of how the failure to attribute humanity is operationalised by different authors, work on PDT shares the assumption that, to the extent that groups are perceptually dehumanized, they are less likely to be the recipients of prosocial behaviour and more likely to be the victims of harm (Deska & Hugenberg, 2017; Fincher & Tetlock, 2016; Fincher et al., 2017). All social psychological theories of dehumanization have implications for interpersonal behaviour. Importantly, however, PDT makes unique predictions. For example, PDT predicts that people who exhibit atypical face processing are more likely to harm others because they fail to fully recognise their humanity (Hugenberg et al., 2016). Similarly, PDT predicts that we are more likely to harm others in situations that prevent typical face processing. For example, orientation inversion is thought to impede configural face processing, leading observers to be more reliant on serial parts-based processing also engaged by non-social objects (Farah et al., 1998; McKone & Yovel, 2009; Rossion, 2008). Thus, when we encounter someone in an unusual (non-canonical) orientation, PDT predicts that we will attribute less humanity to them and be more likely to cause them harm (Deska & Hugenberg, 2017; Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al., 2016).

Importantly, PDT cannot be reduced to the relatively uncontroversial observation that a configural analysis yields a more accurate perceptual description of a target face than does a serial parts-based analysis (Farah et al., 1998; McKone & Yovel, 2009; Rossion, 2008). The key claim of PDT relates to the way in which the perceptual description is derived, not its quality *per se*. If perceptual dehumanization was determined by the accuracy of the perceptual description, this theory would also predict that we attribute less humanity to others when we view them without our glasses (vs. when our visual acuity is corrected) or in poor light conditions (vs. in a well-lit environment). This is clearly not the argument being made. Rather, PDT asserts that the means by which the representation is derived is critical. Specifically, the attribution of humanity is determined by the extent to which domain-specific mechanisms of visual face processing (e.g., configural analysis) are recruited. To illustrate this distinction, it is useful to consider a hypothetical observer who is an expert at serial parts-based processing. Such a person might be able to encode local regions in a virtually noiseless manner and combine information obtained from a serial analysis in such a way as to optimize perceptual decision making. Within the framework endorsed by proponents of PDT, these skills might make this observer an object identification savant. Such an observer would be expected to derive highly accurate perceptual descriptions of inverted faces, encoding subtle differences in facial structure, gaze direction, and expression. Nevertheless, PDT predicts that this observer would still dehumanize inverted faces because the perceptual description, although highly accurate, would be derived from generic, parts-based object processing (Fincher et al., 2017; Hugenberg et al., 2016).

*2.1 Strong and weak versions of the theory*

At least two different versions of PDT can be delineated. We will refer to these as the strong form and the weak form. In its strong form, PDT asserts that dehumanized faces (e.g., those of ethnic outgroup members, members of stigmatised groups, and abhorrent criminals) fail to engage domain-specific face processing mechanisms. Instead, they recruit generic piecemeal processing that is also engaged when viewing objects and non-human animals. Whether a face engages one mode of processing or the other is determined by our knowledge of the individual, their group identity, and the context (e.g., Fincher et al., 2017). This strong form of PDT is evident in the following quotes from Fincher et al. (2017):

> *“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"* (Fincher et al., 2017, p288).

*"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"* (Fincher et al., 2017, p289).

References to the strong form of PDT are commonplace in the wider dehumanization and social psychological literatures. To give one recent example, the following is taken from a review of the dehumanization literature published in *Trends in Cognitive Sciences* (Kteily & Landry, 2022):

*"Other recent work introduces a role for perceptual processes in dehumanization, suggesting that we sometimes literally overlook our counterparts' humanity, processing their faces piecemeal (as we would an object) rather than holistically (as we would with humans […]). This subtly dehumanizing mode of perception occurs in a variety of contexts (e.g., when encountering norm violators […]) and predicts downstream consequences like moralistic punishment […]"* (Kteily & Landry, 2022, p223).

According to the weak form of PDT, dehumanized faces engage face-specific processing, but do so to a lesser extent than ingroup faces. For example, the faces of certain outgroup individuals are hypothesized to engage less configural processing than the faces of ingroup members (Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al., 2016). Whereas the strong form of the theory implies a qualitative difference between the visual processing engaged by 'humanized' and 'dehumanized' faces, the weaker form implies a quantitative difference. This weak form of PDT is evident in the following quotes:

*"The results of the studies presented suggest that the faces of African Americans and members of out-groups and low-status groups are processed less configurally and that these individuals are more often subject to harm and mistreatment"* (Fincher et al., 2017, p290).

*“Thus, our hypotheses begin with the observation that configural processing is attenuated for racial out-groups and for members of stigmatized groups—the very groups who are likely to be dehumanized in naturalistic contexts”* (Hugenberg et al., 2016, p169).

It is important to note that a qualitative distinction between the processing engaged by human faces and that engaged by objects and non-human animals is fundamental to *both* the strong and weak versions of PDT. The existence of domain-specific face processing is the key premise on which all work on PDT rests. The close contingency between the domain-specific visual processing engaged by faces and the presence of another human being is thought to allow the recruitment of the former to provide a signal of humanity. The strong form of PDT asserts that domain-specific face processing is not engaged by outgroup faces. According to the weak form of PDT, outgroup faces engage less domain-specific face processing. In other words, the difference between the strong and weak forms of the theory relates to what happens within a modular architecture, not whether the architecture itself is modular (Figure 1).

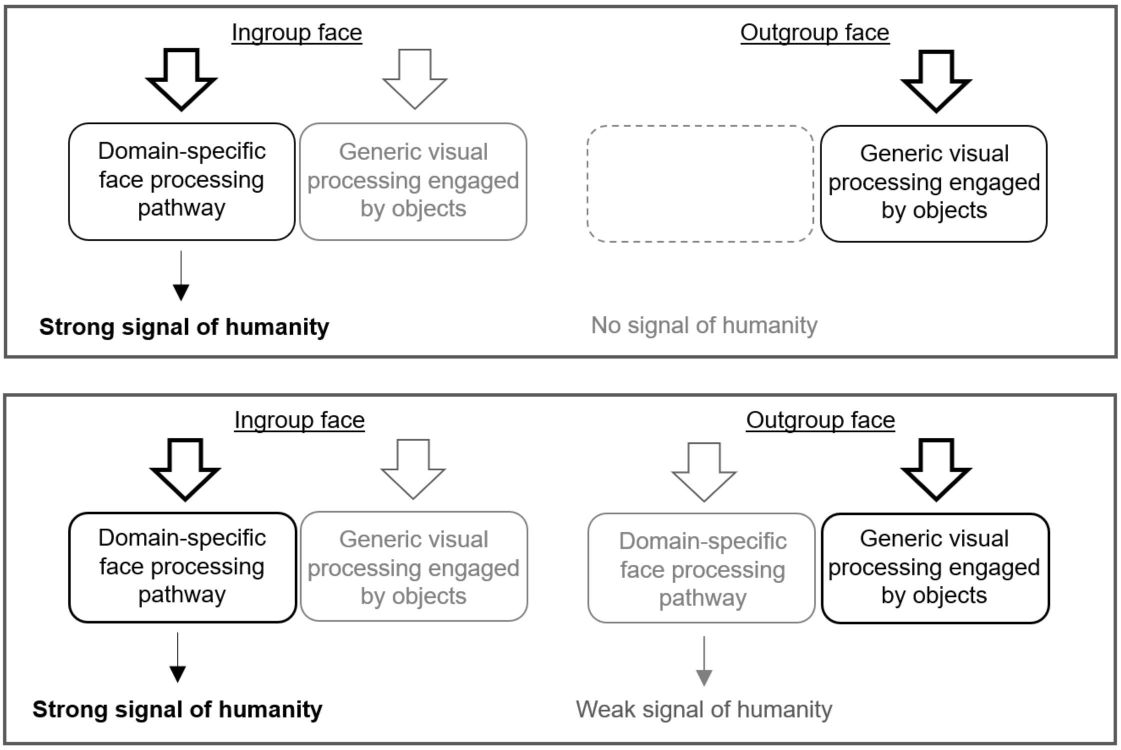


**Figure 1.** Schematic illustration of the strong (top) and weak (bottom) versions of Perceptual Dehumanization Theory. According to the strong form of the theory, dehumanized faces (e.g., those of outgroup members) do not engage domain-specific mechanisms of face processing. According to the weak form of the theory, dehumanized faces engage domain-specific mechanisms of face processing but to a lesser degree. Note, both versions are predicated on the view that faces engage domain-specific visual processing

*2.2 Other aspects of the theory that require clarification*

The foregoing discussion of strong and weak forms of PDT reveals some uncertainty as to whether dehumanized faces fail to engage domain-specific face processing (strong form), or do so to a lesser degree than ingroup faces (weak form). Below, we highlight four further aspects of PDT that also require clarification.

The first is the nature of the front-end processing that determines whether or not a face engages or continues to engage domain-specific processing. It is widely agreed within the PDT literature that the faces of members of ethnic outgroups and members of stigmatized groups fail to fully engage domain-specific face processing (Fincher et al., 2017; Hugenberg et al., 2016). According to some authors, this is also the case for the faces of abhorrent criminals (Fincher & Tetlock, 2016; Fincher et al., 2017). In each case, it would appear that some face processing is required in order to determine the identity and/or group membership of the individual depicted. In particular, one might assume that the full engagement of typical face processing – including a configural analysis – would be required in order to disengage or attenuate domain-specific face processing on the basis of person-specific knowledge (e.g., that the individual depicted is a violent criminal). The extent of this front-end processing and its implications for the attribution of humanity require clarification. Such clarification may elucidate differences between the particular versions of PDT endorsed by leading authors in field.

A second and somewhat related issue, is the extent to which perceptual dehumanization is under observers' control. At points in this literature, perceptual dehumanization is said to be "bottom-up" (Hugenberg et al., 2016) and the result of a "rapid-fire perceptual categorization" (Fincher et al., 2017). This characterisation suggests that perceptual dehumanization occurs automatically where faces are identified as outgroup. Consistent with this suggestion, proponents of PDT assert that that domain-specific face processing is reserved for ingroup faces (Hugenberg et al., 2016). Similarly, the faces of members of minimal outgroups (i.e., people identified as outgroup, but for whom there is no additional incentive to inflict harm) are thought to be subject to perceptual dehumanization (e.g., Fincher et al., 2017). Elsewhere in the literature, however, proponents of PDT imply that observers can dehumanize others intentionally where they wish to inflict harm (Deska & Hugenberg, 2017; Fincher et al., 2017); for example, observers may be able to intentionally disengage or attenuate the configural face processing of outgroup faces (Fincher et al., 2017). This suggestion is at odds with the prevailing view in the face processing literature that the configural processing of upright faces is obligatory – that it cannot be prevented even in situations where it impairs perceptual decision-making (Murphy & Cook, 2017; Rossion, 2013).

A third aspect of the theory that requires clarification is the extent to which engagement of generic piecemeal processing provides a positive signal of "objecthood". One possibility is that, because outgroup faces recruit piecemeal "object processing", they are perceived as being similar to objects (e.g., as lacking animacy). There is some precedence for this interpretation in previous work on mechanistic dehumanization in which certain groups are thought to share attributes in common with robots (Haslam, 2006) and in work on the objectification of sexualised women (Bernard, Gervais, Allen, Campomizzi, & Klein, 2012). Another possibility is that, although outgroup faces recruit a form of processing commonly used for objects, they are not seen as having the qualities of objects themselves. Compatible with this latter possibility, proponents of PDT assume that non-human animals, as well as inanimate objects, engage generic piecemeal processing (Hugenberg et al., 2016; S. G. Young et al., 2019). Clarification of the downstream consequences of "object processing" is important for understanding how PDT relates to other social psychological theories of dehumanization.

Fourth, it is unclear whether the engagement of domain-specific face processing mechanisms is necessary in order to fully attribute humanity to others. According to PDT, faces that afford typical domain-specific face processing (e.g., the faces of ingroup members) are attributed more humanity than faces that do not (e.g., inverted faces, the faces of outgroup members). It is less clear whether or not it is possible to fully attribute humanity in situations where no face is visible (e.g., when we see feet protruding from behind a curtain or when we hear a child refugee crying on the radio). One possibility is that humanity can only be fully attributed when domain-specific face processing is recruited. Another possibility is that humanity may be fully attributed when no face is visible via an alternative means – for example, via a more flexible cognitive route. However, if typical engagement of domain-specific face processing is not deemed necessary (i.e., if there are other routes through which we may attribute humanity to others), then proponents of PDT must explain our apparent failure to attribute humanity to inverted faces and the faces of outgroup members. That is, why we are unable to compensate for the lack of typical face processing in these situations.

## 3. The existence of domain-specific face processing is hotly contested

Both the strong and weak forms of PDT are predicated on the assumption that faces typically engage domain-specific processing not recruited by objects and non-human animals. This feature is crucial as it allows the authors of PDT to distinguish 'humanizing' and 'dehumanizing' modes of perception. There are certainly members of the vision science community who would endorse the view that faces engage domain-specific visual processing (e.g., configural or holistic representation) and dedicated neural substrates (Duchaine & Yovel, 2015; Haxby et al., 2000; Kanwisher, 2000; Kanwisher & Yovel, 2006; McKone et al., 2007; Robbins & McKone, 2007; Rossion, 2013; Rossion, Jacques, & Jonas, 2018). However, the relationship between the processing of faces and objects has been the subject of intense debate within the vision science community over the last forty years, and many authors contest the view that faces engage domain-specific processing (Bukach, Gauthier, & Tarr, 2006; Gauthier & Bukach, 2007; Richler, Wong, & Gauthier, 2011; Tarr & Gauthier, 2000). When introducing PDT, however, proponents frequently cite the face perception literature selectively, and fail to acknowledge the extent to which the existence of domain-specific face processing is debated.

Below, we briefly discuss some of the findings that have cast doubt on the domain-specificity of face processing – results that have prompted some authors to reject the idea of domain-specific face processing. Although clear, unambiguous findings are few and far between, there appears to be considerable overlap between the perceptual mechanisms engaged by faces and objects. In particular, many of the effects characterised as 'markers' of face-specific processing are also produced by non-face "objects of expertise"; i.e., stimulus classes that comprise numerous exemplars, share a prototypical arrangement and orientation, and with which participants have extensive individuation experience (Bukach et al., 2006; Richler et al., 2011). As will become clear, regardless of which side of the debate opinion falls, the discussion raises awkward questions for proponents of PDT.

### *3.1 The face inversion effect*

One source of evidence that faces are processed in a qualitatively different fashion from objects comes from the face inversion effect. It is well-established that faces are processed more accurately and with greater efficiency when shown upright than when shown upside-down (Derntl, Seidel, Kainz, & Carbon, 2009; Murphy & Cook, 2017; Murphy, Gray, & Cook, 2020; Prkachin, 2003; Schwaninger, Lobmaier, & Fischer, 2005; Yin, 1969). This is effect is sometimes called the disproportionate face inversion effect because the performance decrement induced by inversion is greater for faces than for most other visual stimuli (e.g., McKone et al., 2007). Where observed, substantial decrements induced by orientation inversion are often taken as an indirect measure of configural processing (Maurer et al., 2002; McKone & Yovel, 2009; Rossion, 2008). The received narrative is that upright faces are processed configurally, a type of analysis that affords fast, accurate stimulus encoding. When turned upside-down, however, faces no longer engage configural processing. Instead, observers must rely on a feature-by-feature analysis of the target face – a strategy that is slow, effortful, and prone to error. Performance decrements seen when stimuli are inverted are therefore assumed to reflect the switch from configural to piecemeal processing, and are used to index the relative advantage conveyed by the former (Maurer et al., 2002; McKone & Yovel, 2009; Rossion, 2008).

Importantly, however, substantial inversion effects are also seen with certain non-face objects. For example, observers who have been trained to discriminate Greebles (Figure 2a; a class of synthetic object with a prototypical configuration and canonical orientation) appear to develop perceptual expertise that augments the individuation of upright exemplars but not inverted exemplars (Ashworth, Vuong, Rossion, & Tarr, 2008; Gauthier & Tarr, 1997). Similarly, following extensive experience with checkerboard patterns, checkerboard experts show substantial inversion effects when identifying exemplars (Civile et al., 2014; McLaren, 1997). Convergent findings have been obtained with studies of naturally occurring perceptual expertise. For example, dog experts show substantial inversion effects when identifying dogs (Diamond & Carey, 1986), budgerigar experts show substantial inversion effects when identifying budgerigars (Campbell & Tanaka, 2018), and expert radiographers show substantial inversion effects when interpreting mammograms (Chin, Evans, Wolfe, Bowen, & Tanaka, 2018).

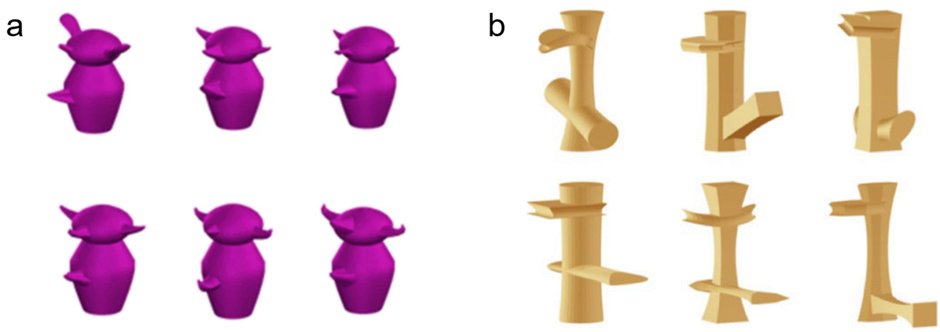


**Figure 2.** The perceptual processing engaged by so-called "objects-of-expertise" has been studied using sets of synthetic objects, including (a) Greebles, and (b) Ziggerins.[1]

*3.2 Composite effects*

A second putative marker of face-specific processing is the composite effect (Abbas & Duchaine, 2008; Calder, Young, Keane, & Dean, 2000; Gray et al., 2020; Hole, 1994; A. W. Young, Hellawell, & Hay, 1987). Composite face paradigms pair the top half of one face with the bottom half of another face. Studies using this approach have revealed that perceptual decisions about one face region (e.g., the top half) are biased by the content of the other region (e.g., the bottom half). As a result, when different composite arrangements pair the same top half with different bottom halves, the two top halves appear to differ (Figure 3a). Interestingly, this effect is not seen when composite arrangements are shown upside-down or when there is a horizontal offset between the top and bottom halves (for reviews, see: Murphy, Gray, & Cook, 2017; Rossion, 2013). Once again, however, similar effects are seen with non-face stimuli with which participants have acquired perceptual expertise, including words (Anstis, 2005; A. C. N. Wong et al., 2011), Chinese characters (Hsiao & Cottrell, 2009; A. C. N. Wong et al., 2012), fingerprints (Vogelsang, Palmeri, & Busey, 2017), and synthetic 'Ziggerin' objects (Figure 2b; A. C. N. Wong, Palmeri, & Gauthier, 2009).

*3.3 Thatcher effects*

Further evidence for the existence face-specific processing has been drawn from the Thatcher effect. Faces are said to be "Thatcherized" when the eye and mouth regions are inverted and superimposed back on the original (unaltered) face context at their canonical locations (Figure 3b; Thompson, 1980). The key finding obtained with Thatcherized faces is that participants are less able to detect the image manipulation when stimuli are shown upside-down than when shown upright (Donnelly et al., 2011; Thompson, 1980; Utz & Carbon, 2016). The Thatcher effect has been taken by some as evidence of configural face processing (e.g., Lewis & Johnston, 1997). Once again, however, comparable effects are produced by a range on non-face stimuli, including cars, buildings, bikes, letter strings (Y. K. Wong, Twedt, Sheinberg, & Gauthier, 2010) and maps (Johnston, Baker, Stone, & Kaufman, 2014).


**Figure 3.** (a) When different composite face arrangements pair the same top half with different bottom halves, the top halves appear to differ (left). However, this effect is not seen when composite arrangements are shown upside-down (right). (b) It is easier to tell when a face has been "Thatcherized" when the wider face context is upright (left) than inverted (right).[2]

*3.4 N170 event-related potential*

Neural data has also been used to support the claim that face processing is qualitatively different from object processing. The N170 is an electrophysiological marker seen during the visual processing of faces (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Eimer, 1998, 2000; Eimer, Kiss, & Nicholas, 2010; Towler & Eimer, 2015). Specifically, it is a negative event related potential (ERP) detected via electroencephalography (EEG), that occurs ~170 ms after the onset of a face stimulus. The magnitude of the N170 is delayed when faces are turned upside-down (Rossion et al., 2000). Again, however, the evidence is more equivocal than it first appears. The N170 component appears to be produced by a range of non-face stimuli including cars and butterflies (Thierry, Martin, Downing, & Pegna, 2007). Similarly, Greeble experts show an N170 when viewing Greebles (Rossion et al., 2000) and expert fingerprint identifiers exhibit an N170 when viewing fingerprints (Busey & Vanderkolk, 2005). Interestingly, similar to findings with faces, the N170 seen in fingerprint experts is also delayed by stimulus inversion (Busey & Vanderkolk, 2005).

*3.5 Activation of the Fusiform Face Area*

Neuroimaging experiments have revealed various regions in the human visual system that are involved in the perception and recognition of faces (Duchaine & Yovel, 2015; Haxby et al., 2000). Perhaps the best known of these regions is the Fusiform Face Area (FFA; Ganel, Valyear, Goshen-Gottstein, & Goodale, 2005; Golby, Gabrieli, Chiao, & Eberhardt, 2001; Hoffman & Haxby, 2000; Kanwisher, McDermott, & Chun, 1997; Kanwisher & Yovel, 2006). It has been argued that the FFA may form part of an innate face processing module (e.g., Kanwisher, 2000). Despite its characterisation as a “face” area, however, this region also responds strongly when Greeble experts view Greebles (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999), when ornithologists view birds (Gauthier, Skudlarski, Gore, & Anderson, 2000), when car enthusiasts view cars (Gauthier et al., 2000; Xu, 2005), when chess experts view chess boards (Bilalić, Langner, Ulrich, & Grodd, 2011) and when expert radiologists view chest X-rays (Bilalić, Grottenthaler, Nägele, & Lindig, 2016).

*3.6 Neuropsychology*

Neuropsychological patients sometimes present with a pattern of deficits consistent with the view that face processing and object processing can be fractionated. For example, patient “CK” exhibited broadly typical face recognition despite severe object recognition deficits (Moscovitch, Winocur, & Behrmann, 1997). Conversely, patient “Herschel” exhibits impaired face recognition, but broadly typical object recognition and individuation (Rezlescu, Barton, Pitcher, & Duchaine, 2014). More typically, however, neuropsychological patients (e.g., Barton, 2008; Barton, Albonico, Susilo, Duchaine, & Corrow, 2019) and developmental cases (Geskin & Behrmann, 2018) who present with impaired face recognition also show signs of impaired object individuation. For example, individuals with developmental prosopagnosia often have difficulty individuating non-face objects including cars and guns (Biotti, Gray, & Cook, 2017; Duchaine, Germine, & Nakayama, 2007; Gray, Biotti, & Cook, 2019). This pattern suggests some overlap between the neurocognitive mechanisms engaged by faces and objects.

### *3.7 Interim summary*

Both the strong and weak forms of PDT are predicated on the view that faces engage domain-specific visual processing that is not recruited by objects and non-human animals (Figure 1). This feature is crucial because it allows proponents to distinguish humanizing and dehumanizing perception. For example, if only faces are processed configurally, the engagement of configural processing may serve as a signal for humanity (Deska & Hugenberg, 2017; Hugenberg et al., 2016; S. G. Young et al., 2019). To date, the authors of PDT have failed to acknowledge the extent to which this key premise is contested within the vision science literature. The view that faces engage domain-specific processing is frequently presented as a matter of fact, rather than as an assumption that disregards one of the most hotly contested debates in cognitive science.

We do not necessarily favour one view (e.g., that faces engage domain-specific processing) over the other (e.g., what appears to be face-specific processing is actually domain-general processing engaged by objects of expertise). Nevertheless, the foregoing behavioural, neuroscientific and neuropsychological findings suggest considerable overlap between the visual processing of faces and objects. In particular, many effects characterised as markers of face processing – including effects used to evidence configural processing (e.g., substantial inversion effects, composite face effects, and Thatcher effects) – are elicited by non-social objects of expertise. In light of these findings, it is unclear whether it is possible to delineate ‘humanizing’ from ‘dehumanizing’ modes of visual processing. Furthermore, findings like these pose awkward questions for proponents of PDT. For example, where observers process maps, cars, bicycles, letter strings, and fingerprints configurally, do they also attribute humanity to these stimuli? To date, proponents of PDT have not engaged with these questions.

In the coming sections, we discuss empirical evidence for the two central claims of PDT - that group membership modulates face-specific processing and that modulation of configural processing affects attributions of humanity.

## 4. Does group membership modulate face-specific processing?

According to PDT, the faces of outgroup members do not fully engage domain-specific face processing and are instead processed in a manner akin to how the brain processes objects. In an attempt to evidence this claim, proponents of PDT cite studies which have manipulated the group membership of the people depicted in stimulus images and examined the extent to which these manipulations influence measures of face processing.

### *4.1 Findings from groups created in the lab*

Authors have sought to examine the effects of group membership and social identity on face processing by employing so-called minimal group paradigms, in which participants are allocated to different groups on a random or arbitrary basis (Tajfel, 1970). Having randomly assigned participants to one of two groups based on their “numerical estimation style” (a fictitious construct), Ratner and Amodio (2013) found that ingroup faces elicited a larger N170 than outgroup faces. Similarly, having randomly assigned participants to one of two teams (the Leopards and the Tigers), Van Bavel et al. (2008) found that FFA responses were greater when participants viewed the faces of their minimal ingroup than those of their minimal outgroup. In a follow-up study Van Bavel et al. (2011) found that the difference in FFA response seen to ingroup and outgroup faces reflected greater activation to ingroup faces rather than reduced activation to outgroup faces. While not necessarily designed to test PDT, these findings have been cited in support of it (e.g., Fincher et al., 2017). These results are consistent with the view that outgroup faces engage typical mechanisms of face processing less than do ingroup faces.

There is reason to question this interpretation, however. It is well known that neural responses within face processing areas exhibit attentional modulation. For example, Bird and colleagues (2006) asked participants to view a pair of faces and a pair of houses presented simultaneously. On some trials participants were told to attend to the faces, on other trials they were told to attend to the houses. On attend-face trials, a significant increase in signal was seen in FFA, while on attend-house trials, a significant increase in signal was seen in parahippocampal place area. Similarly, Hoffman and Haxby (2000) investigated the response of FFA and posterior superior temporal sulcus (pSTS) in a task that required participants to selectively orient their attention to either facial identity or eye gaze. They observed stronger

FFA responses when participants attended to identity and stronger pSTS responses when participants attended to gaze. In light of these findings, it seems plausible that minimal group manipulations modulate neural markers of face processing by directing participants' attention to a particular set of faces or particular attribute within those faces.

There is also evidence that group membership can modulate behavioural markers of face processing. For example, Civile et al. (2018) manipulated the perceived group membership of otherwise identical faces by labelling them as "Autistic" or "Regular" during a study phase (these are the verbatim labels employed in the study). The authors examined the effects of this manipulation on the ability of neurotypical participants to identify the individuals depicted in a subsequent old-new task, when their faces were presented upright and inverted. Participants showed smaller inversion effects for those faces labelled as "Autistic". Civile et al. (2018) suggest that, unlike the "Regular" faces, the faces labelled as "Autistic" were processed in a piecemeal fashion during the study phase, hence stimulus inversion had little impact on recognition at test.

An alternative interpretation seems at least as likely, however. Recognition of the "Autistic" and "Regular" faces was similar when the stimuli were inverted (in both cases, performance approached chance levels). The apparent difference in susceptibility to inversion was thus driven by superior recognition of upright faces labelled "Regular" relative to the faces labelled "Autistic". Once again, this finding can be easily understood in terms of attentional modulation. Autistic individuals are frequently characterised as cold, unemotional, and unempathetic in the popular media (Yergeau, 2013). As a result of the prevailing stereotype, the presence of the "Autistic" label may have led participants to attend to the expression of the person depicted rather than their identity, yielding relatively poor recognition in the upright condition of the old-new test. Indeed, it is possible that participants paid less attention to the Autistic stimuli overall.

*4.2 The influence of person-specific knowledge*

A particularly influential paper has argued that the faces of norm violators and criminals engage less configural processing than do the faces of norm-following ingroup members (Fincher & Tetlock, 2016). In their first study, Fincher and Tetlock examined participants' ability to learn pairings between individual faces and various positive actions (e.g., cared for sick friend), neutral actions (e.g., took clothing to a local dry cleaner), and serious norm violations/crimes (e.g., kidnapped and held someone hostage for a year). Faces were presented either upright or upside-down (both during training and at test). Overall, face-action learning was poorer for upside-down faces. When given an action at test, participants were less able to identify the corresponding face out of a line-up of 12 options. Seemingly consistent with PDT, however, participants' performance was less affected by stimulus inversion when the actions described were negative (i.e., crimes or norm violations).

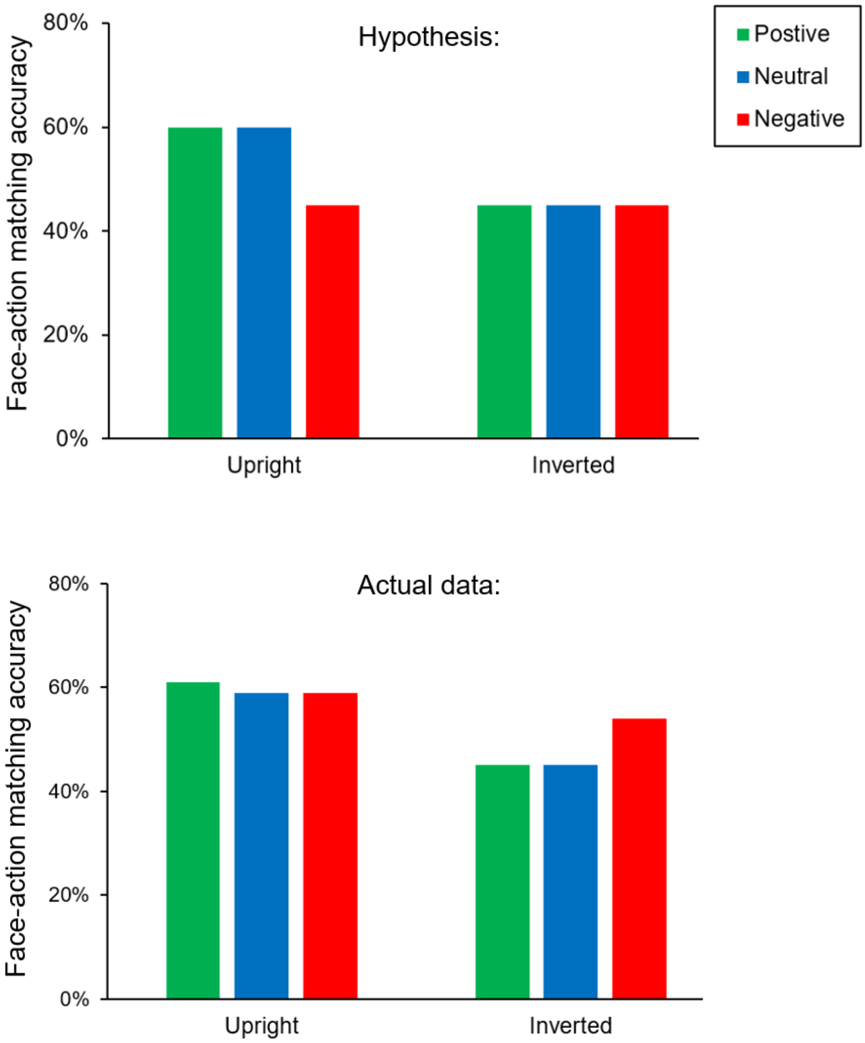


**Figure 4.** Predicted and observed results from the first study described by Fincher & Tetlock (2016). Top panel: PDT predicts a difference in face-action learning between the faces of norm violators and norm followers in in the upright condition because the faces of norm violators are hypothesised not to recruit configural processing. Bottom panel: Fincher and Tetlock (2016) observe an unexpected difference in face-action learning between the faces of norm violators and norm followers in the inverted condition where configural processing is thought to make little or no contribution to processing.

According to the authors, this demonstrates that when we learn that an individual has been guilty of a norm violation their face is processed in a feature-based (non-configural) manner which is less sensitive to inversion (see Section 3.1). The specific pattern of results does not support this conclusion, however. If Fincher and Tetlock's hypothesis is correct, then one would expect to see differences between face-action learning for norm violators and norm followers in the upright conditions where configural processing should augment face-action learning for those who committed neutral and positive actions, but not for those who committed crimes (Figure 4, top). This is not what the authors observed. Instead, the authors describe no differences in face-action learning in the upright conditions but find superior face-action learning for negative actions in the inverted conditions (Figure 4, bottom). This pattern of results is more consistent with the view that the detrimental effects of inversion may have been somewhat offset by the highly salient and graphic nature of the negative actions (including descriptions of rape and murder) – features that may have aided face-action learning in the more challenging inverted condition.

It has also been argued that the faces of criminals produce weaker composite face effects (see Section 3.2) than those of norm-following ingroup members, suggestive of weaker configural face processing (Fincher & Tetlock, 2016; Study 3). Fincher and Tetlock told participants about a series of violent crimes, each involving a perpetrator and a victim (e.g., Dylan shot Jeff during a robbery). Participants then performed sequential matching judgements about the faces of perpetrators and victims. Each sequential judgement began with a ‘target’ face shown for 400 ms. A name was shown underneath to indicate whether it was the victim or the perpetrator. Shortly after, a second ‘sample’ face was presented. The bottom half of the sample face was always different from the target face. The top half of the sample face matched the top half of the target face on 50% of trials. On the other 50% of trials, the top halves did not match. Participants’ task was to judge whether or not the two top halves were identical. 50% of the trials were ‘intact’ where the top and bottom halves were spatially aligned. 50% of the trials were ‘offset’ whereby the top and bottom halves were misaligned horizontally – a manipulation thought to abolish configural processing and composite interference.

Broadly typical composite face effects were seen when the target face was a victim: As expected, discrimination sensitivity was far superior in the offset condition than in the intact condition. However, when the face depicted was a perpetrator, no composite face effect was observed. Fincher and Tetlock (2016) claim that this finding accords with the view that the faces of violent criminals engage configural face processing to a lesser degree than the faces of norm-following ingroup members.

Closer inspection reveals that the results are actually inconsistent with the authors’ interpretation, however. To illustrate the problem, we have plotted the results of Fincher and Tetlock (2016; Study 3) in Figure 5. The authors’ hypothesis is that participants fail to process the faces of violent criminals configurally. If this were the case, then one would expect to observe differences between the victim and perpetrator conditions when composite arrangements are intact because only the faces of victims should recruit configural processing. When the arrangements are offset, on the other hand, there should be no difference between the victim and perpetrator conditions because neither condition would be expected to recruit configural processing (Murphy et al., 2017; Rossion, 2013). This is not what Fincher and Tetlock observe. Instead, they describe a substantial and surprising drop in discrimination sensitivity between the offset-victim and offset-perpetrator conditions. This is extremely difficult to interpret within the context of PDT because, as outlined above, configural processing is thought to make little or no contribution to the processing of misaligned composite arrangements (Murphy et al., 2017; Rossion, 2013).

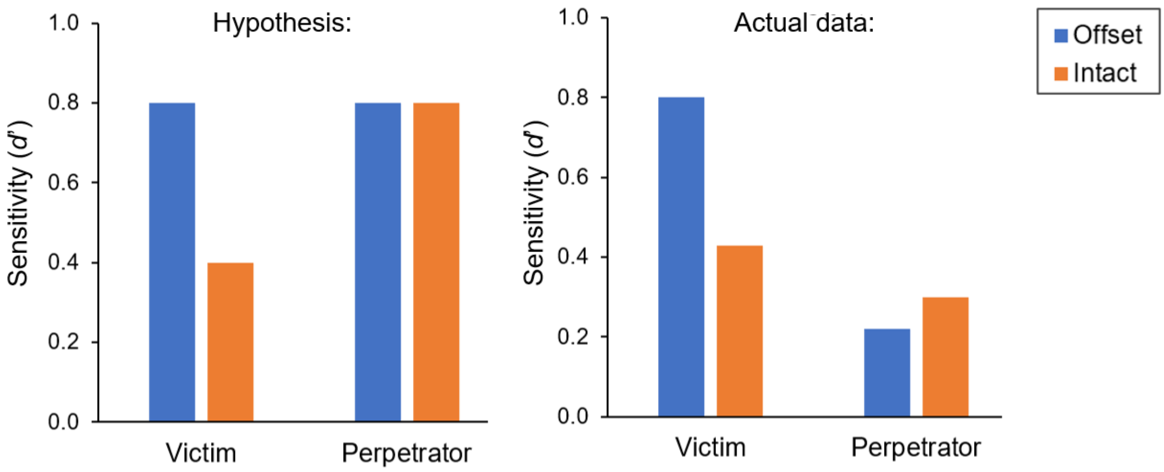


**Figure 5.** Predicted and observed results from the third study described by Fincher & Tetlock (2016). Left panel: PDT predicts that performance in the offset-victim and offset-perpetrator conditions will be similar because configural processing is not thought to be recruited in either case. Right panel: Fincher and Tetlock (2016) observed a surprising difference in performance between the offset-victim and offset-perpetrator conditions even though configural processing is thought to make little or no contribution to the processing of misaligned composite arrangements.

From a methodological point of view, we note that it is highly unorthodox to present a name or other text underneath a composite face arrangement. This feature is likely to draw attention away from the target region of the stimulus arrangement. Given the brief stimulus presentation (400 ms), this extraneous information may well have affected the manifestation of the illusion and participants' perceptual decisions. It is also unclear how the authors were able to calculate *d′* sensitivity indices from only 32 trials, having employed a complex 2 × 2 × 2 factorial design – this would appear to leave only 4 trials per cell (see also Fincher & Tetlock, 2016; Study 5). Finally, the order of the victim and perpetrator were not counterbalanced in the description of the crime; the victim was always described second (e.g., Dylan shot Jeff during a robbery). It is possible these confounds and methodological choices contributed to the low levels of discrimination sensitivity seen in the offset and intact perpetrator conditions.

*4.3 Findings from ethnic minority faces*

Several studies have compared markers of face processing seen when participants view faces drawn from ethnic ingroups and ethnic outgroups (Golby et al., 2001; Hugenberg et al., 2016; Michel, Rossion, Han, Chung, & Caldara, 2006; Ratner & Amodio, 2013; Van Bavel et al., 2011). For example, White Americans showed a stronger FFA response when viewing White faces than when viewing Black faces (Golby et al., 2001). Similarly, White participants showed stronger composite face effects for arrangements constructed from White faces than arrangements constructed from East Asian faces (Michel et al., 2006). These effects do not always replicate reliably (e.g., Horry, Cheong, & Brewer, 2015; Mondloch et al., 2010; Natu & O'Toole, 2013; H. K. Wong, Estudillo, Stephen, & Keeble, 2021). Where observed, however, such findings are potentially consistent with a weaker form of PDT (Fincher et al., 2017; Hugenberg et al., 2016).

It is impossible to draw any strong conclusions from this line of evidence, however, because the manipulation of ethnic ingroup vs. ethnic outgroup is strongly confounded with visual experience. It is well established that humans often struggle to individuate types of faces with which they have little experience. The mechanisms of face perception appear to be calibrated to optimize identification of the faces that the visual system encounters most often (Furl, Phillips, & O'Toole, 2002; Valentine, 1991). For example, while White and East Asian participants often exhibit poor identification of East Asian and White faces, respectively, adults of Korean origin adopted by White families living in France showed better recognition of White faces than of Asian faces (Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005). Reduced responses in FFA when viewing ethnic minority faces (e.g., Golby et al., 2001) may thus reflect the fact that neural populations become tuned to features that best allow us to encode the types of face we encounter most frequently. Similarly, reduced composite face effects for ethnic minority faces, where observed (e.g., Michel et al., 2006), may reflect the fact that our visual system is slightly less able to predict the structure of one facial region from another when viewing types of face with which we are less familiar.

*4.4. The visual system treats outgroup faces as faces*

According to the strong form of PDT, the faces of certain outgroup individuals – including the members of ethnic minorities, stigmatized groups, and criminals – fail to engage domain-specific face processing typically recruited by ingroup faces. Instead, the faces of these individuals recruit generic piecemeal processing also engaged when viewing objects and non-human animals. Contrary to this view, however, there is considerable evidence that the visual system processes outgroup faces in a qualitatively similar way to ingroup faces. For example, participants show clear FFA responses when viewing faces from minimal ingroups and minimal outgroups (Van Bavel et al., 2011). Similarly, faces from minimal ingroups and minimal outgroups also produce clear N170 responses (Ratner & Amodio, 2013).

Similarly, the faces of members of ethnic minorities elicit numerous effects considered by many to be markers of face-specific processing when viewed by members of the ethnic majority. For example, behavioural studies indicate that the faces of members of ethnic minorities produce inversion effects (Megreya, White, & Burton, 2011; Rhodes, Hayward, & Winkler, 2006), composite face effects (Horry et al., 2015; Michel et al., 2006; Mondloch et al., 2010; H. K. Wong et al., 2021; Zhao, Hayward, & Bülthoff, 2014) and Thatcher effects (Hahn, Jantzen, & Symons, 2012) when viewed by members of ethnic majorities. Neuroscientific investigations suggest a similar conclusion. For example, the faces of members of ethnic minorities elicit the N170 ERP component (Caldara, Rossion, Bovet, & Hauert, 2004; Caldara et al., 2003) and strong activation of the FFA (Golby et al., 2001; Natu & O'Toole, 2013; Natu, Raboy, & O'Toole, 2011) when viewed by members of ethnic majorities.

Not only is there little evidence of dissociation between the face-processing mechanisms engaged by ethnic-majority and ethnic-minority faces, there is positive evidence for association. For example, those who exhibit excellent recognition of ethnic-majority faces (so-called “super-recognisers) also exhibit better than average recognition of ethnicity-minority faces (Bate et al., 2019). Similarly, those who exhibit poor recognition of ethnicity-majority faces (developmental prosopagnosics) also exhibit worse than average recognition of ethnicity-minority faces (Cenac, Biotti, Gray, & Cook, 2019).

### *4.5. Interim summary*

PDT has claimed support from studies that manipulate the group membership of the people depicted in stimulus images and examine the influence of this manipulation on measures of face processing. Studies have been run with groups created in the lab as well as real-world groups. We have argued that what appears to be a body of convergent evidence in favour of the theory is, in fact, considerably weaker than it first appears. Methodological problems make some of the findings in this area difficult to interpret. In several cases, the interpretations offered by proponents of PDT are incompatible with the empirical results they seek to explain.

## 5. Does the modulation of configural processing affect attributions of humanity?

According to PDT, the engagement of domain-specific mechanisms for the visual perception of faces – in particular, configural processing – conveys a signal of humanity or personhood. If this view is correct, manipulations that modulate the engagement of those mechanisms (e.g., upright vs. inverted presentation) ought to affect the attribution of humanity to the people depicted in stimulus images. Proponents of PDT have described several results that appear consistent with this prediction.

*5.1 Effects of orientation inversion on trait attributions and punishment*

One of the most prominent lines of evidence for PDT comes from experiments that examine how orientation manipulations affect the attribution of uniquely human traits and participants' willingness to punish the people depicted (Hugenberg et al., 2016; Wilson et al., 2018). For example, Hugenberg et al. (2016; Experiment 3) found that inverted faces were judged to possess positive human traits such as thoughtfulness, empathy, consideration, and creativity to a lesser extent than were upright faces. Similarly, inverted faces were judged less trustworthy (Wilson et al., 2018). In related work, Fincher and Tetlock (2016; Study 6) found that participants endorsed harsher sentences for supposed criminals when their faces were shown upside-down. Inversion is thought to disrupt configural face processing (McKone & Yovel, 2009; Rossion, 2008). Thus, these findings potentially accord with the view that those faces that fail to fully engage typical face processing are attributed fewer human qualities, and are consequently at greater risk of harm.

In fact, these findings are not particularly surprising when considered in the context of the wider face perception literature. Orientation inversion is thought to disrupt the ability of observers to form structural descriptions of faces (McKone & Yovel, 2009; Rossion, 2008). Because observers are less able to encode stimulus variation, face shapes appear less distinctive (i.e., closer to the population average) and expressions appear more neutral. Consistent with this view, inversion adversely affects a wide range of perceptual decisions made about faces that are unrelated to theoretical models of dehumanization including identity recognition (Yin, 1969), unfamiliar face matching (Duchaine & Nakayama, 2006a), judgements of facial similarity (Biotti, Gray, & Cook, 2019; Duchaine et al., 2007), decisions about facial sex (Murphy et al., 2020), judgements of facial age (Murphy & Cook, 2017), the categorisation of facial emotion (McKelvie, 1995; Prkachin, 2003), ratings of attractiveness (Bäuml, 1994; Cook & Duchaine, 2011), and judgements about facial adiposity (Thompson & Wilson, 2012).

When we spontaneously infer the character traits of strangers from facial appearance, our attributions can be based on subtle differences in face shape and configuration – for example, whether a stranger's eyes are close together – and/or the perception of expression cues – for example, whether someone is, or appears to be, smiling, scowling or sorrowful (Cook, Eggleston, & Over, 2022). Given that these kinds of cue are harder to detect and encode when faces are shown upside-down, it is unsurprising that certain individuals who appear thoughtful, empathetic, considerate, and creative when their faces are viewed upright, exhibit fewer of these "human" traits when their faces are viewed upside-down (Hugenberg et al., 2016; Wilson et al., 2018). Similarly, people who appear friendly, trustworthy, and remorseful when viewed upright, are likely to appear less friendly, less trustworthy, and less remorseful when shown upside-down. As a result, participants may endorse less severe punishments when these faces are shown upright, than when inverted.

In light of these competing explanations for the existing data, it is important that future research addresses two key questions. The first relates to the types of trait judgement affected by orientation inversion. If the effects of inversion on the attribution of traits reflects perceptual dehumanization, then the traits affected should be limited to, or especially strong for, those that distinguish humans from animals (so-called animalistic dehumanization) and machines (so-called mechanistic dehumanization) (Haslam, 2006). Examples of such uniquely human traits include creativity, thoughtfulness, and consideration. 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 unrelated to human uniqueness (i.e., shared with other animals and machines) such as aggression, strength, and endurance. To date, there has been no systematic effort to distinguish these possibilities.

The second question relates to inter-stimulus differences. According to the perceptual dehumanization view, orientation should exert a similar effect on all faces. For example, all faces should appear less trustworthy, creative, and considerate when shown upside-down because configural processing is always impeded by inversion. In contrast, our alternative explanation predicts that the effect of face inversion on social evaluations will vary depending on the particular faces used. Under our rival interpretation, inverted faces appear less distinctive than upright faces 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. Sometimes this 'regression to the mean' may lead to more favourable social evaluations. For example, faces deemed extremely untrustworthy when viewed upright, may be found to appear more trustworthy when viewed upside-down (Figure 6).

**Figure 6**. Whereas individuals with genuine, Duchenne smiles may appear more trustworthy when presented upright than inverted, individuals with fake smiles may appear more trustworthy when presented inverted than upright.[3]

In line with this view, previous findings confirm that some faces are judged more favorably when inverted. For example, faces that have had their feature configurations distorted are judged to be more "grotesque" when shown upright, than when inverted (Donnelly et al., 2011; Hahn et al., 2012; Searcy & Bartlett, 1996). Presumably, observers are less able to detect the unsettling alterations when these faces are shown upside-down. Similarly, faces are often judged more attractive when shown upside-down, than when shown upright (Cook & Duchaine, 2011; Leder, Goller, Forster, Schlageter, & Paul, 2017). Again, it is possible that facial variation deemed unattractive by observers goes undetected when faces are inverted.

*5.2 Effects of orientation inversion on explicit attributions of humanity*

Arguably more direct evidence for PDT comes from studies which measure the effects of inversion on explicit attributions of humanness. For example, participants appear to judge inverted faces as "less human-like" than upright faces (Cassidy et al., 2021; Hugenberg et al., 2016). While this superficially appears to be strong evidence for PDT, again a crucial control condition has not been appropriately incorporated. At present, it is impossible to discern whether inversion disproportionately interferes with attributions of humanness or whether all objects appear more like themselves when presented in their canonical orientation. For example, cats and sleeping bats may appear more 'cat-like' and 'bat-like' when encountered in their expected orientation (Figure 7).

**Figure 7.** According to PDT, human faces appear more 'human-like' when shown upright than when shown inverted. We speculate that other entities, like cats and sleeping bats, will also be judged more 'cat-like' and 'bat-like' when shown in their canonical orientation.[4]

*5.3 Effects of orientation inversion on implicit measures of humanity*

A complementary line of research has employed more implicit measures to argue that participants are slower to associate inverted faces with the concept "human". In one study described by Hugenberg et al. (2016; Experiment 1), participants were asked to categorise words (e.g., person, individual, computer, robot) as 'related to humans' or 'related to machines' in a speeded task. People were faster to categorise words as 'related to human' when they were preceded by an upright face than when they were preceded by an inverted face. In a second study described by Hugenberg et al. (2016; Experiment 2), participants were faster to categorise stimulus images as 'human' or 'chimpanzee' when faces were shown upright. In a further study described by Young et al. (2019), participants had to indicate whether an upright or inverted image depicted a human or a non-human – either a robot (Experiment 1) or a Chimpanzee (Experiment 2) – by moving a mouse cursor from a starting position at the bottom-centre of the display, to either the right (to indicate 'human') or left (to indicate 'robot' or 'chimpanzee'). When human target faces were inverted, the cursor trajectory was less linear and appeared to show greater 'attraction' to the non-human label.

Again, this line of evidence is unconvincing. It is well established that the visual processing of faces is delayed by stimulus inversion. For example, the N170 ERP component is delayed when stimulus faces are shown upside-down (Rossion et al., 2000). It is therefore unsurprising that decisions about faces (e.g., whether they are human or not) are a little slower when stimulus images are shown upside-down (Hugenberg et al., 2016; Experiment 2). Moreover, it seems likely that this different processing time-course may impact priming effects inferred from RT measures. For example, priming effects might well be induced by inverted faces but slightly delayed relative to those induced by upright faces (Hugenberg et al., 2016; Experiment 1). It also seems possible that faster responses (i.e., those to upright faces) are less prone to response competition as there is less opportunity for interference induced by a rival category to accumulate (S. G. Young et al., 2019). On a methodological note, it is striking that the response mappings (e.g., human response on the right, non-human response on the left) were not counterbalanced in these studies (Hugenberg et al., 2016; S. G. Young et al., 2019). It is unclear how this impacted responding on these implicit measures.

*5.4 Effects of image filtering on attributions of humanity*

Another line of evidence potentially consistent with PDT comes from studies employing image filtering techniques rather than inversion (Fincher & Tetlock, 2016). Fincher and Tetlock (2016; Study 2) examined participants' ability to learn pairings between individual faces and various positive actions, neutral actions, and negative actions similar to those described above. Upright faces were presented either filtered (their high-spatial frequency content was removed) or full-spectrum, both during training and at test. Participants found it particularly hard to learn pairings between the filtered faces and negative actions. In a follow-up experiment, Fincher and Tetlock (2016; Study 7) found that participants also endorsed more lenient sentences for individuals depicted in filtered images than in unaltered full-spectrum images.

Upright facial images that have had their high-spatial frequency content removed are thought to afford more configural processing than unaltered 'full-spectrum' faces (Goffaux & Rossion, 2006). Thus, Fincher and Tetlock (2016) argue that observers attribute even more human qualities, and experience even more empathy for, the filtered images than they do for unaltered images because they engage more configural processing. However, a plausible alternative explanation for these findings is that the spatial filtering gave the faces a smooth, blemish-free appearance and thereby made the individuals look more attractive (Jaeger, Wagemans, Evans, & van Beest, 2018). It is well-established that attractive faces secure positive social evaluation - the "what is beautiful is good" stereotype (Dion, Berscheid, & Walster, 1972). This may explain why participants found it harder to learn negative face-action associations (Fincher & Tetlock, 2016; Study 2), and endorsed more lenient sentencing (Fincher & Tetlock, 2016; Study 7), when faces had their high-spatial frequency content removed.

*5.5. Interim summary*

PDT maintains that when faces are processed configurally, they are attributed greater humanity. We have argued that the evidence in favour of this claim is considerably weaker than it first appears. Studies designed to directly test PDT often lack crucial control conditions. In many cases, plausible alternative interpretations present themselves. For example, inversion may impair the efficiency and accuracy of many judgments about faces not only those associated with humanness. Similarly, spatial-frequency filtering may influence behavioural responses by enhancing differences in perceived attractiveness, not by modifying perceptions of humanness *per se*.

**6. Comments on the broader logic of PDT**

In this penultimate section, we consider the broader logic underlying PDT. Careful consideration reveals several further reasons to be sceptical of this account.

*6.1 Do people who process faces atypically recognise the humanity of others?*

Proponents of PDT suggest that members of ethnic minorities and other stigmatized outgroups are dehumanized, in part, because their faces fail to recruit configural processing when viewed by members of the majority group (Fincher et al., 2017; Hugenberg et al., 2016). Similarly, it is argued that we attribute less humanity to inverted faces because they do not engage typical face processing mechanisms (Fincher et al., 2017; Hugenberg et al., 2016). The apparent implication is that typical face processing plays a necessary, causal role in the attribution of humanity to others – at least when their face is visible (see Section 2.2).

Consistent with this suggestion, Hugenberg and colleagues (2016) assert that people with atypical face processing may be incapable of fully appreciating the humanity of others. This possibility was raised with particular reference to Autistic individuals:

> *"First, autism is associated with both impairments in theory of mind (…) and abnormal face processing, including failures to configurally process faces (…). The impairments linked to autism share features of (mechanistic) dehumanization (…), suggesting that theory of mind deficits central to autism may be related to face processing. From our perspective, it may not be happenstance that those with chronic inabilities to process the complex mental states of others also have chronic inabilities to process faces in a manner distinct from objects"* (Hugenberg et al., 2016, p173).

In our view, this is a misleading and offensive characterisation of Autistic individuals. There is no evidence that Autistic individuals fail to recognise the humanity of others. While some Autistic individuals experience problems interpreting the emotions and mental states of others, they clearly recognise that others *have* emotions and mental states (Bird & Cook, 2013). Recent findings suggest that, once emotion recognition and interoceptive problems are accounted for, Autistic individuals show typical levels of empathy (Bird et al., 2010;

Santiesteban et al., 2021). Moreover, there is no credible evidence that Autistic individuals are more likely to discriminate against or harm others (Heeramun et al., 2017), as would be predicted by PDT.

It is beyond doubt that some Autistic individuals exhibit poor face recognition (e.g., Boucher, Lewis, & Collis, 1998; Gehdu, Gray, & Cook, 2022; Hedley, Brewer, & Young, 2011; Ipser, Ring, Murphy, Gaigg, & Cook, 2016; Stantić, Ichijo, Catmur, & Bird, 2021) and neural markers of face processing sometimes behave differently in Autistic and non-Autistic participants (Kang et al., 2018; Nomi & Uddin, 2015). However, few within the vision science community would endorse the view that Autistic individuals *"have chronic inabilities to process faces in a manner distinct from objects"* (Hugenberg et al., 2016, p173). Autistic individuals exhibit many markers of typical face processing (for a systematic review, see Weigelt, Koldewyn, & Kanwisher, 2012), including the face inversion effect (Teunisse & De Gelder, 2003), composite face effect (Brewer, Bird, Gray, & Cook, 2019), strong responses to face stimuli in FFA (Bird et al., 2006; Nomi & Uddin, 2015), and the N170 marker (Kang et al., 2018).

To our knowledge, proponents of PDT are yet to suggest that individuals with developmental prosopagnosia fail to fully attribute humanity to others. Like Autism, developmental prosopagnosia is a neurodevelopmental condition associated with face recognition difficulties (Cook & Biotti, 2016; Duchaine & Nakayama, 2006b; Susilo & Duchaine, 2013). There is considerable evidence that people with developmental prosopagnosia process faces atypically (Avidan & Behrmann, 2009; Avidan et al., 2014; Duchaine & Nakayama, 2006b; Towler, Fisher, & Eimer, 2017, 2018), and sometimes show atypical configural face processing (Avidan, Tanzer, & Behrmann, 2011; Liu & Behrmann, 2014; Palermo et al., 2011). The clear implication is that – like Autistic individuals – they would also be expected to dehumanize others; i.e., fail to ascribe full humanness and personhood to observed individuals. However, people with developmental prosopagnosia frequently perform typically on measures of social cognition (Duchaine, Murray, Turner, White, & Garrido, 2009) and show typical attribution of character traits to others (Todorov & Duchaine, 2008).

### *6.2 Why would we disengage social cognition when we encounter faces in non-canonical orientations?*

PDT maintains that when faces are viewed upside-down, they fail to engage domain-specific face processing, or engage less domain-specific face processing than do upright faces. As a result, individuals are attributed less humanity and are more vulnerable to harm when encountered in non-canonical orientations (Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al., 2016).

Although faces are typically observed in a canonical orientation, there are many situations in which we view faces in other orientations. For example, when a father watches his daughter do a handstand, when a doctor observes her patient from the head of the bed, when an athlete observes their team-mate lunge for a ball, and when a mother breastfeeds her baby. It seems distinctly unlikely that in these cases, and others like them, the observers' unusual viewpoint causes a reduction in mental state attribution, a deficit in empathy, or a willingness to harm the observed. Indeed, to disengage social cognition in such situations would be distinctly maladaptive: understanding faces and bodies in action, across a range of orientations, is crucial for successful social interaction and coordination. In short, it seems essential that the attribution of humanity and the operation of social cognition is independent of – not tethered to – orientation-specific perceptual processing.

*6.3 Why would we disengage social cognition when we encounter outgroup members?*
In common with some other theoretical perspectives on dehumanization from social psychology (e.g., Harris & Fiske, 2006), PDT asserts that when individuals are dehumanized, social cognition is not engaged or is engaged to a lesser extent (Deska & Hugenberg, 2017; Fincher & Tetlock, 2016; Fincher et al., 2017; Hugenberg et al., 2016). In different contributions to the literature on PDT, dehumanization has been characterised as a failure to fully attribute human mental states (e.g., Deska & Hugenberg, 2017), human character traits (e.g., Hugenberg et al., 2016; Wilson et al., 2018), human emotions (Cassidy et al., 2021), and human essence (e.g., Hugenberg et al., 2016; S. G. Young et al., 2019) to the outgroup.

This work stands in tension with a parallel body of work in developmental and evolutionary psychology. According to the Machiavellian Intelligence Hypothesis, mentalising abilities emerged partly as a means by which to maximise the chance of success in competitive interactions (Whiten & Byrne, 1997). For example, mentalising abilities help individuals infer whether or not they are being deceived by a competitor (Sperber et al., 2010). On this account, failing to attribute mental states to certain outgroup members (e.g., rivals and aggressors) could be seriously deleterious to survival and reproduction.

Indeed, there is a substantial body of empirical research suggesting that we do attribute mental states, traits and emotions to outgroup members. Whereas outgroup members might be thought of as lacking socially desirable human qualities such as warmth and rationality, they are often viewed as possessing undesirable human qualities, such as arrogance, spite and jealousy to a greater extent than do the ingroup (Enock, Flavell, Tipper, & Over, 2021; Enock, Tipper, & Over, 2021; Over, 2021a, 2021b). These findings accord with research on stereotype content showing that outgroups are often thought of as possessing negative yet uniquely human characteristics, such as being devious, deceptive and cunning (Over, 2021b).

**7. Conclusion**

PDT argues that the faces of outgroup members fail to fully engage domain-specific visual processing and instead recruit the same generic piecemeal processing used to individuate objects and non-human animals. As a result of this ‘perceptual dehumanization’, outgroup members are not afforded full humanity and are thus rendered vulnerable to harm (reviewed in Section 2). Two versions of this theory can be delineated. According to the strong form of the theory, domain-specific face-processing is engaged only by ingroup faces. According to the weak form, the faces of outgroup members may also engage domain-specific mechanisms for face processing, but do so less than ingroup faces.

Crucially, both forms of the theory are predicated on the assumption that faces typically engage qualitatively different domain-specific processing not recruited by objects and non-human animals. In Section 3, we argued that this view is based on a selective, one-sided reading of the face perception literature. In particular, many behavioural and neuroscientific findings suggest considerable overlap between the visual processing of faces and non-social objects of expertise. In light of this literature, it is unclear whether it is possible to distinguish ‘humanizing’ from ‘dehumanizing’ processing pathways. Proponents of PDT have not engaged extensively with the debate regarding the domain-specificity of face processing mechanisms. Nevertheless, this literature poses some challenging questions for the theory. For example, are objects-of-expertise (non-social stimuli that recruit configural processing) attributed humanity and personhood?

In Sections 4 and 5, we consider whether the existing evidence base accords with PDT. We find that that the strong version of PDT has little or no empirical support. While the weaker form of the theory is harder to refute, we find that there is little compelling evidence for it either. Studies in this area often lack crucial control conditions, fail to incorporate appropriate counterbalancing and/or make highly unorthodox methodological choices. Time and again, plausible alternative interpretations of the data present themselves but are overlooked. For example, inversion may impair the efficiency/accuracy of many judgments about faces not only those associated with humanness. Similarly, spatial filtering may give faces a smooth blemish-free appearance, and thereby make them more attractive, leading to more positive treatment. In sum, the existing evidence base provides far less support for PDT than it superficially appears to.

Looking beyond methodological problems, PDT also faces serious conceptual challenges (reviewed in Section 6). PDT maintains that in order to recognise the full humanity of an individual or group, the typical mechanisms of face processing must be engaged fully. One implication of this view is that people who are unable to process faces typically should be unable to fully recognise the humanity of others. However, this suggestion does not accord with previous findings from studies of developmental prosopagnosia. A second implication is that social cognition is not fully engaged in situations where it would appear to be essential; for example, in competitive interactions with out-group members.

We do not deny that perceiving a human face likely engages neurocognitive mechanisms for social cognition. Crucially, however, we suggest that it is not necessary to encounter a face in a particular orientation or process it in a particular way (i.e., configurally), in order to attribute humanity to it. Faces are clearly a rich source of information about the emotions and mental states of others. However, visual face processing is just one of many ways we can infer the presence of a conspecific. For example, we are able to infer the presence of another when their face is occluded (e.g., when we see a pair of feet protruding from behind a curtain or when we are approached by someone wearing a surgical mask) or from auditory cues (e.g., if we hear footsteps on the path behind us or hear someone crying). In each case, we speculate that social cognition is engaged and humanity is attributed to the other person, thereby enabling us to reason about their intentions, emotions, traits, and perspective.

In the wider social psychological literature, it has long been argued that dehumanizing processes play an important role in understanding discrimination and intergroup harm (Haslam, 2006; Haslam & Loughnan, 2014; Leyens et al., 2000; Leyens et al., 2001). Increasingly, however, the utility of social psychological models of dehumanization has been called into question, following conceptual (Bloom, 2017; Lang, 2010; Manne, 2016; Over, 2021a, 2021b; Smith, 2011) and empirical (Enock, Flavell, et al., 2021; Enock, Tipper, et al., 2021) challenges. To date, the literature on PDT has largely escaped critical scrutiny. However, our analysis suggests that, in its current formulation, PDT lacks empirical support and leaves a number of important questions unanswered.

**Footnotes**

[1]The images used to create panel (2a) were cropped from Lee et al. (2021; Figure 1) published under CC BY-SO 4.0 (https://creativecommons.org/licenses/by-sa/4.0/). The images used to create panel (2b) were cropped from Sunday, Donnelly and Gauthier (2017; Figure 2) published under CC BY-SO 4.0 (https://creativecommons.org/licenses/by-sa/4.0/).

[2]The images used to create panel (3a) were cropped from Gray, Murphy, March & Cook (2017; Figure 2) published under CC BY-SO 4.0 (https://creativecommons.org/licenses/by-sa/4.0/). The images used to create panel (3b) were cropped from Utz and Carbon (2016; Figure 1) published under CC BY-SO 4.0 (https://creativecommons.org/licenses/by-sa/4.0/).

[3]The images used to create Figure 6 were cropped from Manera, Del Giudice, Grandi and Colle (2011; Figure 1) published under CC BY-SO 4.0 (https://creativecommons.org/licenses/by-sa/4.0/).

[4]The image used to create the left most panel of Figure 7 was provided by the author. The image used to create the central panel was published under a CC-BY-SA 3.0 license https://commons.wikimedia.org/wiki/File:Cat_November_2010-1a.jpg. The image used to create the right most panel was published under a CC-BY-SA 4 license https://commons.wikimedia.org/wiki/File:Greater_Indian_fruit_bat_@_Kanjirappally_01.jpg

## References


Abbas, Z. A., & Duchaine, B. (2008). The role of holistic processing in judgments of facial attractiveness. *Perception, 37*(8), 1187-1196.

Anstis, S. (2005). Last but not least. *Perception, 34*(2), 237–240.

Ashworth, A., Vuong, Q., Rossion, B., & Tarr, M. (2008). Recognizing rotated faces and Greebles: What properties drive the face inversion effect? *Visual Cognition, 16*(6), 754–784.

Avidan, G., & Behrmann, M. (2009). Functional MRI reveals compromised neural integrity of the face processing network in congenital prosopagnosia. *Current Biology, 19*(13), 1146-1150.

Avidan, G., Tanzer, M., & Behrmann, M. (2011). Impaired holistic processing in congenital prosopagnosia. *Neuropsychologia, 49*(9), 2541-2552.

Avidan, G., Tanzer, M., Hadj-Bouziane, F., Liu, N., Ungerleider, L. G., & Behrmann, M. (2014). Selective dissociation between core and extended regions of the face processing network in congenital prosopagnosia. *Cerebral Cortex, 24*, 1565–1578.

Barton, J. J. (2008). Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. *Journal of Neuropsychology, 2*(1), 197-225.

Barton, J. J., Albonico, A., Susilo, T., Duchaine, B., & Corrow, S. L. (2019). Object recognition in acquired and developmental prosopagnosia. *Cognitive Neuropsychology, 36*(1-2), 54-84.

Bate, S., Bennetts, R., Hasshim, N., Portch, E., Murray, E., Burns, E., & Dudfield, G. (2019). The limits of super recognition: An other-ethnicity effect in individuals with extraordinary face recognition skills. *Journal of Experimental Psychology: Human Perception and Performance, 45*(3), 363-377.

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

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8*(6), 551-565.

Bernard, P., Gervais, S. J., Allen, J., Campomizzi, S., & Klein, O. (2012). Integrating sexual objectification with object versus person recognition: The sexualized-body-inversion hypothesis. *Psychological Science, 23*(5), 469-471.

Bilalić, M., Grottenthaler, T., Nägele, T., & Lindig, T. (2016). The faces in radiological images: fusiform face area supports radiological expertise. *Cerebral Cortex, 26*(3), 1004-1014.

Bilalić, M., Langner, R., Ulrich, R., & Grodd, W. (2011). Many faces of expertise: fusiform face area in chess experts and novices. *Journal of Neuroscience, 31*(28), 10206-10214.

Biotti, F., Gray, K. L. H., & Cook, R. (2017). Impaired body perception in developmental prosopagnosia. *Cortex, 93*, 41-49.

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

Bird, G., Catmur, C., Silani, G., Frith, C., & Frith, U. (2006). Attention does not modulate neural responses to social stimuli in autism spectrum disorders. *NeuroImage, 31*(4), 1614-1624.

Bird, G., & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry, 3*(7), e285.

Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain, 133*(5), 1515-1525.

Bliuc, A. M., McGarty, C., Thomas, E. F., Lala, G., Berndsen, M., & Misajon, R. (2015). Public division about climate change rooted in conflicting sociopolitical identities. *Nature Climate Change, 5*, 226-229.

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

Boucher, J., Lewis, V., & Collis, G. (1998). Familiar face and voice matching and recognition in children with autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*(2), 171-181.

Brewer, R., Bird, G., Gray, K. L., & Cook, R. (2019). Face perception in autism spectrum disorder: Modulation of holistic processing by facial emotion. *Cognition, 193*, e104016.

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.

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research, 45*(4), 431–448.

Caldara, R., Rossion, B., Bovet, P., & Hauert, C. A. (2004). Event-related potentials and time course of the 'other-race'face classification advantage. *Neuroreport, 15*(5), 905-910.

Caldara, R., Thut, G., Servoir, P., Michel, C. M., Bovet, P., & Renault, B. (2003). Face versus non-face object perception and the 'other-race'effect: a spatio-temporal event-related potential study. *Clinical Neurophysiology, 114*(3), 515-528.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance, 26*(2), 527-551.

Campbell, A., & Tanaka, J. W. (2018). Inversion impairs expert budgerigar identity recognition: A face-like effect for a nonface object of expertise. *Perception, 47*, 647-659.

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. (2021). Decoding complex emotions and humanization show related face processing effects. *Emotion*.

Cenac, Z., Biotti, F., Gray, K. L. H., & Cook, R. (2019). Does developmental prosopagnosia impair identification of other-ethnicity faces? *Cortex, 119*, 12–19.

Chin, M. D., Evans, K. K., Wolfe, J. M., Bowen, J., & Tanaka, J. W. (2018). Inversion effects in the expert classification of mammograms and faces. *Cognitive Research: Principles and Implications, 3*.

Civile, C., Colvin, E., Siddiqui, H., & Sukhvinder, O. (2018). Labelling faces as 'Autistic' reduces the inversion effect. *Autism, 23*, 1596-1600.

Civile, C., Zhao, D., Ku, Y., Elchlepp, H., Lavric, A., & McLaren, I. P. L. (2014). Perceptual learning and inversion effect: Recognition of prototype-defined familiar checkerboards. *Journal of Experimental Psychology: Animal Learning and Cognition, 40*, 144–161.

Cook, R., & Biotti, F. (2016). Developmental prosopagnosia. *Current Biology, 26*(8), R312-R313.

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*.

Derntl, B., Seidel, E. M., Kainz, E., & Carbon, C. C. (2009). Recognition of emotional expressions is affected by inversion and presentation time. *Perception, 38*(12), 1849-1862.

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.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*, 107-117.

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

Donnelly, N., Zürcher, N. R., Cornes, K., Snyder, J., Naik, P., Hadwin, J., & Hadjikhani, N. (2011). Discriminating grotesque from typical faces: evidence from the Thatcher illusion. *PloS One, 6*(8), e23340.

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology, 24*(4), 419-430.

Duchaine, B., Murray, H., Turner, M., White, S., & Garrido, L. (2009). Normal social cognition in developmental prosopagnosia. *Cognitive Neuropsychology, 26*(7), 620-634.

Duchaine, B., & Nakayama, K. (2006a). The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia, 44*, 576-585.

Duchaine, B., & Nakayama, K. (2006b). Developmental prosopagnosia: a window to content-specific face processing. *Current Opinion in Neurobiology, 16*, 166-173.

Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual Review of Vision Science, 1*, 393-416.

Eimer, M. (1998). Does the face-specific N170 component reflect the activity of a specialized eye processor? *Neuroreport, 9*(13), 2945-2948.

Eimer, M. (2000). The face-specific N170 component reflects late stages in the structural encoding of faces. *Neuroreport, 11*(10), 2319-2324.

Eimer, M., Kiss, M., & Nicholas, S. (2010). Response profile of the face-sensitive N170 component: a rapid adaptation study. *Cerebral Cortex, 20*(10), 2442-2452.

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.

Enock, F. E., Tipper, S. P., & Over, H. (2021). Intergroup preference, not dehumanization, explain social biases in emotion attribution. *Cognition, 216*, 104865.

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.

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.

Ganel, T., Valyear, K. F., Goshen-Gottstein, Y., & Goodale, M. A. (2005). The involvement of the “fusiform face area” in processing facial expression. *Neuropsychologia, 43*(11), 1645-1654.

Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition, 103*(2), 322-330.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*(2), 191-197.

Gauthier, I., & Tarr, M. J. (1997). Becoming a “Greeble” expert: exploring mechanisms for face recognition. *Vision Research, 37*, 1673–1681.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform'face area'increases with expertise in recognizing novel objects. *Nature Neuroscience, 2*(6), 568-573.

Gehdu, B. K., Gray, K. L. H., & Cook, R. (2022). Imapired grouping of ambient facial images in autism *Scientific Reports*.

Geskin, J., & Behrmann, M. (2018). Congenital prosopagnosia without object agnosia? A literature review. *Cognitive Neuropsychology, 35*(1-2), 4-54.

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.

Golby, A. J., Gabrieli, J. D., Chiao, J. Y., & Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience, 4*(8), 845-850.

Gray, K. L. H., Biotti, F., & Cook, R. (2019). Evaluating object recognition ability in developmental prosopagnosia using the Cambridge Car Memory Test. *Cognitive Neuropsychology, 36*(1-2), 89-96.

Gray, K. L. H., Guillemin, Y., Cenac, Z., Gibbons, S., Vestner, T., & Cook, R. (2020). Are the facial gender and facial age variants of the composite face illusion products of a common mechanism? *Psychonomic Bulletin & Review, 27*(1), 62-69.

Gray, K. L. H., Murphy, J., Marsh, J. E., & Cook, R. (2017). Modulation of the composite face effect by unintended emotion cues. *Royal Society Open Science, 4*(4), 160867.

Hahn, A. C., Jantzen, K. J., & Symons, L. A. (2012). Thatcherization impacts the processing of own-race faces more so than other-race faces: An ERP study. *Social Neuroscience, 7*(1), 113-125.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science, 17*(10), 847–853.

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.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*, 223-233.

Hedley, D., Brewer, N., & Young, R. (2011). Face recognition performance of individuals with Asperger syndrome on the Cambridge Face Memory Test. *Autism Research, 4*(6), 449-455.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience, 3*(1), 80-84.

Hole, G. (1994). Configurational factors in the perception of unfamiliar faces. *Perception, 23*(1), 65-74.

Horry, R., Cheong, W., & Brewer, N. (2015). The other-race effect in perception and recognition: insights from the complete composite task. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 508-524.

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., 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.

Ipser, A., Ring, M., Murphy, J., Gaigg, S. B., & Cook, R. (2016). Similar exemplar pooling processes underlie the learning of facial identity and handwriting style: Evidence from typical observers and individuals with Autism. *Neuropsychologia, 85*, 169-176.

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.

Johnston, P., Baker, D. H., Stone, R., & Kaufman, J. (2014). Thatcher's Britain: A new take on an old illusion. *Perception, 43*(12), 1400-1403.

Kang, E., Keifer, C. M., Levy, E. J., Foss-Feig, J. H., McPartland, J. C., & Lerner, M. D. (2018). Atypicality of the N170 event-related potential in autism spectrum disorder: a meta-analysis. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(8), 657-666.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3*, 759–763.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361*(1476), 2109-2128.

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

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

Leder, H., Goller, J., Forster, M., Schlageter, L., & Paul, M. A. (2017). Face inversion increases attractiveness. *Acta Psychologica, 178*, 25-31.

Lee, R., Flavell, J. C., Tipper, S. P., Cook, R., & Over, H. (2021). Spontaneous first impressions emerge from brief training. *Scientific Reports, 11*(1), 1-13.

Lewis, M. B., & Johnston, R. A. (1997). The Thatcher illusion as a test of configural disruption. *Perception, 26*(2), 225-227.

Leyens, J. P., Paladino, P. M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The emotional side of prejudice: The attribution of secondary emotions to ingroups and outgroups. *Personality and Social Psychology Review, 4*, 186–197.

Leyens, J. P., Rodriguez-Perez, A., Rodriguez-Torres, R., Gaunt, R., Paladino, M. P., Vaes, J., & Demoulin, S. (2001). Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology, 31*, 395–411.

Liu, T. T., & Behrmann, M. (2014). Impaired holistic processing of left-right composite faces in congenital prosopagnosia. *Frontiers in Human Neuroscience, 8*, 750.

Manera, V., Del Giudice, M., Grandi, E., & Colle, L. (2011). Individual differences in the recognition of enjoyment smiles: No role for perceptual–attentional factors and autistic-like traits. *Frontiers in Psychology, 2*, e143.

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.

McLaren, I. P. L. (1997). Categorisation and perceptual learning: An analogue of the face inversion effect. *Quarterly Journal of Experimental Psychology, 50A*, 257–273.

Megreya, A. M., White, D., & Burton, A. M. (2011). The other-race effect does not rely on memory: Evidence from a matching task. *Quarterly Journal of Experimental Psychology, 64*(8), 1473-1483.

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.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience, 9*(5), 555-604.

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.

Murphy, J., Gray, K. L. H., & Cook, R. (2017). The composite face illusion. *Psychonomic Bulletin & Review, 24*(2), 245-261.

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

Natu, V., Raboy, D., & O'Toole, A. J. (2011). Neural correlates of own-and other-race face perception: Spatial and temporal response differences. *NeuroImage, 54*(3), 2547-2555.

Nomi, J. S., & Uddin, L. Q. (2015). Face processing in autism spectrum disorders: from brain regions to brain networks. *Neuropsychologia, 71*, 201-216.

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

Over, H. (2021b). Seven challenges for the dehumanisation hypothesis. *Perspectives on Psychological Science, 16*(1), 3–13.

Over, H., & McCall, C. (2018). Becoming us and them: Social learning and intergroup bias. *Social and Personality Psychology Compass, 12*(4), e12384.

Palermo, R., Willis, M. L., Rivolta, D., McKone, E., Wilson, C. E., & Calder, A. J. (2011). Impaired holistic coding of facial expression and facial identity in congenital prosopagnosia. *Neuropsychologia, 49*(5), 1226-1235.

Pascoe, E. A., & Smart Richman, L. (2009). Perceived discrimination and health: a meta-analytic review. *Psychological Bulletin, 135*, 531-554.

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.

Ratner, K. G., & Amodio, D. M. (2013). Seeing “us vs. them”: Minimal group effects on the neural encoding of faces. *Journal of Experimental Social Psychology, 49*(2), 298-301.

Rezlescu, C., Barton, J. J., Pitcher, D., & Duchaine, B. (2014). Normal acquisition of expertise with greebles in two cases of acquired prosopagnosia. *Proceedings of the National Academy of Sciences, 111*(14), 5123-5128.

Rhodes, G., Hayward, W. G., & Winkler, C. (2006). Expert face coding: Configural and component coding of own-race and other-race faces. *Psychonomic Bulletin & Review, 13*(3), 499-505.

Richeson, J. A. (2020). The mythology of racial progress. *The Atlantic*. https://www.theatlantic.com/magazine/archive/2020/09/the-mythology-of-racial-progress/614173/

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.

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition, 103*(1), 34-79.

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

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition, 21*(2), 139-253.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport, 11*(1), 69-72.

Rossion, B., Jacques, C., & Jonas, J. (2018). Mapping face categorization in the human ventral occipitotemporal cortex with direct neural intracranial recordings. *Annals of the New York Academy of Sciences, 1426*(1), 5-24.

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.

Santiesteban, I., Gibbard, C., Drucks, H., Clayton, N., Banissy, M. J., & Bird, G. (2021). Individuals with autism share others' emotions: evidence from the continuous affective rating and empathic responses (CARER) task. *Journal of Autism and Developmental Disorders, 51*(2), 391-404.

Schwaninger, A., Lobmaier, J. S., & Fischer, M. H. (2005). The inversion effect on gaze perception reflects processing of component information. *Experimental Brain Research, 167*(1), 49-55.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial–relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 904-915.

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

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.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 25*(4), 359-393.

Stantić, M., Ichijo, E., Catmur, C., & Bird, G. (2021). Face memory and face perception in autism. *Autism*, 13623613211027685.

Sunday, M. A., Donnelly, E., & Gauthier, I. (2017). Individual differences in perceptual abilities in medical imaging: the Vanderbilt Chest Radiograph Test. *Cognitive Research: Principles and Implications, 2*(1), 1-10.

Susilo, T., & Duchaine, B. (2013). Advances in developmental prosopagnosia research. *Current Opinion in Neurobiology, 23*, 423-429.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American, 223*(5), 96-103.

Tarr, M. J., & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience, 3*(8), 764-769.

Teunisse, J. P., & De Gelder, B. (2003). Face processing in adolescents with autistic disorder: The inversion and composite effects. *Brain and Cognition, 52*(3), 285-294.

Thierry, G., Martin, C. D., Downing, P., & Pegna, A. J. (2007). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nature Neuroscience, 10*(4), 505-511.

Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception, 9*(4), 483-484.

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

Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology, 25*(3), 395-410.

Towler, J., & Eimer, M. (2015). Early stages of perceptual face processing are confined to the contralateral hemisphere: Evidence from the N170 component. *Cortex, 64*, 89-101.

Towler, J., Fisher, K., & Eimer, M. (2017). The cognitive and neural basis of developmental prosopagnosia. *Quarterly Journal of Experimental Psychology, 70*(2), 316-344.

Towler, J., Fisher, K., & Eimer, M. (2018). Holistic face perception is impaired in developmental prosopagnosia. *Cortex, 108*, 112-126.

Utz, S., & Carbon, C. C. (2016). Is the Thatcher illusion modulated by face familiarity? Evidence from an eye tracking study. *PloS One, 11*(10), e0163933.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology, 43*(2), 161-204.

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The neural substrates of in-group bias: A functional magnetic resonance imaging investigation. *Psychological Science, 19*(11), 1131-1139.

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2011). Modulation of the fusiform face area following minimal exposure to motivationally relevant faces: evidence of in-group enhancement (not out-group disregard). *Journal of Cognitive Neuroscience, 23*(11), 3343-3354.

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.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience & Biobehavioral Reviews, 36*(3), 1060-1084.

Whiten, A., & Byrne, R. W. (1997). *Machiavellian intelligence II: Extensions and evaluations*: Cambridge University Press.

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., Hsiao, J., Greenspon, E., Ahern, E., Duan, Y., & Lui, K. F. (2012). Holistic processing as a hallmark of perceptual expertise for nonface categories including Chinese characters. *Journal of Vision, 12*(13), e7.

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.

Wong, Y. K., Twedt, E., Sheinberg, D., & Gauthier, I. (2010). Does Thompson's Thatcher effect reflect a face-specific mechanism? *Perception, 39*(8), 1125-1141.

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex, 15*(8), 1234-1242.

Yergeau, M. (2013). Clinically significant disturbance: On theorists who theorize theory of mind. *Disability Studies Quarterly, 33*(4), e3876.

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

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 42*(11), 1166-1178.

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.

Zhao, M., Hayward, W. G., & Bülthoff, I. (2014). Holistic processing, contact, and the other-race effect in face recognition. *Vision Research, 105*, 61-69.