Article:
OVER, HARRIET and Cook, Richard (2026) Understanding the role of cultural learning in the emergence of first impressions from facial appearance. European Review of Social Psychology. ISSN: 1479-277X

https://doi.org/10.1080/10463283.2025.2606574

Understanding the role of cultural learning in the emergence of first impressions from facial appearance

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

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

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

*Corresponding author:
harriet.over@york.ac.uk
Department of Psychology
University of York,
York, U.K., YO10 5DD

Funding statement: This research was supported by the European Research Council under the European Union's Horizon 2020 Programme, grant no. ERC-STG-755719.

Disclosure of interest: The authors have no competing interests to declare

Abstract

Humans spontaneously form character judgments about strangers from their facial appearance. While these ‘first impressions’ typically have little or no factual basis, they exert a significant influence over behaviour. Whereas some authors argue that first impressions have an innate origin, we propose that first impressions arise from learned associations between representations of facial appearance – conceived of as locations in multidimensional face space – and representations of the trait profiles that others may possess – characterised as locations in multidimensional trait space. Cultural messages, including those conveyed by propaganda, illustrations, art, iconography, films, and television, as well as interactions with caregivers and peers, teach children a range of face-trait mappings, some of which may be widely shared within their community. We review the emerging evidence base, much of which supports the TIM framework. However, we argue that previous research may have inadvertently ‘stacked the deck’ in favour of evolutionary accounts of first impressions by systematically confounding facial appearance cues (i.e., facial features, face shape) and facial behaviour cues (expressions, head tilt, gaze direction). To advance the origins debate, we argue that researchers must do more to distinguish between first impressions based on stable appearance cues and those based on actual or perceived facial behaviours.

Keywords

First impressions; Cultural learning, Trait inference mapping; Trustworthiness, Dominance

1. Introduction

When we encounter a stranger, we spontaneously attribute to them a wide variety of character traits based on their facial appearance. For example, we frequently make inferences about the extent to which they are trustworthy, dominant, or intelligent (Oosterhof & Todorov, 2008; Sutherland & Young, 2022; Todorov, 2017; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015; Zebrowitz, 2017). These 'first impressions' are rarely accurate and yet they exert a striking influence over our behaviour (Olivola, Funk, & Todorov, 2014; Todorov, 2017). Research has shown that first impressions of faces can influence hiring decisions, criminal sentencing and even the outcome of elections (Porter, ten Brinke, & Gustaw, 2010; Todorov, Mandisodza, Goren, & Hall, 2005; J. P. Wilson & Rule, 2015).

The origin of these first impressions from facial appearance remains controversial. According to nativist accounts, at least some first impressions are the product of innate face-trait mappings (Sutherland, Burton, et al., 2020; Van Vugt & Grabo, 2015; Zebrowitz, 2004, 2017; Zebrowitz & Zhang, 2011). According to the Trait Inference Mapping framework (TIM), on the other hand, first impressions are the products of learned associations between representations of facial appearance – conceived of as locations in multidimensional face space – and representations of the trait profiles that others may possess – characterised as locations in multidimensional trait space (Cook, Eggleston, & Over, 2022; Cook & Over, 2020; Over & Cook, 2018; Over, Eggleston, & Cook, 2020). TIM posits a key role for cultural messages, such as those common within storybooks, film, TV and political propaganda, in teaching children that certain facial characteristics are predictive of certain trait profiles. In focusing on the role of cultural learning in the acquisition of first impressions, TIM draws inspiration from both the cognitive and neuroscientific literature on face perception and from the social psychological literature on stereotyping and discrimination.

In this article, we review the literature on first impressions from facial appearance and evaluate the relative merits of the nativist and cultural learning accounts. In short, we believe that many findings – although not all – are compatible with the cultural learning account. Following this, we argue that previous research has systematically confounded stable appearance cues (e.g., interocular distance, nose

shape, mouth size) and transient behaviour cues (expressions, head tilt). We discuss the implications of this confounding for the origins debate and offer recommendations for future research.

Throughout the article, our focus is the attribution of traits (e.g., personality characteristics, intelligence) to others based on facial cues. We note, however, that the study of trait inferences from facial cues is part of a wider research endeavour – to understand the social evaluation of faces more broadly (Todorov, Olivola, et al., 2015). In this context, our tendency to spontaneously attribute traits to others is sometimes studied alongside other kinds of social evaluation, such as impressions of physical attractiveness, age, and emotional states (e.g., Sutherland et al., 2013). Consistent with our previous work (Cook et al., 2022; Over & Cook, 2018), we use the term "first impressions" to refer specifically to spontaneous trait attributions. This usage does not encompass impressions of physical attractiveness, age, or emotional state, which we regard as distinct phenomena.

2. First impressions from facial appearance

First impressions from facial appearance have been studied by scholars from various research traditions including cognitive, developmental and social psychology, vision science, and cognitive neuroscience. In a typical study, participants are presented with a series of faces, one at a time, and asked to make trait judgments about each in turn (e.g., Eggleston, Flavell, Tipper, Cook, & Over, 2021; Jones et al., 2021; Lavan, Mileva, Burton, Young, & McGettigan, 2021; Mileva, Young, Kramer, & Burton, 2019; Sutherland, Burton, et al., 2020; Sutherland et al., 2013; Todorov et al., 2005; Tsantani, Over, & Cook, 2023; Willis & Todorov, 2006). Stimulus presentation duration may be unlimited (e.g., Jones et al., 2021; Sutherland, Burton, et al., 2020) or restricted (e.g., Eggleston, Flavell, et al., 2021; Willis & Todorov, 2006). The results of studies employing limited presentation durations suggest that participants form reliable first impressions of faces presented very briefly – for 100 ms or less (Bar, Neta, & Linz, 2006; Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006).

The nature of the facial images presented in these studies varies considerably. Some studies use images of real faces, either standardised images from established

databases (Figure 1a; Eggleston, Flavell, et al., 2021; Jones et al., 2021; Tsantani et al., 2023; Willis & Todorov, 2006), or so-called 'ambient images' which vary widely in pose, facial expression, lighting conditions, viewpoint, and the presence of make-up and jewellery (Figure 1b; Collova, Sutherland, & Rhodes, 2019; Sutherland et al., 2013; Vernon, Sutherland, Young, & Hartley, 2014). Other studies use synthetic facial images such as those generated by FaceGen Modeller software (e.g., Charlesworth, Hudson, Cogsdill, Spelke, & Banaji, 2019; Cogsdill, Todorov, Spelke, & Banaji, 2014; Jessen & Grossmann, 2016; Oosterhof & Todorov, 2008). These images can be manipulated to exaggerate facial cues associated with particular traits (e.g., apparent trustworthiness, competence and dominance; Figure 1c).

Figure-1

Most studies have measured the influence of facial cues on explicit trait ratings. Commonly assessed traits include trustworthiness (e.g., DeBruine, 2002; Falvello, Vinson, Ferrari, & Todorov, 2015; FeldmanHall et al., 2018; Sakuta, Kanazawa, & Yamaguchi, 2018; Stirrat & Perrett, 2010; Thierry & Mondloch, 2021; Todorov et al., 2009; J. P. Wilson & Rule, 2015), dominance (e.g., Fiala, Tureček, Akoko, Pokorný, & Kleisner, 2022; Jones et al., 2021; Main, Jones, DeBruine, & Little, 2009; Mondloch, Gerada, Proietti, & Nelson, 2019; Oosterhof & Todorov, 2008; Pandeirada, Madeira, Fernandes, Marinho, & Vasconcelos, (2022; Torrance, Wincenciak, Hahn, DeBruine, & Jones, 2014; Witkower, Hill, Koster, & Tracy, 2022; Witkower & Tracy, 2019), and intelligence (e.g., Bar et al., 2006; Eggleston, Flavell, et al., 2021; Talamas, Mavor, Axelsson, Sundelin, & Perrett, 2016; Tsantani et al., 2023).

Until quite recently, the overwhelming majority of studies in this field focussed on the first impressions formed by White individuals from WEIRD cultures (western, educated, industrialised, rich and democratic) about White faces (for review, see: Cook & Over, 2021). While the reasons for the emergence and maintenance of this problematic convention are likely complex, there are positive signs that more recent first impressions research is seeking to incorporate greater diversity (e.g., Charbonneau, Robinson, Blais, & Fiset, 2020; Jones et al., 2021; Oh, Dotsch, Porter,

& Todorov, 2020; Peterson, Uddenberg, Griffiths, Todorov, & Suchow, 2022; Xie, Flake, & Hehman, 2019; Xie, Flake, Stolier, Freeman, & Hehman, 2021).

Different first impressions of faces are often described as either 'idiosyncratic' or 'consensus' (sometimes 'consensual' or 'shared'). Idiosyncratic impressions differ between individuals (e.g., Sutherland, Burton, et al., 2020). For example, we tend to attribute trustworthiness to strangers whose faces resemble our own (DeBruine, 2002). Because we all differ in facial appearance, this self-resemblance effect yields idiosyncratic preferences. Consensus impressions are broadly consistent across individuals (e.g., Zebrowitz, 2017). For example, baby-faced features – round face, large eyes, high eyebrows, small chin and nose – elicit attributions of naivety in observers across many cultures (Zebrowitz McArthur & Berry, 1987). To date, much of the extant work in this field has sought to describe and understand consensus impressions. However, a growing body of evidence suggests that existing research may have underestimated the extent to which our first impressions are idiosyncratic (Albohn, Martinez, & Todorov, 2024; Hehman, Sutherland, Flake, & Slepian, 2017).

2.1 Accuracy

There is considerable debate regarding the extent to which first impressions from facial appearance accurately reflect the character traits of the individuals being judged (Bonnefon, Hopfensitz, & De Neys, 2015; Olivola et al., 2014; Todorov, 2017; Todorov, Funk, & Olivola, 2015). Some results suggest that certain first impressions may contain a 'kernel of truth'. For example, Stirrat and Perrett (2010) measured the extent to which male participants could be trusted within the context of an economic game. In an initial experiment, they found that players with wider faces – specifically, those with greater bizygomatic width – were more likely to exploit the trust of other players by dividing the resources unfairly than were players with narrower faces. In a second experiment, a new group of participants spontaneously judged the players with wide faces to be less trustworthy when shown photographs of their faces. Similarly, Bonnefon, Hopfensitz, and De Neys (2013) found that participants could detect those individuals who acted in an untrustworthy manner in a trust game from cues present in facial photographs.

Where observed, however, empirical effects suggestive of a kernel of truth tend to be small (Foo, Sutherland, Burton, Nakagawa, & Rhodes, 2022; Olivola et al., 2014; Todorov, 2017; Todorov, Olivola, et al., 2015). Although participants may sometimes distinguish trustworthy and untrustworthy individuals at rates that statistically exceed chance, their judgements remain extremely error prone (Todorov, Funk, et al., 2015). Moreover, many studies have failed to find any relationship between how trustworthy an individual appears based on their facial appearance and their actual character traits (e.g., Efferson & Vogt, 2013; Jaeger et al., 2022; Olivola & Todorov, 2010). It has been noted that different images of the same person afford very different trait ratings depending on their pose and expression and the lighting conditions (Lavan et al., 2021; Todorov & Porter, 2014). Such findings are hard to reconcile with the view that facial photographs convey accurate information about stable personality characteristics (Todorov, Funk, et al., 2015).

It has been suggested that some first impressions exhibit a kernel of truth as a result of a self-fulfilling prophecy (Hong, Suk, Choi, & Na, 2021; Li, Heyman, Mei, & Lee, 2019). Some individuals who are treated as though they are untrustworthy may experience a higher proportion of negative interactions and may therefore be more likely to engage in untrustworthy and/or aggressive behaviours. Other individuals may exhibit the opposite pattern, however. For example, some who tend to be judged trustworthy (e.g., those with baby-faced features) may learn they can exploit the first impressions of others for their own advantage.

As we have argued elsewhere, it is likely that systematically excluding the faces of people of colour from first impressions research may have artificially inflated apparent evidence for the kernel of truth hypothesis (Cook & Over, 2021). For example, first impressions based on racist stereotyping are widely thought to have no basis in fact (e.g., Fiske, Cuddy, Glick, & Xu, 2002). As such, incorporating more diverse stimulus sets into research on first impressions would likely reveal the true scale of their inaccuracy.

2.2 Consequences

The consequences of first impressions have been studied in a range of applied contexts. Investigating the influence of first impressions within the criminal justice

system, J. P. Wilson and Rule (2015) asked participants to rate the faces of convicted murderers for apparent trustworthiness. The authors found that those with untrustworthy appearance were more likely to receive death sentences (vs. life imprisonment) than were those with trustworthy appearance (though see Kramer & Gardner, 2020). In closely related lab-based research, Porter et al. (2010) presented participants with vignettes describing crimes accompanied by photographs of defendants who varied in apparent trustworthiness. Participants required less evidence to arrive at a guilty verdict when the defendants appeared untrustworthy. Participants were also more confident of their guilty verdicts when the defendants appeared untrustworthy (Porter et al., 2010).

First impressions from facial appearance have also been shown to influence political and economic decision making. Todorov et al. (2005) found that electoral candidates who appeared more competent to participants from their photographs were more likely to be elected to the US senate. Crucially, participants in these studies were unfamiliar with the candidates and their campaigns suggesting that first impressions of faces may bias election results. This result has since been replicated in a variety of cultural contexts (Ballew & Todorov, 2007; Brusattin, 2012; Lawson, Lenz, Baker, & Myers, 2010; Sussman, Petkova, & Todorov, 2013).

Duarte et al. (2012) found that individuals who appeared trustworthy in their photographs on a peer-to-peer lending site were more likely to have their loans funded. In related lab-based research, authors have found that facial appearance can influence how participants allocate resources in economic games (Chang, Doll, van't Wout, Frank, & Sanfey, 2010; Rezlescu, Duchaine, Olivola, & Chater, 2012; Van't Wout & Sanfey, 2008; R. K. Wilson & Eckel, 2006).

The research discussed in this section has emphasised the many consequences of first impressions from appearance, including their influence on election results (e.g., Ballew & Todorov, 2007) and criminal sentencing (e.g., J. P. Wilson & Rule, 2015). Important as these results are, they are focused primarily on consequences experienced by White people. All too often, the first impressions formed about people of colour contribute to fatal outcomes (e.g., Correll, Park, Judd, & Wittenbrink, 2002). It has long been recognised that police officers in the US are more likely to fatally

injure Black suspects than White suspects. Correll et al. investigated this so-called "shooter bias" in a lab-based setting. They presented participants with a computerised game in which they were instructed to fire at armed targets and decide "not to shoot" unarmed targets. Participants were more likely to mistakenly shoot unarmed Black targets than unarmed White targets. This result can be explained by stereotypical associations between Black people, crime and aggression.

3. The origins of facial impressions: Evolutionary adaptations?

Until recently, there was a dearth of detailed discussion regarding the origins of first impressions. Where the origins question was considered, authors typically sought to explain the existence of consensus impressions – face-trait attributions that are shared within and between cultures – using arguments grounded in evolutionary psychology. According to this view, certain first impressions of others are instinctive adaptations because they conveyed a competitive advantage on our ancestors and were therefore favoured by natural selection.

According to the Babyface Overgeneralisation Hypothesis (Zebrowitz, 2004; Zebrowitz, Fellous, Mignault, & Andreoletti, 2003; Zebrowitz & Montepare, 2006), baby-faced features instinctively elicit attributions of physical weakness, intellectual incompetence, submissiveness, and naivety. Proponents argue that these innate face-trait associations were favoured by natural selection because it was (and is) adaptive to identify and care for young children. In one sense baby-faced appearance can be thought of as an honest signal of physical weakness, intellectual incompetence, submissiveness, and naivety, insofar as young children possess these characteristics. However, these instinctive face-trait mappings are also triggered erroneously by the faces of adults who possess baby-faced facial features, affording attributions of submissiveness and naivety.

Similarly, the so-called Anomalous-Face Overgeneralisation Hypothesis (Zebrowitz, 2004; Zebrowitz et al., 2003) – also referred to as the 'Fitness Overgeneralization Hypothesis' (Zebrowitz & Montepare, 2006) – posits that certain atypical facial features (e.g., asymmetry, flat upturned nose, a smooth philtrum, pronounced epicanthal folds, a thin upper lip, cleft palate, low-set ears, upwards slanting eyes) instinctively elicit attributions of poor health, weak social skills, and intellectual

incompetence. Proponents argue that these innate face-trait associations have been favoured by natural selection because i) these features are characteristic of certain genetic and developmental conditions (e.g., schizophrenia, foetal alcohol syndrome, Down syndrome), and ii) such conditions are associated with poor health, weak social skills, and intellectual incompetence. In other words, these features are an honest signal of what the authors term “low genetic quality” (Zebrowitz & Montepare, 2006, p97). However, these instinctive face-trait mappings are thought to be triggered erroneously by the faces of healthy, typically developing individuals whose facial features bear some subtle resemblance to those characteristic of genetic and developmental conditions; i.e., faces that might be judged unattractive, though not atypical or anomalous.

Another suggestion is that sexually dimorphic characteristics instinctively elicit attributions about male strangers’ likely dominance and aggression (Carré & McCormick, 2008; Geniole, Denson, Dixson, Carré, & McCormick, 2015; Puts, Jones, & DeBruine, 2012). Relative to female faces, male faces tend to have a strong square jawline, a prominent brow ridge, a stronger nasal bridge, greater bizygomatic width, a higher and wider forehead, thinner lips, and slightly darker skin tone. These masculine facial features are thought to elicit attributions of dominance and aggression (e.g., Swaddle & Reierson, 2002). Masculine faces, it is argued, are a product of higher levels of testosterone (e.g., Penton-Voak & Chen, 2004), and high-levels of testosterone are in-turn predictive of greater aggression and physical and social dominance (Eisenegger, Haushofer, & Fehr, 2011; Mazur & Booth, 1998). Instinctive face-trait associations between facial masculinity and dominance / aggression may have been favoured by natural selection because they allowed individuals to quickly infer the relative superiority of males.

According to the Evolutionary-Contingency Hypothesis (Van Vugt & Grabo, 2015), certain face-trait mappings may have been favoured by natural selection, in part, because they helped our ancestors identify the best leaders in different situations: Van Vugt and Grabo (2015) suggest that the instinctive attribution of dominance and aggression to masculine faces may have helped our ancestors identify the best leaders during times of war and conflict. Similarly, the instinctive attribution of trustworthiness to feminine faces may have helped our ancestors identify the best

leaders during periods of peace and cooperation. The attribution of competence to older-looking faces may have helped our ancestors identify the best leaders in knowledge domains (e.g., deciding which ritual should be employed or which medicine to use).

Finally, we note that a number of authors (e.g., Jessen & Grossmann, 2016; Oosterhof & Todorov, 2008; Schaller, 2008; Sutherland, Burton, et al., 2020; Sutherland, Collova, et al., 2020) have alluded to the particular evolutionary significance of valence evaluation, frequently operationalised as attributions of trustworthiness (and also approachability, warmth, agreeableness, and likeability). For example, according to Schaller (2008):

> *"Behaviors that promote genetic reproduction (mating, provision of resources to offspring, etc.) are more difficult to produce if one is injured, destitute, dying, or dead. For that reason, some of the most evolutionarily fundamental psychological goals pertain to the avoidance of (or defense against) other people who might harm us, cheat us, or kill us. That requires that we know – or at least make a reasonable first guess – whether someone is nasty or nice. From an evolutionary perspective, no other kind of inference probably matters quite so much"* (Shaller, 2008, p19-20).

According to this perspective, the need to quickly distinguish friends from foe – and thereby infer whom to approach and whom to avoid – drove the evolution of an adaptive mechanism for the detection of trustworthy faces. Consequently, certain facial features, such as attractiveness, babyface features and facial femininity (lower hairline, smaller eyebrows, bigger eyes, fuller cheeks and lips, a smaller rounded jaw and chin) may instinctively elicit positive attributions (e.g., trustworthiness, likeability).

3.1 Nativist accounts of inaccurate first impressions

The various nativist views outlined above are all predicated on the assumption that certain consensus impressions were favoured by natural selection because they were useful; i.e., that our ancestors were more likely to survive and pass on their genes because they inferred particular traits from certain facial characteristics

(Schaller, 2008; Van Vugt & Grabo, 2015; Zebrowitz, 2004; Zebrowitz & Montepare, 2006). However, the view that the mechanisms responsible conveyed an evolutionary advantage on our ancestors is somewhat at odds with evidence that first impressions typically exhibit little or no accuracy.

Some authors maintain that certain first impressions have enough validity to be adaptive – that facial signals are 'honest enough'. For example, two meta-analyses investigated whether facial-width-to-height-ratio – a sexually dimorphic characteristic thought to inform attributions of dominance and aggression – is an honest signal of these traits (Geniole et al., 2015; Haselhuhn, Ormiston, & Wong, 2015). Although the effect sizes were small, both studies found statistically significant relationships between facial-width-to-height-ratio and threatening and dominant traits. The authors of both studies present their findings as consistent with an "evolved cueing system of intra-sexual threat, dominance, and aggressiveness in men" (Geniole et al., 2015, p15). However, other authors have questioned the interpretation of these results. According to Todorov (2017), for example, "A close reading of the evidence finds little support for evolved honest signals of character in the face" (Todorov, 2017, p187). In line with this view, Kosinski (2017) found no meaningful association between facial-width-to-height ratio and self-reported personality in a sample of 137,136 participants.

A closely related line of argument is that, while trait attributions from faces may be error-prone, the benefits of correct trait attributions outweigh the costs of misattributions. For example, according to the Anomalous-Face Overgeneralisation Hypothesis (Zebrowitz, 2004; Zebrowitz et al., 2003; Zebrowitz & Montepare, 2006), instinctive mappings between atypical facial features and intellectual incompetence are adaptive because they lead people to avoid unfit mates carrying bad genes. While it is unfortunate that these instinctive face-trait mappings are also triggered erroneously by the faces of some healthy, typically developing individuals, these false alarms are a cost worth incurring to avoid unfit mates. Similarly, according to the Babyface Overgeneralisation Hypothesis (Zebrowitz, 2004; Zebrowitz et al., 2003; Zebrowitz & Montepare, 2006), instinctive mappings between baby-faced features and naivety are adaptive because they ensure that we care for our vulnerable young. Once again, it is unfortunate that these instinctive face-trait

mappings are also triggered erroneously by the faces of some adults, but these false alarms are a cost worth incurring to ensure that children are nurtured and protected.

A similar view has been advanced to justify inaccurate inferences of trustworthiness (Schaller, 2008). In particular, it has been suggested that the benefits arising from correct inferences of untrustworthiness outweigh the costs of misattributions of untrustworthiness. For example, Schaller (2008) writes:

> *"Of course, the fact remains that some inferential errors are inevitable. Importantly, different kinds of errors may have different implications for reproductive fitness. When it comes to avoiding social perils, … the failure to detect a real danger (a false-negative error) typically has implications that are far more costly than the detection of a danger that doesn't really exist (a false-positive error). Consequently, just as smoke detectors are calibrated to err on the side on false-positive errors (to trigger an alarm at the merest hint of smoke, even if it that smoke is associated with no real threat whatsoever), psychological mechanisms may have evolved to implicitly err on the side of making false-positive errors when inferring the potentially-dangerous traits or intentions of others"* (Schaller, 2008, p17).

Finally, some authors have argued that, although instinctive face-trait associations may have little or no validity in contemporary society, they may have been accurate in ancestral times (e.g., Van Vugt & Grabo, 2015). For example:

> *"The evolutionary-contingency approach hypothesizes that leadership judgments will vary as a function of the match between environmental demands and the needs of followers. Facial cues may serve as inputs into this system as they are predictive—or probably were, in ancestral times—of the physical and psychological attributes of leaders"* (Van Vugt & Grabo, 2015, p485).

For example, in an environment where conspecifics can detect untrustworthiness from facial appearance, the ability to conceal untrustworthiness from others would convey an evolutionary advantage on would-be bad actors. As such, it is conceivable

that the facial appearance of untrustworthy individuals evolved over time to help them escape detection (e.g., Dawkins & Krebs, 1979). Another possibility is that people who are evolutionarily predisposed to develop certain behavioural profiles (e.g., trustworthy or untrustworthy character traits), do not develop these characteristics in contemporary society because of environmental differences. According to this perspective, instinctive first impressions may be viewed as vestigial as they no longer serve the function for which they evolved.

4. The origins of facial impressions: The cultural learning account

In our recent work, we have argued that first impressions of faces are learned through experience (Cook et al., 2022; Cook & Over, 2020, 2021; Over & Cook, 2018; Over et al., 2020). The TIM framework conceptualises first impressions as mappings between representations of facial appearance – conceived of as locations in multidimensional face space – and representations of the trait profiles that others may possess – characterised as locations in multidimensional trait space. Mappings between points (or regions) in face space and points (or regions) in trait space are acquired through correlated face-trait experience (Figure 2a-c). TIM proposes that when we repeatedly encounter individuals with particular facial features who subsequently reveal themselves to have particular character traits, mappings form between the two representations. Once acquired, these mappings mediate spontaneous first impressions from faces: When we encounter an individual whose face falls close to a mapped region in face space, excitation automatically propagates to the associated representations in trait space (Cook et al., 2022; Over & Cook, 2018). For example, we may learn through experience that a particular teacher is kind and thoughtful. When we subsequently encounter an individual with similar facial features to this teacher, we may assume them to be similarly kind and helpful.

Figure-2

4.1 Cultural learning of consensus impressions

It is widely accepted that learning accounts can explain idiosyncratic first impressions that differ between individuals (Sutherland, Burton, et al., 2020; Sutherland, Collova, et al., 2020). However, learning accounts have traditionally

been criticised for failing to explain consensus impressions – face-trait attributions that are consistent across individuals (e.g., Sutherland, Collova, et al., 2020; Zebrowitz, 2004). According to this critique, consistent first impressions could not arise from learning through direct interaction with individuals because there is little or no relationship between appearance and behavioural traits. When exposed to this kind of environmental input – where appearance is not predictive of character traits and abilities – domain-general associative learning mechanisms would not yield consistent first impressions widely shared by individuals within a community (Todorov, 2017).

TIM resolves this issue through an appeal to cultural learning. TIM proposes that cultural products such as cartoons, illustrated storybooks, films, video games and visual propaganda 'teach' children mappings between particular facial features and certain trait profiles (Figure 2d). Importantly, the face-trait experience that we receive via these cultural products differs considerably from that we encounter in actual social interactions. In particular, facial appearance is far more predictive of heroism and villainy, and physical and intellectual prowess, in these fictional domains than it is in the real world. For example, the depiction of villains in movies frequently – and misleadingly – pairs baldness and facial disfigurement with untrustworthy character (Croley, Reese, & Wagner, 2017). By exposing many people within a community to common sources of face-trait experience, these cultural messages canalise the emergence of erroneous consensus impressions.

It has been suggested previously that the perceptual and cognitive mechanisms responsible for consensus impressions may differ qualitatively from the mechanisms responsible for idiosyncratic impressions (Hehman et al., 2017; Sutherland, Burton, et al., 2020; Xie et al., 2019). Whereas consensus impressions are attributed to "*a target's face*", idiosyncratic impressions are thought to arise "*from our mind*" (Hehman et al., 2017; Xie et al., 2019). Within the TIM framework, however, the mechanisms responsible for consensus impressions are qualitatively identical to those that mediate idiosyncratic impressions. Both are thought to be mediated by mappings acquired during the observer's lifetime that let excitation propagate from a point or region of face space, to points or regions in trait space. Idiosyncratic and consensus impressions may tend to differ in terms of the nature of the correlated

face-trait experience that induces the mapping; for example, many of our idiosyncratic impressions may result from personal experiences of social interaction, while consensus impressions may typically result from exposure to cultural messages. However, in terms of the resulting cognitive mechanisms, these different sources of correlated face-trait experience are thought to afford qualitatively similar face-trait mappings.

TIM places a particular emphasis on the cultural messages encountered by children (Cook et al., 2022; Over & Cook, 2018). Research in the associative learning tradition (e.g., Bouton, 1994, 2002; Bouton & King, 1983; Peck & Bouton, 1990) suggests that 'first learned' associations (i.e., the first rules learned about novel stimuli) are hard to unlearn. Subsequent learning that contradicts the original rule, tends to manifest disproportionately in the learning context. Outside of this context, however, effects of the original learning may still be evident. Thus, the face-trait mappings that children learn early in development may be particularly resistant to change (Over, Lee, Flavell, Vestner, & Cook, 2023). If correct, this feature of TIM suggests that efforts to reduce the malign effects of first impressions should focus on modifying the cultural input available to children.

As explained above, there appears to be little or no systematic relationship between and individual's facial appearance and their character (Olivola et al., 2014; Todorov, 2017; Todorov, Funk, et al., 2015). However, the possibility that certain face-trait judgements possess a kernel of truth remains contested (Bonnefon et al., 2015). In principle, the TIM framework is equally able to explain the emergence of accurate consensus impressions should they exist (Cook et al., 2022; Over & Cook, 2018). Where veridical relationships exist between appearance and traits, TIM predicts that individuals will learn these accurate face-trait mappings through their own social interactions. Even in this case, however, we anticipate an influential role for cultural learning, whereby cultural messages amplify honest signals (i.e., the strength of any face-trait contingency is exaggerated) leading individuals within that society to over-estimate the ability of facial appearance to predict individuals' traits.

4.2. Separate face and trait spaces shaped by experience

TIM hypothesises two distinct multidimensional representation spaces: face space, within which we represent the facial appearance of others (Valentine, 1991; Valentine, Lewis, & Hills, 2016), and trait space, within which we represent the trait profiles that others possess (Conway, Catmur, & Bird, 2019; Fiske, Cuddy, & Glick, 2007; Hassabis et al., 2013). Each dimension within face space is thought to encode a particular source of facial variation; for example, one dimension might code interocular distance, such that faces with relatively large and small interocular distances fall on opposing sides of face space. Similarly, each dimension within trait space is thought to encode a particular characteristic or ability; for example, one dimension might code trustworthiness, such that relatively trustworthy and untrustworthy individuals fall on either side of trait space. Within each space, representations are conceived of as points or mean-relative vectors. The respective centres of face space and trait space represent the average face shape and the average trait profile.

TIM assumes that face space and trait space are both, themselves, shaped by experience and may therefore differ substantially across individuals. A large body of research in vision science converges on the view that the dimensionality of face space is strongly influenced by the perceptual experience of the individual (Furl, Phillips, & O'Toole, 2002; Rhodes & Anastasi, 2012; Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005; Valentine, 1991; Valentine et al., 2016; Webster & MacLeod, 2011). Inspired by data reduction algorithms – for example, principal components analysis – researchers have suggested that the visual system extracts relevant modes of variation from the particular 'diet of faces' encountered by an individual (Calder & Young, 2005; Furl et al., 2002). Where an individual encounters many faces of a particular type, for example primarily East Asian faces, the resulting dimensionality may be optimized to describe this variation (e.g., Furl et al., 2002).

Similarly, the nature of an individual's trait space is likely to be heavily shaped by experience. For example, the acquisition of trait vocabulary likely scaffolds the development of trait space (Over & Cook, 2018). Moreover, many trait constructs are understood differently by individuals in different cultures (John & Srivastava, 1999; Sternberg, 2004; Sternberg & Grigorenko, 2004; Yang & Bond, 1990). Indeed, the

imperfect translation of WEIRD / non-WEIRD trait constructs is a difficulty regularly encountered by authors conducting cross-cultural first impressions research (e.g., Sutherland et al., 2018; Zebrowitz et al., 2012). Relatedly, understanding the extent to which an individual is 'honest' or 'sneaky' clearly relies on sophisticated theory of mind abilities. Inferring the character traits of others represents a substantial challenge – social behaviour is often unpredictable and varies widely according to the context. Even relatively simple judgments such as deciding who is 'nice' and 'mean' often relies on intention understanding, a capacity that is not observed until the second year of life (Carpenter, Akhtar, & Tomasello, 1998).

4.3 Innate contributions to trait inference mapping

Our account does not deny a role for evolutionary factors in the development of first impressions. In particular, TIM allows that certain innate biases may canalise the emergence of common face-trait mappings. For example, so-called 'infant schema' may elicit positive emotions and encourage nurturing responses towards individuals with large eyes and round faces (Glocker et al., 2009). These positive responses may then promote more favourable interpretations of the individuals' behaviour in range of ambiguous situations. Consequently, observers may be more likely to map cute, baby-faced features to positively valanced traits such as trustworthiness (Over & Cook, 2018).

Crucially, however, TIM does not posit an innate mapping between baby-faced appearance and trustworthiness – though many people around the world may acquire such a mapping ontogenetically. Rather, infant schema is assumed to act in the background, exerting a non-specific influence on a range of evaluative judgements. This view is based on two observations. First, the presence of infant schema elicits nurturing behaviours in a range of vertebrate species, including birds (Lorenz, 1943, 1971). This is important because it suggests that this instinctive stimulus-response behaviour emerged long before the capacity to represent others' traits in phylogenetic history. Second, the effects of infant schema do not appear to be face-specific; a host of inanimate objects and simple shapes – including cars (Miesler, Leder, & Herrmann, 2011), watches and sofas (Bar & Neta, 2006), and rectangles (Cho, Dydynski, & Kang, 2022) are evaluated more favourably when given soft, rounded features.

4.4. A cognitive gadget (or cognitive malware?)

The formulation of TIM owes much to the "cognitive gadgets" framework advance by Heyes (2018). Heyes (2018) argues that many cognitive mechanisms previously attributed to genetic inheritance (e.g., those for language, imitation, and mindreading), may instead be products of cultural learning. These 'cognitive gadgets' are specialised mechanisms built by general cognitive processes (e.g., associative learning, attention, executive functions) using information from the sociocultural environment. These culturally acquired cognitive mechanisms are thought to be "gadget-like" in that they are useful (i.e., they do their job reasonably well) and the products of human rather than genetic agency.

Heyes (2018) proposes that cultural change may have shaped and reshaped these mechanisms in the recent past. For example, the invention of cultural artefacts such as mirrors and cultural practices such as group dance promoted the development of the visuomotor associations thought to mediate imitation. In a similar vein, we propose that the capacity to illustrate stories with still (storybooks, comics) and moving (television and film dramatizations, cartoons) pictures caused a surge in correlated face-trait experience in industrialised societies. We speculate that these cultural innovations served to canalize the emergence of consensus impressions, that they led to greater inter-observer consistency in the traits inferred from facial cues by different observers. Before industrialisation, the inferences of traits from faces in these societies may have been more idiosyncratic (Cook et al., 2022; Over & Cook, 2018).

The mechanism responsible for first impressions from faces is a curious example of a cognitive gadget, however, because this culturally acquired mechanism does its job *badly*. If the mechanism responsible for first impressions from faces did its job well it would be extremely useful: it would enable 'zero-trial' learning about the character traits of other individuals. We could, for example, infer that someone was untrustworthy simply from their facial appearance, without the need for potentially costly social interaction. In the vast majority of cases, however, our first impressions are inaccurate and impair decision making (Olivola et al., 2014; Todorov, 2017; Todorov, Funk, et al., 2015). For example, we vote for the candidate who only appears trustworthy (Ballew & Todorov, 2007; Todorov et al., 2005).

The cognitive mechanisms highlighted by Heyes (2018), including those responsible for imitation and theory of mind, may be thought of as cognitive gadgets insofar as they are products of human agency and useful. In contrast, the mechanism responsible for first impressions from faces might be better characterised as "cognitive malware" – also a product of human agency, but one that is deleterious to the individual. Our error-strewn guesses about the likely character traits of strangers exert such a malign influence on our decision making and social behaviour, perhaps we would be better-off without this particular 'gadget'.

The view that dubious 'wisdom' or deleterious practices may be transmitted culturally from one generation to another may strike some as counter-intuitive, but there are many precedents. Superstitions are a good example. The popular wisdom that certain events are unlucky (e.g., Fridays that fall on the 13th day of the month, walking under ladders, breaking mirrors) has been passed from generation to generation for centuries despite questionable validity. Similarly, the view that earthly and human events can be forecasted by observing and interpreting the movements of the moon, the stars, the planets, and comets has been around for millennia in one form or another (e.g., astrology). Indeed, many leading newspapers still publish a daily horoscope.

The potential for such misleading and inaccurate cultural input to shape the development of cognitive mechanisms is an intriguing feature of the Cognitive Gadgets framework. For example, it is hard to see how natural selection and genetic inheritance could produce cognitive mechanisms that are fundamentally deleterious to the individual. However, cultural evolution and transmission might conceivably endow individuals with 'cognitive malware'.

5. The existing evidence base

When seeking to understand the origins of first impressions from facial appearance, researchers have turned to several lines of evidence, including developmental data, cross-cultural findings and the results of training studies. The resulting body of research offers a complex picture in which much, but not all, of the evidence base supports the learning account (Cook et al., 2022).

5.1 Developmental trajectory

If first impressions from facial appearance emerge early in development, before children have had extensive opportunities for learning, this would lend considerable weight to nativist accounts (Cogsdill et al., 2014; Ewing, Sutherland, & Willis, 2019; Sutherland, Collova, et al., 2020). Conversely, a slower, more protracted developmental trajectory would accord well with the cultural learning view.

Developmental data suggest that consensus impressions are measurable in Western children by the preschool years (Cogsdill et al., 2014; Ewing et al., 2019; Jessen & Grossmann, 2016). Cogsdill et al. (2014) presented US children aged between 3 and 10 with pairs of computer generated faces and asked them to make judgments about their relative trustworthiness (which individual appeared nicer), dominance (which individual appeared stronger) and competence (which individual appeared smarter). From the age of 3, children's judgments accorded with those of adults, and by 5- to 6-years-of-age, their judgments showed similar consistency to those of adults. Similarly, Charlesworth et al. (2019) found adult-like levels of consensus in first impressions of trustworthiness, competence and dominance in US children by 5 years of age. According to some authors, these data demonstrate that "extended cultural learning of appearance trait-mappings […] is not necessary for adult-like appearance biases to emerge" (Ewing et al., 2019, p1699). Rather, the reported findings are "more consistent with evolutionary-based accounts, wherein selection pressures to rapidly establish whether others appear likely to help or harm us may have shaped social biases that emerge relatively early in life" (Ewing et al., 2019, p1700).

Five-year-olds have had considerable opportunities for social learning, however. Within Western cultural contexts, many five-year-olds are able to engage in elaborate activities that are uncontroversially the product of learning, for example reading a written language and navigating an iPad (Heyes & Frith, 2014). Evidence that infants < 12 months-of-age exhibit spontaneous first impressions would afford a more convincing "poverty of the stimulus" argument (Thomas, 2002). It is noteworthy, therefore, that German infants as young as 7 months preferred to look at faces previously rated by adults as trustworthy (Jessen & Grossmann, 2016). In a follow-up study, Sakuta et al. (2018) found that 6- to 8-month-old Japanese infants

preferentially attended to faces judged trustworthy by adults relative to those judged untrustworthy, albeit only when the faces were also judged high in dominance. Superficially at least, these data are suggestive of innate mechanisms that enable first impressions from appearance.

Other studies, however, suggest a protracted developmental trajectory broadly compatible with a learning account. For example, a systematic review and meta-analysis of 10 studies (representing 1,325 children aged 3–12, and 851 adults aged 17–81) concluded that facial impressions of trustworthiness continue to develop throughout childhood reaching adult-like levels of consistency only between 10 and 13 years of age (Siddique et al., 2022). Attributions of competence and dominance from facial appearance appear to follow a similar developmental trajectory in US samples (Cogsdill et al., 2014).

5.2 Cross-cultural approaches

Evidence for broad cross-cultural agreement in first impressions, despite variable learning experiences, would lend support to nativist accounts. Several authors have reported findings broadly consistent with this possibility (e.g., Hester, Xie, & Hehman, 2021; Jones et al., 2021; Sutherland et al., 2018; Walker, Jiang, Vetter, & Sczesny, 2011; Zebrowitz et al., 2012). For example, Zebrowitz et al. (2012) compared first impressions formed by US undergraduate students and the culturally isolated Tsimane living in Bolivia, of US and Tsimane male faces. The results revealed signs of cross-cultural agreement. For example, both the US students and the Tsimane judged individuals with more attractive faces to be more warm (sociable) and more intelligent (knowledgeable). Moreover, this was true irrespective of the type of face being judged (American vs. Tsimane).

Having developed image processing algorithms that accentuated the perception of certain facial traits in the eyes of Western observers, Walker et al. (2011) sought to determine whether Asian observers would respond to these facial manipulations in a similar way. Western and Asian participants were shown pairs of faces (both Western and Asian in appearance) that had manipulated to appear high and low in aggressiveness, extroversion, likeability, risk seeking, social skill, and trustworthiness. Although the algorithms had been designed to manipulate the

perception of these traits in Western observers, the Asian observers were also able to recognise the high and low variants for each trait, suggestive of some cross-cultural similarity in first impressions of face.

Sutherland et al. (2018) examined the first impressions made by British and Chinese observers when judging White British and Chinese faces. Once again, there were clear signs of cross-culture similarities. In particular, the perception of approachability – a latent dimension that encapsulates attributions of friendliness, niceness, warmth, kindness – was positively influenced by greater facial femininity and the presence of a smile. Conversely, faces judged less approachable tended to be more masculine and depict a sullen expression or a scowl. A similar pattern was seen for British and Asian observers irrespective of the type of face being judged.

Finding such as these are often cited as decisive evidence in the context of the origins debate. For example:

> *"The evidence provided for similar trait impressions from faces across Tsimane' and U.S. judges indicates that some universal mechanism guides these impressions"* (Zebrowitz et al., 2012, p132).

> *"… these remarkably different cultures [US students and Tsimane] formed highly similar impressions, suggesting that at least some contingencies between cues and impressions are evolutionarily predisposed"* (Sutherland, Collova, et al., 2020, p16115).

This conclusion seems premature, however. Importantly, there is also striking evidence of cultural variability (Scott et al., 2014; Sofer et al., 2017; Sutherland et al., 2018; Walker et al., 2011; Zebrowitz et al., 2012).

Consider the study described by Zebrowitz et al. (2012). When rating the White faces for dominance (respect) and warmth (sociability), the authors report that the US undergraduates exhibited extremely high inter-rater agreement (Figure 3a). When judging the same faces, however, the inter-rater agreement exhibited by the Tsimane people failed to reach statistical significance (Zebrowitz et al., 2012). When

judging the White faces, the US students judged babyfaced targets to be relatively warm and submissive. An effect of target babyfacedness on perceived warmth was also seen when the US students judged the Tsimane faces. However, the Tsimane raters showed no effect of target babyfacedness on any trait rating (knowledgeability, respect, or sociability), for either the Tsimane or White US faces (Zebrowitz et al., 2012). This is particularly striking as the 40 US faces used in this study purposely included 10 examples of high and low babyfacedness, respectively.

It is also noteworthy that Zebrowitz et al. (2012) needed to use different trait terms to measure first impressions when studying US American and Tsimane participants. Whereas US participants were asked to rate the faces on apparent warmth, intelligence and dominance, Tsimane participants were asked to rate the faces on sociability, knowledge and respect. These contrasting terms were necessary because abstract trait terms such as 'intelligent' were deemed to be culturally irrelevant to Tsimane.

Similarly, consider the study described by Walker et al. (2011). As described above, the Asian observers were able to recognise the high and low variants for each trait at rates that clearly exceeded chance. The fact that Western trait manipulations are somewhat effective on Asian participants is suggestive of some cross-cultural agreement. However, the Asian participants also exhibited significantly lower identification scores than Western participants – particularly for extroversion, social-skill, and trustworthiness (Figure 3b) – suggestive of cross-cultural variability. The Asian participants also took longer to make trait inferences from faces, suggesting that their trait inferences may have been less automatic. This possibility accords with the view that the understanding of social behaviour is less reliant on trait-based explanations in Asian cultures (Walker et al., 2011).

A close reading of Sutherland et al. (2018) also reveals evidence of cross-cultural differences. When judging the White British faces, the trait ratings of the Chinese and British participants accorded closely: A similar 3-factor structure emerged in the judgements made by the two groups (characterised as Approachability, Youthful-attractiveness, and Capability). Moreover, British and Chinese participants positioned White faces at similar locations within this dimensionality. However, when

judging Chinese faces the ratings of the Chinese participants showed higher dimensionality than those of the British participants, consistent with more differentiated impressions. In an earlier phase of their study, Sutherland et al. (2018) asked British and Chinese participants to describe the traits of own-race faces in their own words. Interestingly, while the Chinese descriptions were quite variable, there was greater consistency in the descriptors offered by the British group: certain terms (e.g., friendly, kind, intelligent, warm) were frequently offered by different participants.

We know from research in social psychology that there are systematic individual and cultural differences in intergroup biases (Over & McCall, 2018). These individual and cultural differences appear to influence first impressions of diverse faces in predictable ways (Sofer et al., 2017; Tsantani et al., 2023; Xie et al., 2021; Zebrowitz, Montepare, & Lee, 1993). For example, a set of Black faces were judged less intelligent by White British participants than by Black British participants, while a set of White faces were judged less likable by Black British participants than by White British participants (Tsantani et al., 2023). Similarly, Sofer et al. (2017) had Israeli and Japanese participants rate the trustworthiness of faces that varied systematically from Japanese-typical to Israeli-typical. Both groups of participants judged own-culture typical faces to be more trustworthy than other-culture typical faces.

Consistent with the cultural learning account, some consensus impressions appear to manifest more strongly in cultures in which individuals are more likely to encounter cartoons, illustrated storybooks, films, video games and visual propaganda. Scott et al. (2014) compared the extent to which facial masculinity was associated with perceived aggression in 12 societies with very diverse levels of economic development. Participants from highly industrialized / urbanized communities (e.g., Shanghai, U.K., Canada) attributed aggression to masculine faces more strongly, than those from smaller non-industrialized communities (e.g., the Aka – a foraging community from the Central African Republic, and the Tchimba – a pastoral community from Namibia). Across the 12 societies studied, the authors found that urbanization (the % of the population living in urban areas) was highly predictive of the strength of the masculinity-aggression stereotype, accounting for ~90% of the variance observed. Note, the results described by Zebrowitz et al. (2012; Figure 3a)

suggest this possibility for other first impressions, including attributions of dominance (respect) and warmth (sociability).

Figure-3

5.3 Training studies

Training studies with adults lend further credence to the claim that first impressions can be acquired or modified through experience. For example, Verosky and Todorov (2010) had participants complete a training procedure during which particular facial identities were paired with positive (e.g., “He gave his balloon to a child who had let hers go”) and negative (e.g., “He stole money and jewellery from the relatives he was living with”) behavioural descriptions. During a subsequent test phase, participants were asked to judge faces that looked similar to those used in training. Test faces resembling training faces paired with positive behaviours were deemed more trustworthy than test faces resembling training faces paired with negative behaviours. A number of similar results have been reported elsewhere, both with faces (Chua & Freeman, 2022; FeldmanHall et al., 2018; Gawronski & Quinn, 2013) and synthetic ‘Greeble’ stimuli with which participants have little or no perceptual expertise at the outset (Lee, Flavell, Tipper, Cook, & Over, 2021; Over et al., 2023).

Training paradigms have also been used to examine whether first impressions can be acquired through cultural learning mechanisms. Eggleston, Geangu, Tipper, Cook, and Over (2021) presented 5- to 7-year-old participants with images of emotional-neutral ‘target faces’ flanked on either side by expressive ‘context faces’. All the facial images depicted children. Some target faces were paired with smiling context faces, while others were paired with fearful context faces. At test, participants judged the trustworthiness of new faces that were similar in appearance to the target faces on which they had been trained. The to-be-judged faces that resembled the target faces shown with smiling peers were rated more trustworthy than to-be-judged faces that resembled the target faces shown with fearful peers. The implication is that children may infer aspects of a stranger’s character by observing their interactions with others (e.g., their friends and/or caregivers). This ‘social referencing’ might contribute to the transmission of face-trait stereotypes within a community (Over & Cook, 2018).

There has also been considerable interest in whether first impressions can be 'unlearned' through periods of counter-stereotype training – as would be predicted by the cultural learning account (Over & Cook, 2018). To date, results have been mixed: While some findings suggest that training interventions can weaken face-trait associations (Chua & Freeman, 2021), others suggest that face-trait associations may be resistant to training interventions (Jaeger, Todorov, Evans, & van Beest, 2020). Where counter-stereotype training is ineffective, one potential explanation comes from the study of 'renewal' in the associative learning literature (Bouton, 1994, 2002). In short, new learning that contradicts old learning often manifests selectively in the context in which the new learning occurs (Over et al., 2023; Rydell & Gawronski, 2009). Similar ideas have been advanced to understand why people addicted to drugs are prone to relapse when they leave clinical rehabilitation settings and return to their home environments (e.g., Bouton, 2002).

5.4 Other lines of evidence

Two further lines of evidence are worth mentioning. First, as described in Section 2, adults form consistent first impressions from faces even when those faces are presented for as little as 100 milliseconds (Bar et al., 2006; Todorov et al., 2009; Willis & Todorov, 2006). Related research demonstrates that these first impressions occur automatically. For example, facial appearance informs first impressions of an individual's character even when observers are instructed to ignore the facial images and focus only on their voice (Mileva, Tompkinson, Watt, & Burton, 2018). Some researchers have argued that the speed and automaticity of first impressions from faces is most compatible with an evolutionary account (Schaller, 2008; Zebrowitz & Montepare, 2006; Zebrowitz & Zhang, 2011). According to this point of view, if someone has aggressive or nefarious intentions, it serves an organism to detect those intentions as quickly as possible (e.g., Schaller, 2008).

However, speed and automaticity are not uniquely compatible with a nativist account (Over & Cook, 2018). Some learned skills can become fast and automatic with practice. In adults, reading is a prototypical example of a learned skill and yet it occurs quickly and is difficult to inhibit (Heyes & Frith, 2014; Stroop, 1935). Moreover, first impressions that must be learned also exhibit these features.

Eggleston, Flavell, et al. (2021) asked participants to judge the relative intelligence of individuals, some of whom were wearing glasses (Figure 3). Replicating previous research, the individuals wearing glasses were judged to be more intelligent than those who were not (Fleischmann, Lammers, Stoker, & Garretsen, 2019). Importantly, this effect also occurred automatically (participants' intelligence judgements were biased despite explicit instructions to ignore the presence of glasses) and after brief presentation (when stimuli were shown for only 100 ms). It is implausible that the fast, automatic influence of glasses on attributions of intelligence could be the product of innate mechanisms because glasses were created relatively recently in human history (Eggleston, Flavell, et al., 2021; Over & Cook, 2018).

Figure-4

Second, neuroimaging studies have identified several neural regions that appear to be involved in the formation and analysis of facial impressions including the amygdala (Engell, Haxby, & Todorov, 2007), posterior cingulate cortex (Schiller, Freeman, Mitchell, Uleman, & Phelps, 2009), and the middle temporal gyrus (Chwe, Vartiainen, & Freeman, 2024). Findings that first impressions are associated with activity in specific regions of the brain has been cited as consistent with an evolutionary origin (Schaller, 2008). Again, however, this argument does not withstand scrutiny. Many learned skills also engage specific cortical regions (Heyes, 2018). For example, reading engages the visual word form area, a region of the left fusiform gyrus (in the ventral occipitotemporal cortex) thought to play a crucial role in visual word recognition (Dehaene & Cohen, 2011). In the context of the origins debate, the fact that making trait attributions elicits neural activity in specific regions does not favour one view or the other.

6. Trait inferences from appearance vs. trait inferences from behaviour

When asked to report their first impression of an actor depicted in a photograph, observers can base their judgement on two types of cues: stable appearance cues – what does the actor look like? – and /or transient behaviour cues – what is the actor doing? (Figure 4). That an individual depicted smiling is more trustworthy than an individual depicted scowling is conceptually similar to the inference that someone depicted waving a switch blade is less trustworthy than someone depicted holding a

bouquet of flowers – in both case the judgement is based on ‘thin slices’ of observed behaviour (Ambady & Rosenthal, 1992). Whereas first impressions based on stable appearance cues may be likened to *judging a book by its cover*, impressions based on behaviour cues are more akin to *judging a book from its contents*, albeit only a few lines of the first page (Cook et al., 2022).

The origins debate described above relates to the inference of traits from stable appearance cues (e.g., why individuals with large eyes and round faces are often judged more trustworthy than individuals with more angular features), not trait inferences from behaviour cues (e.g., why someone shown scowling is judged less trustworthy than someone shown smiling). To decide between the two accounts in the origins debate, it is therefore necessary to clearly distinguish first impressions based on facial appearance from first impressions based on behavioural cues. Regrettably, however, existing research on first impressions has not distinguished between these different types of inference as clearly as it could have. Below we illustrate the nature of this problem and the resulting difficulties, with reference to three key lines of evidence in the origins debate: developmental emergence, cross-cultural agreement and training studies.

6.1 Implications for developmental emergence

It has previously been reported that the tendency to make inferences about the apparent trustworthiness of others from their facial appearance emerges during middle infancy (Section 5.1). For example, Jessen and Grossmann (2016) report that 7-month-old infants prefer to look at computer-generated faces that adults deem to be trustworthy than similar faces that adults deem to be untrustworthy. These findings were replicated and extended by Sakuta et al. (2018) who found that found that 6–8 month-old infants attended to trustworthy faces relative to untrustworthy faces – replicating the results of Jessen and Grossmann – but only when faces were high in dominance. The stimuli used in these studies are shown in Figure 5.

Figure-5

Jessen and Grossmann claim that these results reflect first impressions from physical appearance rather than expression:

> "*While it is unlikely that infants possess an elaborate concept of trustworthiness, they do differentiate between trustworthy and untrustworthy faces based on subtly different featural combinations… In this context, it is important to consider that facial trustworthiness detection is based on invariant (stable) facial information rather than the variant (transient) facial information"* (Jessen & Grossmann, 2020, p457).

If the characterisation offered by Jessen and Grossmann is correct, this finding would be hard to reconcile with the view that first impressions of faces are learned culturally. Because ~7-month-old infants have had little opportunity to have learned face-trait mappings culturally, such a result would imply that some first impressions have an innate basis.

However, follow-up work by Eggleston, Tsantani, Over, and Cook (2022) demonstrated that the faces described as trustworthy by Jessen and Grossman were perceived by adults as happier than the faces described as neutral and untrustworthy. The faces described as untrustworthy by Jessen and Grossman were also perceived by adults as angrier than the faces described as neutral and trustworthy. Similarly, the high-dominance trustworthy faces used by Sakuta et al. (2018) were judged to be more happy and less angry than the high-dominance untrustworthy faces. This confound was much weaker in the authors' low-dominance stimuli that failed to induce the same preferential looking bias in infants (Eggleston et al., 2022).

Thus, contrary to the view of Jessen & Grossman, there is every reason to believe the preferential looking effects described are products of variant (transient) facial information rather than the invariant (stable) facial information. It has long been known that by 5 months infants show some basic understanding of expression valence as revealed through their gaze behaviour (Bornstein & Arterberry, 2003; Montague & Walker-Andrews, 2001). Evidence that ~7-month-old infants fixate more on trustworthy faces that show signs of positive emotion (Jessen & Grossmann, 2016; Sakuta et al., 2018) might very well reflect a preference for smiling faces, not trustworthy facial structure.

Figure-6

6.2 Implications for cross-cultural agreement

Another common assertion is that individuals from different cultures make broadly similar trait inferences when judging the same faces (Jones et al., 2021; Sutherland, Collova, et al., 2020; Sutherland et al., 2018; Walker et al., 2011; Zebrowitz et al., 2012; Zebrowitz & Zhang, 2011; Zebrowitz McArthur & Berry, 1987). Apparent evidence that face-trait judgements exhibit cultural universality potentially accords with an innate origin and argues against the cultural learning view, which predicts substantial cross-cultural variability (Section 5.2).

Once again, however, the stimuli used in these cross-cultural studies are often rich in behaviour cues. For example, the ambient facial stimuli used by Sutherland et al., not only depict a host of different facial identities, but also vary widely in the actors' expression and degree of head-tilt – a mixture of appearance and behaviour cues. The presence of salient behaviour cues may artificially inflate levels of cross-cultural agreement in situations where observers derive very different impressions from stable facial appearance cues (e.g., mouth shape, nose size, facial width, interocular distance, skin tone). People from different cultures around the world produce broadly similar emotional expressions and infer similar meanings from these displays (Cowen et al., 2020; Jack, Sun, Delis, Garrod, & Schyns, 2016). Thus, it is unsurprising that people around the world also judge smiling strangers to be more trustworthy and more approachable than strangers who scowl. People around the world also perceive head-tilt as a cue to dominance (Witkower et al., 2022). Thus, the presence of head-tilt variability may similarly inflate cross-cultural agreement in judgements of leadership and dominance.

6.3 Implications for training studies

According to TIM, first impressions of faces are the product of face-trait mappings acquired within the lifetime of the observer (Cook et al., 2022; Over & Cook, 2018). If this view is correct, there should be some scope to attenuate the malign influence of consensus impressions through counter-stereotypical training. The evidence for this claim is mixed (Section 5.3). While some authors have described positive benefits of

training interventions (e.g., Chua & Freeman, 2021), others have found that first impressions are resistant to training interventions (e.g., Jaeger et al., 2020).

As we allude to above, one potential explanation for training ineffectiveness comes from the study of 'renewal' in the associative learning literature (Bouton, 1994, 2002). In short, new learning that contradicts old learning often manifests only in the context in which the new learning occurs (Over et al., 2023; Rydell & Gawronski, 2009). However, another possibility is that the outcome of counter-stereotypical training depends on the face stimuli used, and consequently the nature of the face-trait inference studied. We speculate that first impressions from appearance may be modified through counter-stereotypical training, because these first impressions are learned within the lifetime of the individual. In contrast, first impressions from expression cues may be more resistant to training interventions, because these first impressions owe more to innate factors.

7. Emotion over-generalization

Thus far, we have argued that researchers should take greater care to distinguish impressions based on stable facial appearance cues (e.g., facial shape or structure) and those based on observed facial behaviours (e.g., facial expressions, head tilt, gaze direction). However, some stable face cues are sometimes mistaken for signs of facial expression. For example, people who have a mouth that naturally curves upwards at its corners may be thought to be smiling when they are not. Likewise, when we encounter a stranger whose eyes are unusually close-together we may erroneously believe they are scowling at us. Where observed, these 'misperceived' expression cues exert a powerful influence on our first impressions. For example, the two individuals described may be judged to be warm and aggressive, respectively (Montepare & Dobish, 2003; Said, Sebe, & Todorov, 2009; Todorov, 2008). In popular vernacular, the influence of misperceived facial expression cues on trait inferences is sometimes referred to as "resting bitch face". However, in the academic literature on first impressions this is known as "emotion over-generalization" (Montepare & Dobish, 2003; Zebrowitz, 2004).

These face-trait inferences are somewhat ambiguous insofar as they possess features of both appearance-based and behaviour-based impressions. From the

point of view of the judged individual, trait inferences based on misperceived facial emotion are appearance-based: the person is judged harshly through no fault of their own, and short of surgery to alter their appearance, there is little they can do about it. These first impressions can result in systematic discrimination against the judged individual (Olivola et al., 2014; Todorov, Olivola, et al., 2015). For example, individuals judged to be untrustworthy because they possess narrow eyes may encounter a lifetime of unfair treatment in a variety of contexts (e.g., educational, financial, professional, interpersonal, judicial). Efforts to understand and remediate this kind of first impression are therefore worthy and should continue.

From the point of view of the observer making the inference, however, this kind of first impression is behaviour-based. In this context, it is important to remember that visual perception is inferential and probabilistic. At any point in time, our subjective perception reflects our brain's best guess at the contents of the environment but one that is frequently wrong and / or incomplete (Clark, 2013; De Lange, Heilbron, & Kok, 2018; Gregory, 1997). The perception of unfamiliar faces – in particular, those depicted in static 2D images – poses an enormous computational challenge. Faces are highly complex 3D shapes defined by a series of convex and concave surfaces. Moreover, the illumination conditions (e.g., the number, direction, and intensity of the light sources) and pose (e.g., degree of head tilt, expression) are highly variable and frequently change from moment-to-moment. Under these conditions, the accurate perception of 3D face shape requires some impressive perceptual algebra whereby incoming sensory information is combined with prior knowledge about likely face shapes, poses, and illumination conditions, to derive a probabilistic solution – a best guess.

Were it possible to monitor the muscles of the to-be-judged face or carefully examine how it changes over time in response to different situations, one could establish whether the individual depicted is actually scowling / smiling or whether they simply have an unusual facial shape. However, when viewing a photographic image of a stranger's face, study participants have no way to establish the ground truth empirically. Indeed, when viewing synthetic faces (e.g., Figure X), there simply is no ground truth. Instead, observers in this situation must 'guess' – or rather their visual system must infer – the to-be-judged person's likely face shape, pose and

expression from the available sensory evidence and their previous experience. When presented with a stimulus image depicting an unknown person with an unusual face shape expressing no emotion, the vast majority of observers will quite reasonably perceive a person with a statistically more likely face shape expressing emotion (Cook et al., 2022).

From a mechanistic point of view – and in the context of the origins debate discussed here – trait inferences based on misperceived expression cues must be seen as behaviour-based even when the source of the facial variation is in fact structural (or is intended to be so by stimulus creators). In the previous section, we argued that the inclusion of emotion cues in to-be-judged facial images inflates levels of inter-rater and cross-cultural agreement (e.g., Jones et al., 2021; Sutherland et al., 2018; Zebrowitz et al., 2012) and may be responsible for the differential responses of 7-month-old infants to faces deemed trustworthy and untrustworthy by adults (Jessen & Grossmann, 2016). The same is true of misperceived expression cues. In this context, it makes no difference whether the actor depicted was asked to pose a neutral expression – all that matters is the subjective interpretation of stimuli by study participants.

In our view, the description and discussion of this kind of inference often obscures what is going on inside the head of the observer; namely, the detection and interpretation of (subtle) emotion cues. Consider the following quote from Jessen and Grossmann (2016, p1728): “At the mechanistic level, trustworthiness evaluations are considered to rely on an overextension of our ability to respond to facial expressions.” Similarly, “…face evaluation is an extension of functionally adaptive systems for understanding the communicative meaning of emotional expressions” (Todorov, 2008, p209). In what sense is the expression recognition system is being “extended” or “over-extended” in these situations? As we have argued above, the most likely account would appear to be that study participants are simply perceiving and responding to signs of facial emotion.

8. Recommendations for future research

To decide between the two accounts in the origins debate, it is necessary to distinguish first impressions based on facial appearance from first impressions based

on behavioural cues (e.g., expression, head-tilt). While we recognize this represents a formidable challenge, we believe it is tractable. How might the influence of appearance cues be isolated?

It is impossible to eliminate all behaviour cues from a facial photograph; for example, even a so-called neutral expression is meaningful behaviour (Albohn, Brandenburg, & Adams Jr, 2019; Carrera-Levillain & Fernandez-Dols, 1994; Carvajal et al., 2013; Rohrbeck, Kersting, & Suslow, 2023; Tae, Nam, Lee, Weldon, & Sohn, 2020). Consequently, facial photographs should always be thought of as compound stimuli, simultaneously depicting stable appearance cues – what does the actor look like? – and transient behaviour cues – what is the actor doing? However, it may be possible to control for their influence by standardizing the behaviour cues across all to-be-judged faces.

To date, most authors have sought to achieve this by depicting all to-be-judged faces with neutral expressions (e.g., Figure 1a). Where authors choose to persevere with this approach, it is important that stimulus sets are rigorously rated to determine whether observers perceive all stimuli to be neutral, or whether some items appear to be happy, sad, angry, fearful, etc. Such ratings should be obtained using sensitive procedures capable of revealing the perception of subtle expression cues (Eggleston et al., 2022).

While the use of ‘neutral’ faces may seem intuitive, however, it may not be the most effective. The perception of facial expressions is thought to be categorical (Etcoff & Magee, 1992). In other words, observers exhibit heightened perceptual sensitivity to small physical differences that cause two stimuli to fall into different categories. Conversely, small differences between stimuli that fall within the same category are much harder to detect. For example, the differences between tokens of fear are less salient than the differences between subtle displays of fear and happiness. Importantly, the neutral expression is a ‘tiny island’ in expression space, at the junction of multiple emotion categories. As such, subtle deviations present in a stimulus set can cause observers to categorise one neutral face as angry, another as sad, another as happy, and so on. This makes it hard to standardize facial behaviour cues across to-be-judged faces.

Instead, authors may be better off using happiness as the standard expression (as opposed to ‘neutral’). Provided all to-be-judged stimuli exhibit the same expression, the nature of the emotion signal (e.g., neutral vs. happiness vs. anger) makes little difference. Importantly, however, facial happiness is rarely confused with other emotional expressions; hence, there is a good chance the emotion signal present in each face will be categorised similarly. Moreover, where all exemplars fall within the same emotion category (e.g., happiness), residual inter-stimulus expression differences should be far less salient. While this approach is imperfect – observers are able to detect subtle differences in real vs. fake smiles (Song, Over, & Carpenter, 2016) – it leverages the categorical nature of expression perception to help authors standardize facial behaviour cues across to-be-judged faces.

Another option is to provide participants with an array of photographs simultaneously depicting each to-be-judged face with a set of expressions (e.g., happiness, sadness, surprise, fear, anger, & disgust; Figure 7a). By presenting to-be-judged individuals in a variety of poses, authors may avoid any implicit message that ‘this person tends to scowl’ or ‘this person smiles often’ that may influence trait attributions. Moreover, illustrating how an individual’s facial appearance varies across different poses appears to help observers form an accurate perceptual representation of their facial structure (Burton, Jenkins, Hancock, & White, 2005; Murphy, Ipser, Gaigg, & Cook, 2015).

In the scenario outlined in the previous paragraph, multiple exemplars of each individual face are presented simultaneously to help observers infer the true structure of the target face before they record their trait judgement. However, there may also be value in getting each exemplar rated separately; e.g., each target face could be judged six times, once expressing happiness, once expressing sadness, once expressing surprise, once expression fear, once expressing anger, and once expressing disgust. From the resulting distributions of ratings, the influence of facial structure and facial expression on trait attributions could be modelled separately for each face (Figure 7b).

Figure-7

It is unclear whether the independent contributions of facial appearance and facial behaviour can be isolated using so-called ambient image approaches (Sutherland et al., 2018; Sutherland et al., 2013; Vernon et al., 2014). The stimuli used in these studies depict the naturalistic variation that arises in everyday facial photographs (Figure 1b). Under this approach, there is no attempt to control the expressions, head tilt, and gaze direction of the to-be-judged faces; rather, facial behaviours are allowed to vary freely. Consequently, where an image is judged to depict a trustworthy or dominant individual, it is impossible to know whether these impressions are based on the target's facial appearance or their facial behaviour, or a combination of the two. Similarly, where cross-cultural agreement is observed across raters it is impossible to know whether consensus is a product of targets' facial appearance or their facial behaviour; where adult-like impressions are evident early in development, it is impossible to know whether this emergence is a product of the targets' facial appearance or their facial behaviour. While the use of ambient images has proved useful in understanding the kinds of trait attributions made spontaneously about natural images (Sutherland et al., 2018; Sutherland et al., 2013; Vernon et al., 2014), this kind of approach – where appearance cues and behaviour cues are all mixed up together – is unlikely to yield compelling insights into the origin and mechanisms of appearance-based impressions.

9. Conclusion

TIM seeks to understand the origins of first impressions from facial appearance. According to TIM, first impressions are the product of mappings representations of facial appearance (points in face space) and representations of the trait profiles that others may possess (points in trait space) acquired within the lifetime of the observer following correlated face-trait experience (Cook et al., 2022; Over & Cook, 2018). Many of these mappings are likely idiosyncratic products of our unique social interaction histories. However, some mappings may be acquired through exposure to cultural instruments (e.g., children's storybooks, video games, film, and visual propaganda) that pair particular appearance cues (e.g., round eyes or pale skin tone) with particular trait profiles (e.g., being kind or honest) and therefore widely shared within communities.

A great deal of the existing evidence supports the cultural learning account of first impressions (e.g., signs of cross-cultural differences and protracted development). However, by systematically confounding facial appearance (what does the to-be-judged person look like?) and facial behaviour cues (what is the to-be-judged person doing?) previous research may have inadvertently ‘stacked the deck’ in favour of evolutionary accounts of first impressions. The confounding of appearance and behaviour may have obscured cross-cultural differences and yielded misleading evidence that appearance-based impressions emerge during middle infancy.

Appearance-based and behaviour-based impressions are likely mediated by qualitatively different mechanisms and may have very different origins (Cook et al., 2022; Eggleston et al., 2022). To advance the origins debate, future first impressions research must do more to isolate the independent contributions of facial appearance and facial behaviour cues. This will not be easy and may require the use of new face evaluation procedures. It may, however, be the only way to achieve meaningful progress.

TIM provides an optimistic view of the possibility for social change and stereotype reduction. One important aspect of the cultural learning view is that it suggests cultural products depicting correlations between appearance and behaviour do not merely reflect our pre-existing biases, they are crucial in forming them. If we modify cultural input to reduce or eliminate correlations between appearance and character, then first impressions from appearance may be weakened and their malign social effects reduced (Cook et al., 2022; Over & Cook, 2018).

References

Albohn, D. N., Brandenburg, J. C., & Adams Jr, R. B. (2019). Perceiving emotion in the “neutral” face: a powerful mechanism of person perception. In *In The social nature of emotion expression: What emotions can tell us about the world* (pp. 25-47): Springer International Publishing.

Albohn, D. N., Martinez, J. E., & Todorov, A. (2024). Determinants of shared and idiosyncratic contributions to judgments of faces. *Journal of Experimental Psychology: Human Perception and Performance*.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256-274.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the United States of America, 104*(46), 17948– 17953.

Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science, 17*(8), 645-648.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*(2), 269-278.

Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2013). The modular nature of trustworthiness detection. *Journal of Experimental Psychology: General, 142*(1), 143–150.

Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2015). Trends in Cognitive Sciences. *Face-ism and kernels of truth in facial inferences, 19*(8), 421-422.

Bornstein, M. H., & Arterberry, M. E. (2003). Recognition, discrimination and categorization of smiling by 5-month-old infants. *Developmental Science, 6*(5), 585-599.

Bouton, M. E. (1994). Context, ambiguity, and classical conditioning. *Current Directions in Psychological Science, 3*(2), 49-53.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biological Psychiatry, 52*, 976-986.

Bouton, M. E., & King, D. A. (1983). Contextual control of the extinction of conditioned fear: tests for the associative value of the context. *Journal of Experimental Psychology: Animal Behavior Processes, 9*, 248-265.

Brusattin, L. (2012). Candidate visual appearance as a shortcut for both sophisticated and unsophisticated voters: Evidence from a Spanish online study. *International Journal of Public Opinion Research, 24*(1), 1-20.

Burton, A. M., Jenkins, R., Hancock, P. J., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology, 51*(3), 256-284.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience, 6*(8), 641-651.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21*(2), 315-330.

Carré, J. M., & McCormick, C. M. (2008). In your face: facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proceedings of the Royal Society B: Biological Sciences, 275*(1651), 2651-2656.

Carrera-Levillain, P., & Fernandez-Dols, J. M. (1994). Neutral faces in context: Their emotional meaning and their function. *Journal of Nonverbal Behavior, 18*(4), 281-299.

Carvajal, F., Rubio, S., Serrano, J. M., Ríos-Lago, M., Alvarez-Linera, J., Pacheco, L., & Martín, P. (2013). Is a neutral expression also a neutral stimulus? A study with functional magnetic resonance. *Experimental Brain Research, 228*, 467-479.

Chang, L. J., Doll, B. B., van't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology, 61*(2), 87-105.

Charbonneau, I., Robinson, K., Blais, C., & Fiset, D. (2020). Implicit race attitudes modulate visual information extraction for trustworthiness judgments. *PloS One, 15*(9), e0239305.

Charlesworth, T. E. S., Hudson, S. T. J., Cogsdill, E. J., Spelke, E. S., & Banaji, M. R. (2019). Children use targets' facial appearance to guide and predict social behavior. *Developmental Psychology, 55*(7), 1400–1413.

Cho, S., Dydynski, J. M., & Kang, C. (2022). Universality and specificity of the kindchenschema: A cross-cultural study on cute rectangles. *Psychology of Aesthetics, Creativity, and the Arts, 16*(4), 719-732.

Chua, K. W., & Freeman, J. B. (2021). Facial stereotype bias is mitigated by training. *Social Psychological and Personality Science, 12*(7), 1335-1344.

Chua, K. W., & Freeman, J. B. (2022). Learning to judge a book by its cover: Rapid acquisition of facial stereotypes. *Journal of Experimental Social Psychology, 98*, e104225.

Chwe, J. A. H., Vartiainen, H. I., & Freeman, J. B. (2024). A multidimensional neural representation of face impressions. *Journal of Neuroscience, 44*(39).

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181-204.

Cogsdill, E. J., Todorov, A. T., Spelke, E. S., & Banaji, M. R. (2014). Inferring character from faces: A developmental study. *Psychological Science, 25*(5), 1132-1139.

Collova, J. R., Sutherland, C. A., & Rhodes, G. (2019). Testing the functional basis of first impressions: Dimensions for children's faces are not the same as for adults' faces. *Journal of Personality and Social Psychology, 117*(5), 900-924.

Conway, J. R., Catmur, C., & Bird, G. (2019). Understanding individual differences in theory of mind via representation of minds, not mental states. *Psychonomic Bulletin & Review, 26*(3), 798-812.

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

Cook, R., & Over, H. (2020). A learning model can explain both shared and idiosyncratic first impressions from faces. *Proceedings of the National Academy of Sciences of the USA, 117*(28), 16112-16113.

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

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 83*(6), 1314-1329.

Cowen, A. S., Keltner, D., Schroff, F., Jou, B., Adam, H., & Prasad, G. (2020). Sixteen facial expressions occur in similar contexts worldwide. *Nature, 589*(7841), 251-257.

Croley, J. A., Reese, V., & Wagner, R. F. (2017). Dermatologic features of classic movie villains: The face of evil. *JAMA Dermatology, 153*(6), 559-564.

Dawkins, R., & Krebs, J. R. (1979). Arms races between and within species. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 205*(1161), 489-511.

De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences, 22*(9), 764-779.

DeBruine, L. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society of London B: Biological Sciences, 269*(1498), 1307-1312.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences, 15*(6), 254-262.

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

Eggleston, A., Flavell, J. C., Tipper, S. P., Cook, R., & Over, H. (2021). Culturally learned first impressions occur rapidly and automatically and emerge early in development. *Developmental Science, 24*(2), e13021.

Eggleston, A., Geangu, E., Tipper, S. P., Cook, R., & Over, H. (2021). Young children learn first impressions of faces through social referencing. *Scientific Reports, 11*(1), e14744.

Eggleston, A., Tsantani, M., Over, H., & Cook, R. (2022). Preferential looking studies of trustworthiness detection confound structural and expressive cues to facial trustworthiness. *Scientific Reports, 12*(1), e17709.

Eisenegger, C., Haushofer, J., & Fehr, E. (2011). The role of testosterone in social interaction. *Trends in Cognitive Sciences, 15*(6), 263-271.

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience, 19*(9), 1508-1519.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44*(3), 227-240.

Ewing, L., Sutherland, C. A., & Willis, M. L. (2019). Children show adult-like facial appearance biases when trusting others. *Developmental Psychology, 55*(8), 1694–1701.

Falvello, V. B., Vinson, M., Ferrari, C., & Todorov, A. (2015). The robustness of learning about the trustworthiness of other people. *Social Cognition, 33*, 368-386.

FeldmanHall, O., Dunsmoor, J. E., Tompary, A., Hunter, L. E., Todorov, A., & Phelps, E. A. (2018). Stimulus generalization as a mechanism for learning to trust. *Proceedings of the National Academy of Sciences of the United States of America, 115*(7), E1690-E1697.

Fiala, V., Tureček, P., Akoko, R. M., Pokorný, Š., & Kleisner, K. (2022). Africans and Europeans differ in their facial perception of dominance and sex-typicality: a multidimensional Bayesian approach. *Scientific Reports, 12*, e6821.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77-83.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878-902.

Fleischmann, A., Lammers, J., Stoker, J. I., & Garretsen, H. (2019). You can leave your glasses on: Glasses can increase electoral success. *Social Psychology, 50*(1), 38–52.

Foo, Y. Z., Sutherland, C. A. M., Burton, N. S., Nakagawa, S., & Rhodes, G. (2022). Accuracy in facial trustworthiness impressions: Kernel of truth or modern physiognomy? A meta-analysis. *Personality and Social Psychology Bulletin, 48*(11), 1580-1596.

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.

Gawronski, B., & Quinn, K. A. (2013). Guilty by mere similarity: Assimilative effects of facial resemblance on automatic evaluation. *Journal of Experimental Social Psychology, 49*(1), 120-125.

Geniole, S. N., Denson, T. F., Dixson, B. J., Carré, J. M., & McCormick, C. M. (2015). Evidence from meta-analyses of the facial width-to-height ratio as an evolved cue of threat. *PloS One, 10*(7), e0132726.

Glocker, M. L., Langleben, D. D., Ruparel, K., Loughead, J. W., Gur, R. C., & Sachser, N. (2009). Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults. *Ethology, 115*(3), 257-263.

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 352*(1358), 1121-1127.

Haselhuhn, M. P., Ormiston, M. E., & Wong, E. M. (2015). Men's facial width-to-height ratio predicts aggression: A meta-analysis. *PloS One, 10*(4), e0122637.

Hassabis, D., Spreng, R. N., Rusu, A. A., Robbins, C. A., Mar, R. A., & Schacter, D. L. (2013). Imagine all the people: how the brain creates and uses personality models to predict behavior. *Cerebral Cortex, 24*(8), 1979-1987.

Hehman, E., Sutherland, C. A., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology, 113*(4), 513-529.

Hester, N., Xie, S. Y., & Hehman, E. (2021). Little Between-Region and Between-Country Variance When People Form Impressions of Others. *Psychological Science, 32*(12), 1907-1917.

Heyes, C. M. (2018). *Cognitive gadgets: The cultural evolution of thinking*: Harvard University Press.

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science, 344*(6190), e1243091.

Hong, S., Suk, H. W., Choi, Y., & Na, J. (2021). Face-based judgments: accuracy, validity, and a potential underlying mechanism. *Psychological Science, 32*(9), 1452-1462.

Jack, R. E., Sun, W., Delis, I., Garrod, O. G., & Schyns, P. G. (2016). Four not six: Revealing culturally common facial expressions of emotion. *Journal of Experimental Psychology: General, 145*, 708-730.

Jaeger, B., Oud, B., Williams, T., Krumhuber, E. G., Fehr, E., & Engelmann, J. B. (2022). Can people detect the trustworthiness of strangers based on their

facial appearance? *Evolution and Human Behavior. https://doi.org/10.1016/j.evolhumbehav.2022.04.004*.

Jaeger, B., Todorov, A. T., Evans, A. M., & van Beest, I. (2020). Can we reduce facial biases? Persistent effects of facial trustworthiness on sentencing decisions. *Journal of Experimental Social Psychology, 90*, e104004.

Jessen, S., & Grossmann, T. (2016). Neural and behavioral evidence for infants' sensitivity to the trustworthiness of faces. *Journal of Cognitive Neuroscience, 28*, 1728-1736.

Jessen, S., & Grossmann, T. (2020). The developmental origins of subliminal face processing. *Neuroscience & Biobehavioral Reviews, 116*, 454-460.

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138): Guilford Press.

Jones, B., DeBruine, L., Flake, J., Aczel, B., Adamkovic, M., Alaei, R., . . . Vásquez-Amézquita, M. (2021). To which world regions does the valence-dominance model of social perception apply? *Nature Human Behaviour, 5*(1), 159-169.

Kosinski, M. (2017). Facial width-to-height ratio does not predict self-reported behavioral tendencies. *Psychological Science, 28*(11), 1675-1682.

Kramer, R. S., & Gardner, E. M. (2020). Facial trustworthiness and criminal sentencing: A comment on Wilson and Rule (2015). *Psychological Reports, 123*(5), 1854-1868.

Lavan, N., Mileva, M., Burton, A. M., Young, A. W., & McGettigan, C. (2021). Trait evaluations of faces and voices: Comparing within-and between-person variability. *Journal of Experimental Psychology: General, 150*(9), 1854–1869.

Lawson, C., Lenz, G. S., Baker, A., & Myers, M. (2010). Looking like a winner: Candidate appearance and electoral success in new democracies. *World Politics, 62*(4), 561-593.

Lee, R., Flavell, J. C., Tipper, S. P., Cook, R., & Over, H. (2021). Spontaneous first impressions emerge from brief training. *Scientific Reports, 11*(1), e15024.

Li, Q., Heyman, G. D., Mei, J., & Lee, K. (2019). Judging a book by its cover: Children's facial trustworthiness as judged by strangers predicts their real-world trustworthiness and peer relationships. *Child Development, 90*(2), 562-575.

Lorenz, K. (1943). Die angeborenen formen möglicher erfahrung. *Zeitschrift für Tierpsychologie, 5*(2), 235-409.

Lorenz, K. (1971). *Studies in animal and human behaviour: II* (R. Martin, Trans.). Cambridge, MA: Harvard University Press.

Main, J. C., Jones, B. C., DeBruine, L. M., & Little, A. C. (2009). Integrating gaze direction and sexual dimorphism of face shape when perceiving the dominance of others. *Perception, 38*(9), 1275-1283.

Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *Behavioral and Brain Sciences, 21*(3), 353-363.

Miesler, L., Leder, H., & Herrmann, A. (2011). Isn't it cute: An evolutionary perspective of baby-schema effects in visual product designs. *International Journal of Design, 5*(3).

Mileva, M., Tompkinson, J., Watt, D., & Burton, A. M. (2018). Audiovisual integration in social evaluation. *Journal of Experimental Psychology: Human Perception and Performance, 44*(1), 128-138.

Mileva, M., Young, A. W., Kramer, R. S., & Burton, A. M. (2019). Understanding facial impressions between and within identities. *Cognition, 190*, 184-198.

Mondloch, C. J., Gerada, A., Proietti, V., & Nelson, N. L. (2019). The influence of subtle facial expressions on children's first impressions of trustworthiness and dominance is not adult-like. *Journal of Experimental Child Psychology, 180*, 19-38.

Montague, D. P., & Walker-Andrews, A. S. (2001). Peekaboo: a new look at infants' perception of emotion expressions. *Developmental Psychology, 37*(6), 826-838.

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior, 27*(4), 237-254.

Murphy, J., Ipser, A., Gaigg, S. B., & Cook, R. (2015). Exemplar variance supports robust learning of facial identity. *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 577-581.

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2020). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General, 149*(2), 323-342.

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences, 18*, 566-570.

Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology, 46*(2), 315-324.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the USA, 105*, 11087-11092.

Over, H., & Cook, R. (2018). Where do spontaneous first impressions of faces come from? *Cognition, 170*, 190-200.

Over, H., Eggleston, A., & Cook, R. (2020). Ritual and the origins of first impressions. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 375*(1805), e20190435.

Over, H., Lee, R., Flavell, J., Vestner, T., & Cook, R. (2023). Contextual modulation of appearance-trait learning. *Cognition, 230*, e105288.

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

Pandeirada, J. N., Madeira, M., Fernandes, N. L., Marinho, P., & Vasconcelos, M. ((2022). Judgements of social dominance from faces and related variables. *Frontiers in Psychology, 13*, e873147.

Peck, C. A., & Bouton, M. E. (1990). Context and performance in aversive-to-appetitive and appetitive-to-aversive transfer. *Learning and Motivation, 21*, 1-31.

Penton-Voak, I. S., & Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution and Human Behavior, 25*(4), 229-241.

Peterson, J. C., Uddenberg, S., Griffiths, T. L., Todorov, A., & Suchow, J. W. (2022). Deep models of superficial face judgments. *Proceedings of the National Academy of Sciences of the United States of America, 119*(17), e2115228119.

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law, 16*(6), 477-491.

Puts, D. A., Jones, B. C., & DeBruine, L. M. (2012). Sexual selection on human faces and voices. *Journal of Sex Research, 49*(2-3), 227-243.

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PloS One, 7*, e34293.

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: a meta-analytic and theoretical review. *Psychological Bulletin, 138*(1), 146-174.

Rohrbeck, P., Kersting, A., & Suslow, T. (2023). Trait anger and negative interpretation bias in neutral face perception. *Frontiers in Psychology, 14*, e1086784.

Rydell, R. J., & Gawronski, B. (2009). I like you, I like you not: Understanding the formation of context-dependent automatic attitudes. *Cognition and Emotion, 23*(6), 1118-1152.

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion, 9*(2), 260-264.

Sakuta, Y., Kanazawa, S., & Yamaguchi, M. K. (2018). Infants prefer a trustworthy person: An early sign of social cognition in infants. *PloS One, 13*(9), e0203541.

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.

Schaller, M. (2008). Evolutionary bases of first impressions. In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 15-34): Guilford Press.

Schiller, D., Freeman, J. B., Mitchell, J. P., Uleman, J. S., & Phelps, E. A. (2009). A neural mechanism of first impressions. *Nature Neuroscience, 12*(4), 508-514.

Scott, I. M., Clark, A. P., Josephson, S. C., Boyette, A. H., Cuthill, I. C., Fried, R. L., . . . Penton-Voak, I. S. (2014). Human preferences for sexually dimorphic faces may be evolutionarily novel. Proceedings of the National Academy of Sciences. *Proceedings of the National Academy of Sciences of the United States of America, 111*(40), 14388-14393.

Siddique, S., Sutherland, C. A., Palermo, R., Foo, Y. Z., Swe, D. C., & Jeffery, L. (2022). Development of face-based trustworthiness impressions in childhood: A systematic review and metaanalysis. *Cognitive Development, 61*, e101131.

Sofer, C., Dotsch, R., Oikawa, M., Oikawa, H., Wigboldus, D. H., & Todorov, A. (2017). For your local eyes only: Culture-specific face typicality influences perceptions of trustworthiness. *Perception, 46*(8), 914-928.

Song, R., Over, H., & Carpenter, M. (2016). Young children discriminate genuine from fake smiles and expect people displaying genuine smiles to be more prosocial. *Evolution and Human Behavior, 37*(6), 490-501.

Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist, 59*(5), 325-338.

Sternberg, R. J., & Grigorenko, E. L. (2004). Intelligence and culture: How culture shapes what intelligence means, and the implications for a science of well–being. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359*(1449), 1427-1434.

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science, 21*(3), 349-354.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Sussman, A. B., Petkova, K., & Todorov, A. (2013). Competence ratings in US predict presidential election outcomes in Bulgaria. *Journal of Experimental Social Psychology, 49*(4), 771-775.

Sutherland, C. A., Burton, N. S., Wilmer, J. B., Blokland, G. A. M., Germine, L., Palermo, R., . . . Rhodes, G. (2020). Individual differences in trust evaluations

are shaped mostly by environments, not genes. *Proceedings of the National Academy of Sciences of the United States of America, 117*(19), 10218-10224.

Sutherland, C. A., Collova, J. R., Palermo, R., Germine, L., Rhodes, G., Blokland, G. A., . . . Wilmer, J. B. (2020). Reply to Cook and Over: Social learning and evolutionary mechanisms are not mutually exclusive. *Proceedings of the National Academy of Sciences of the USA, 117*(28), 16114-16115.

Sutherland, C. A., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Personality and Social Psychology Bulletin, 44*, 521– 537.

Sutherland, C. A., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition, 127*, 105-118.

Sutherland, C. A., & Young, A. W. (2022). Understanding trait impressions from faces. *British Journal of Psychology, 113*(4), 1056-1078.

Swaddle, J. P., & Reierson, G. W. (2002). Testosterone increases perceived dominance but not attractiveness in human males. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 169*(1507), 2285-2289.

Tae, J., Nam, Y. E., Lee, Y., Weldon, R. B., & Sohn, M. H. (2020). Neutral but not in the middle: cross-cultural comparisons of negative bias of "neutral" emotional stimuli. *Cognition and Emotion, 34*(6), 1171-1182.

Talamas, S. N., Mavor, K. I., Axelsson, J., Sundelin, T., & Perrett, D. I. (2016). Eyelid-openness and mouth curvature influence perceived intelligence beyond attractiveness. *Journal of Experimental Psychology: General, 145*(5), 603-620.

Thierry, S. M., & Mondloch, C. J. (2021). First impressions of child faces: Facial trustworthiness influences adults' interpretations of children's behavior in ambiguous situations. *Journal of Experimental Child Psychology, 208*, e105153.

Thomas, M. (2002). Development of the concept of "the poverty of the stimulus". *The Linguistic Review, 19*(1-2), 51-71.

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences, 1124*(1), 208-224.

Todorov, A. (2017). *Face value: The irresistible influence of first impressions*: Princeton University Press.

Todorov, A., Funk, F., & Olivola, C. Y. (2015). Response to Bonnefon et al.: Limited 'kernels of truth'in facial inferences. *Trends in Cognitive Sciences, 19*(8), 422-423.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623-1626.

Todorov, A., Olivola, C., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*, 519-545.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*, 813-833.

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

Torrance, J. S., Wincenciak, J., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2014). The relative contributions of facial shape and surface information to perceptions of attractiveness and dominance. *PloS One, 9*(10), e104415.

Tsantani, M., Over, H., & Cook, R. (2023). Does a lack of perceptual expertise prevent participants from forming reliable first impressions of “other-race” faces? *Journal of Experimental Psychology: General, 152*(4), 1134-1145.

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.

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *Quarterly Journal of Experimental Psychology, 69*(10), 1996-2019.

Van Vugt, M., & Grabo, A. E. (2015). The many faces of leadership: an evolutionary psychology approach. *Current Directions in Psychological Science, 24*, 484-489.

Van’t Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108*(3), 796-803.

Vernon, R. J., Sutherland, C. A., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences of the USA, 111*(32), E3353-E3361.

Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science, 21*(6), 779-785.

Walker, M., Jiang, F., Vetter, T., & Sczesny, S. (2011). Universals and cultural differences in forming personality trait judgments from faces. *Social Psychological and Personality Science, 2*, 609-617.

Webster, M. A., & MacLeod, D. I. (2011). *Visual adaptation and face perception. Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1571), 1702-1725.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592-598.

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science, 26*(8), 1325–1331.

Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly, 59*(2), 189-202.

Witkower, Z., Hill, A. K., Koster, J., & Tracy, J. L. (2022). Is a downwards head tilt a cross-cultural signal of dominance? Evidence for a universal visual illusion. *Scientific Reports, 12*, e365.

Witkower, Z., & Tracy, J. L. (2019). A facial-action imposter: How head tilt influences perceptions of dominance from a neutral face. *Psychological Science, 30*(6), 893-906.

Xie, S. Y., Flake, J. K., & Hehman, E. (2019). Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of Personality and Social Psychology, 117*(2), 364-385.

Xie, S. Y., Flake, J. K., Stolier, R. M., Freeman, J. B., & Hehman, E. (2021). Facial impressions are predicted by the structure of group stereotypes. *Psychological Science, 32*(12), 1979-1993.

Yang, K. S., & Bond, M. H. (1990). Exploring implicit personality theories with indigenous or imported constructs: The Chinese case. *Journal of Personality and Social Psychology, 58*(6), 1087-1095.

Zebrowitz, L. A. (2004). The origins of first impressions. *Journal of Cultural and Evolutionary Psychology, 2*, 93-108.

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science, 26*, 237-242.

Zebrowitz, L. A., Fellous, J. M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review, 7*(3), 194-215.

Zebrowitz, L. A., & Montepare, J. M. (2006). The ecological approach to person perception: Evolutionary roots and contemporary offshoots. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and Social Psychology*: Psychology Press.

Zebrowitz, L. A., Montepare, J. M., & Lee, H. K. (1993). They don't all look alike: Individual impressions of other racial groups. *Journal of Personality and Social Psychology, 65*(1), 85-101.

Zebrowitz, L. A., Wang, R., Bronstad, P. M., Eisenberg, D., Undurraga, E., Reyes-García, V., & Godoy, R. (2012). First impressions from faces among US and

culturally isolated Tsimane'people in the Bolivian rainforest. *Journal of Cross-Cultural Psychology, 43*(1), 119-134.

Zebrowitz, L. A., & Zhang, Y. (2011). Origins of impression formation in animal and infant face perception. In D. J. Cacioppo (Ed.), *The Handbook of Social Neuroscience* (pp. 434-444). Oxford: Oxford University Press.

Zebrowitz McArthur, L., & Berry, D. S. (1987). Cross-cultural agreement in perceptions of babyfaced adults. *Journal of Cross-Cultural Psychology, 18*(2), 165-192.

Figures

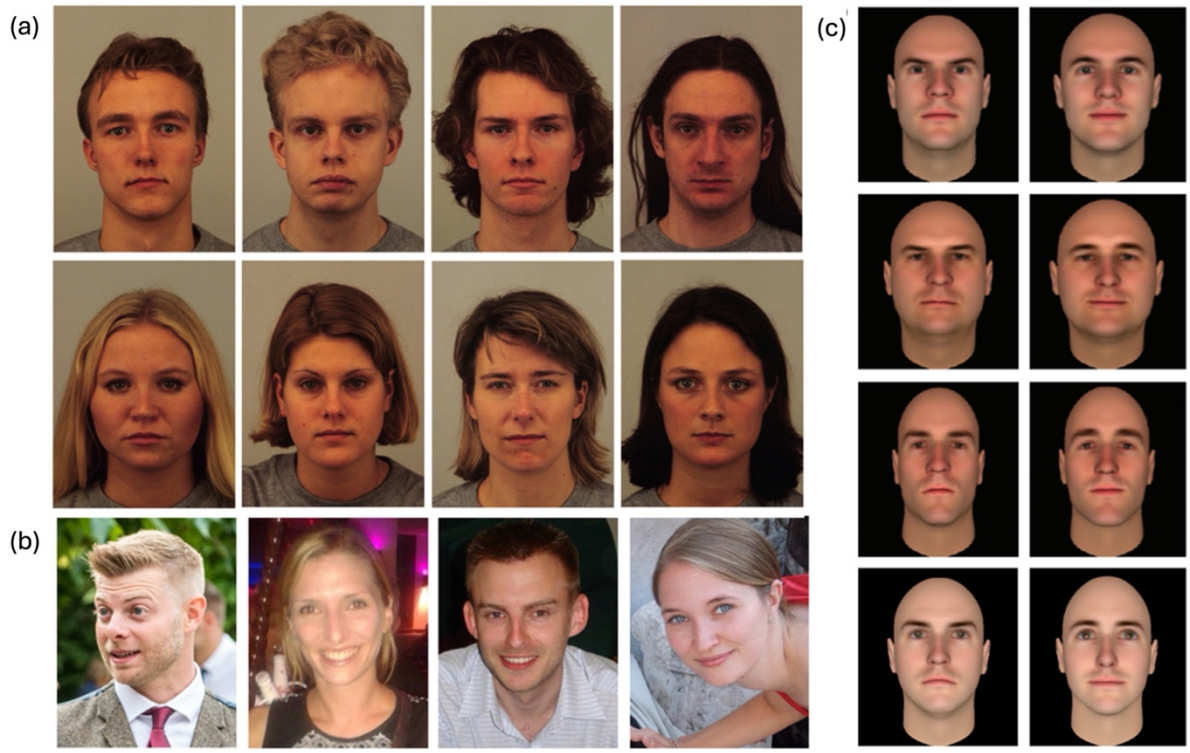

Figure 1. The types of stimuli used in first impressions research. a) Stimuli taken from the set of Karolinska Directed Emotional Faces b) Naturalistic or 'ambient' images supplied by the authors with the permission of the individuals depicted. c) Synthetic facial stimuli generated by the computer model described by Oosterhof & Todorov (2008).

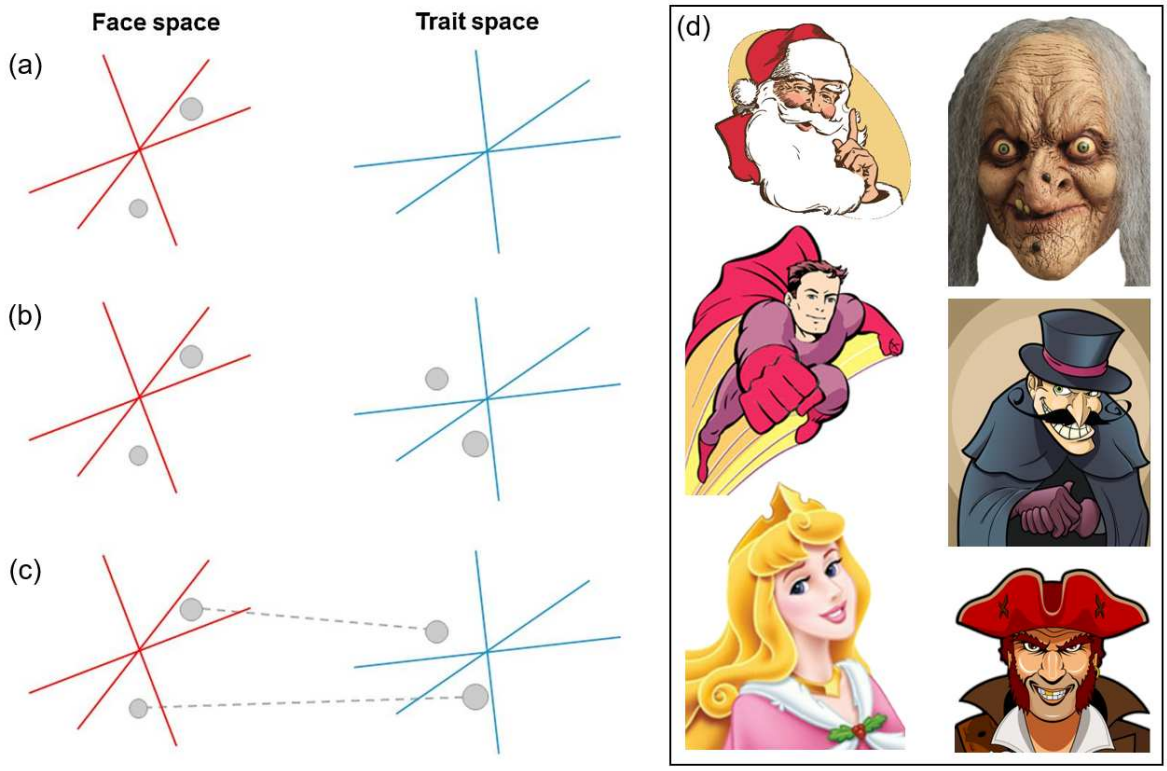

Figure 2. The TIM framework. (a) When encountering a stranger, an observer represents their face as a vector in face space. (b) Having learned about the stranger's behaviours, the observer can reason about their traits, thereby placing them in trait space. (c) Where excitation of representations in face space predicts the nature of representations in trait space, a face-trait mapping emerges. (d) TIM explains consensus impressions by arguing that the depiction of heroes and villains in films, storybooks, and rituals may promote consistent mappings between facial appearance and character.

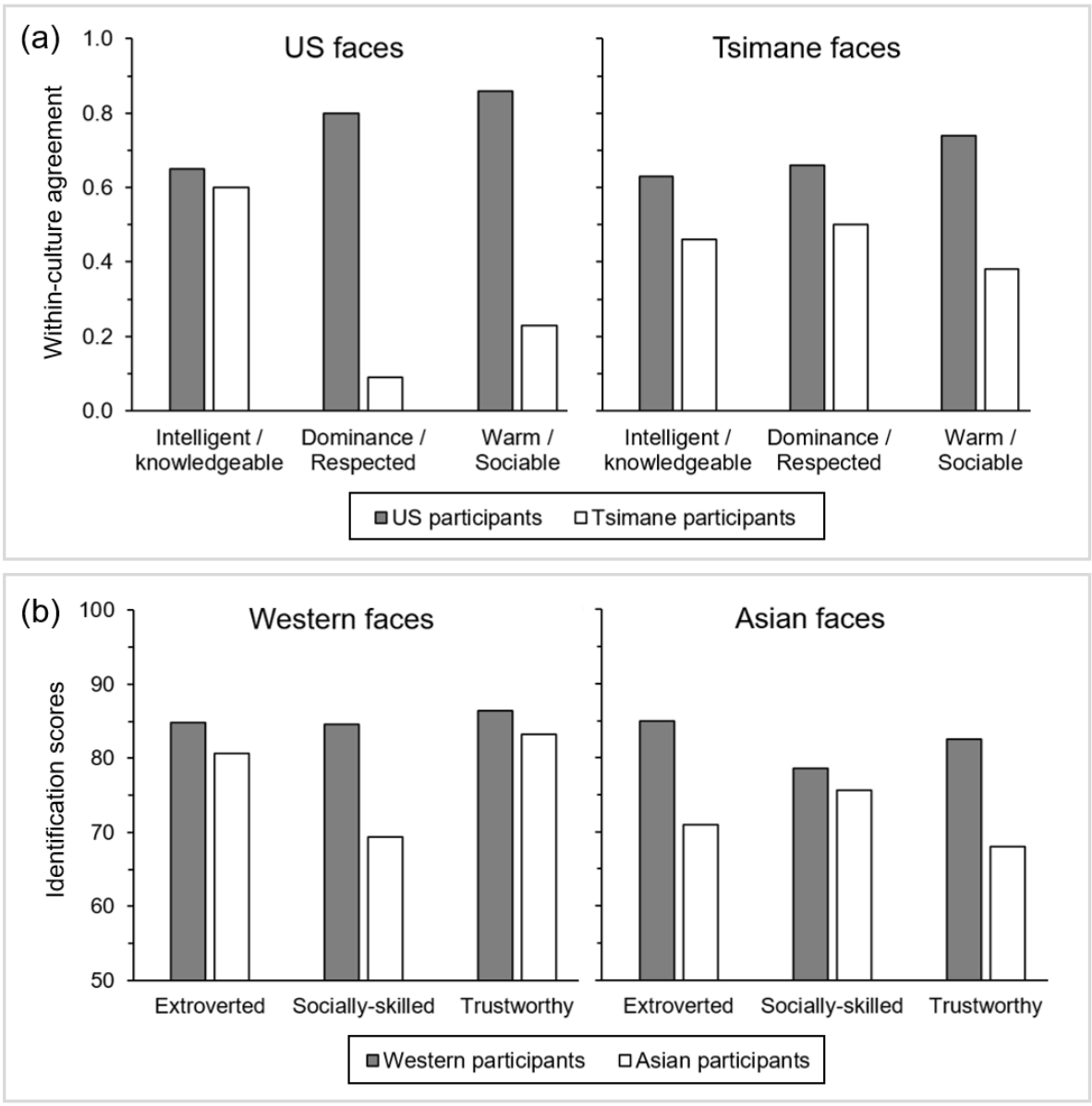

Figure 3. (a) Zebrowitz et al (2012) asked US students and Tsimane participants to judge US and Tsimane faces for intelligence / knowledgeability, dominance / respect and warmth / sociability. In each case, the US raters exhibited greater within-group agreement. When judging the respect and sociability of the US faces, the within group agreement of the Tsimane did not exceed chance. (b) Walker et al (2011) presented Western and Asian participants with pairs of faces that had been manipulated to afford certain trait judgement in Western observers. Although the Asian observers were also able to recognise the high and low variants for each trait, in several cases, they exhibited significantly lower identification scores (notably, for extroversion, social-skill, and trustworthiness).

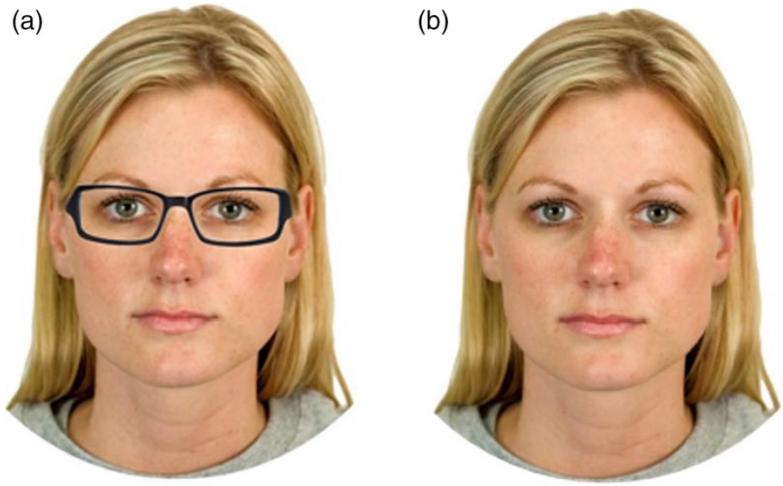

Figure 4. First impressions of intelligence are influenced by whether a person is wearing glasses. These inferences occur quickly and automatically and are present in Western children as young as 6. Thus, speed, automaticity and early emergence are not uniquely compatible with a nativist account. Images adapted from the Chicago Face Database (Ma et al., 2015).

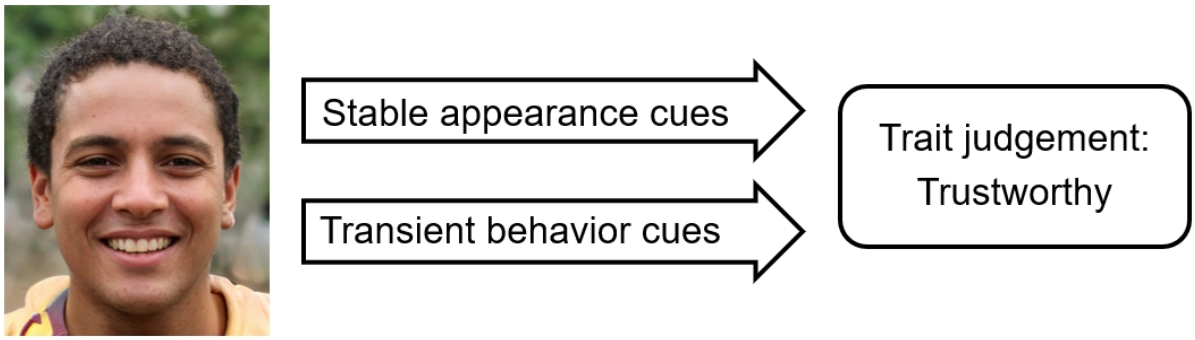

Figure 5. Facial photographs can be thought of as a compound stimulus, simultaneously depicting stable appearance cues – what does the actor look like? – and transient behaviour cues – what is the actor doing? The cultural learning account seeks to explain the origins of first impressions based on stable appearance cues. Image generated by AI at thispersondoesnotexist.org.

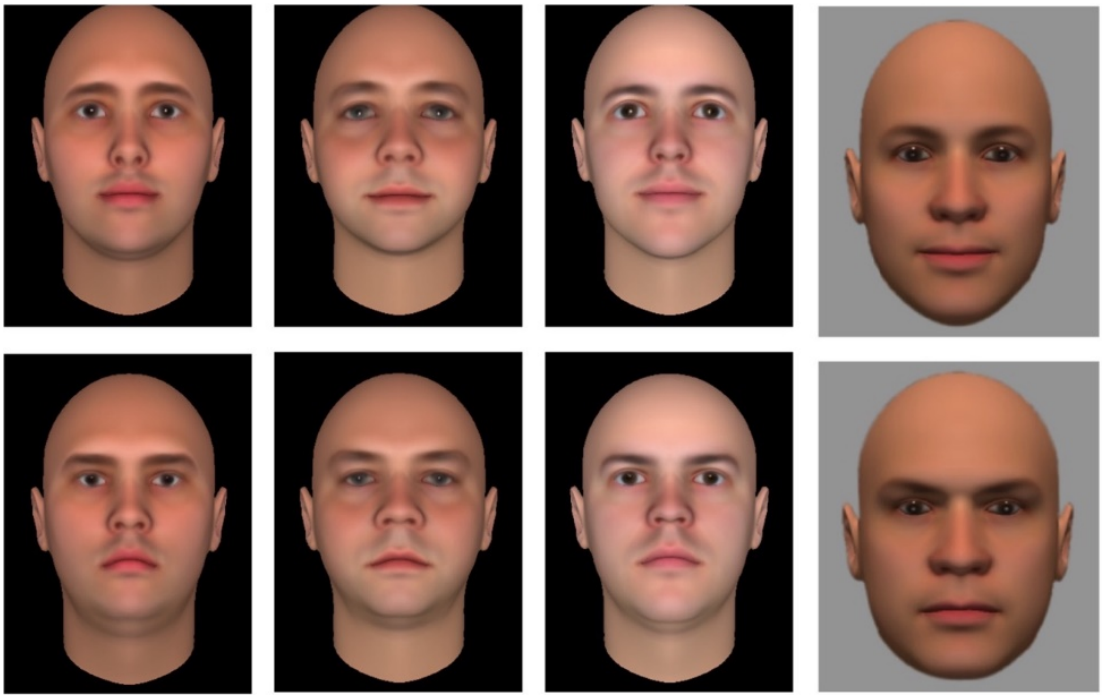

Figure 6. The stimuli used by Jessen and Grossmann (2016, left) and Sakura et al. (2018, right panels) confound structural and expression cues to apparent trustworthiness. Whereas, the trustworthy faces appear to be subtly smiling, the untrustworthy faces do not (Eggleston et al. 2022).

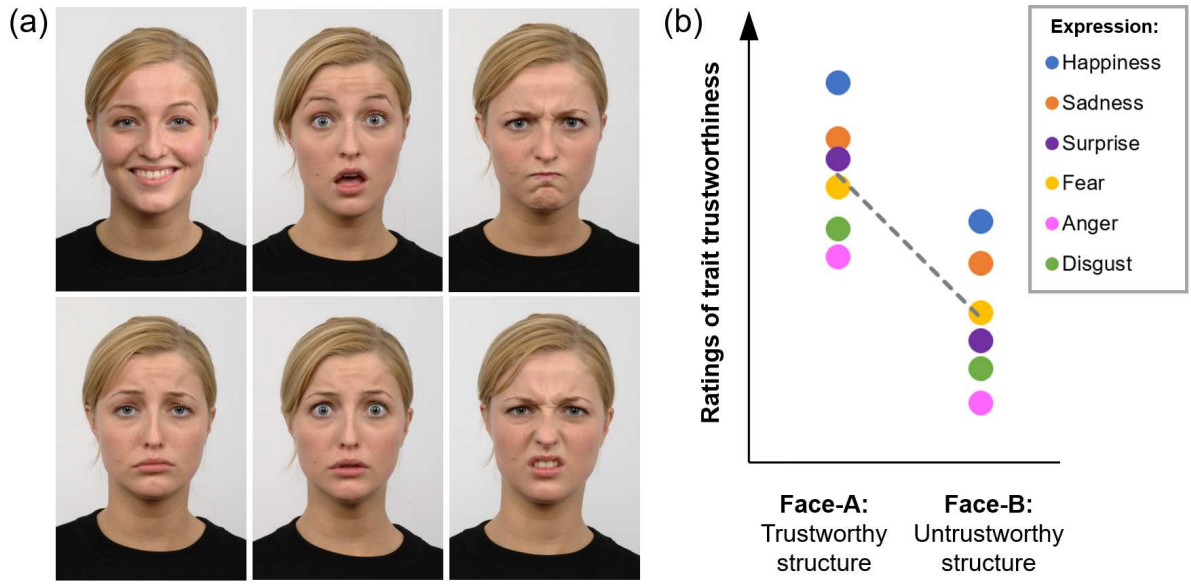

Figure 7. Using multiple exemplars of to-be-judged faces to elicit trait judgments of facial structure. (a) Each to-be-judged face could be shown expressing a range of standard expressions (e.g., happiness, sadness, surprise, fear, anger, & disgust). (b) Where observers estimate the traits for each exemplar, the influence of facial structure (dashed line) and facial expression on trait attributions can be modelled separately for each target face.