Article:

Cook, Richard, Eggleston, Adam and Over, Harriet orcid.org/0000-0001-9461-043X (2022) The cultural learning account of first impressions. Trends in Cognitive Sciences. pp. 656-668. ISSN 1364-6613

https://doi.org/10.1016/j.tics.2022.05.007

In press at *Trends in Cognitive Sciences*

The cultural learning account of first impressions

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

[1]Department of Psychological Sciences,
Birkbeck, University of London, London, U.K.

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

*Correspondence:
richard.cook@bbk.ac.uk
Department of Psychological Sciences,
Birkbeck, University of London,
Malet Street,
London, U.K., WC1E 7HX

Highlights

- When we encounter a stranger, we spontaneously attribute to them a wide variety of character traits (e.g., trustworthiness, dominance, intelligence) based on their facial appearance.

- The Trait Inference Mapping (TIM) account argues that first impressions are the product of domain general associative mappings between representations of facial appearance and representations of the possible trait profiles that others may possess. Many of these mappings are thought to arise through exposure to cultural messages.

- We review the emerging body of evidence that speaks to the origins of first impressions. Many recent findings, though not all, accord with the TIM account.

- We suggest that the distinction between inferences based on invariant facial properties (e.g., shape) and facial behaviours (e.g., expression) may be crucial to understanding these discrepant findings.

Abstract

We spontaneously attribute character traits to strangers based on their facial appearance. Although these 'first impressions' typically have no basis in reality, some authors have assumed they have an innate origin. In contrast, the Trait Inference Mapping (TIM) account proposes that first impressions are products of culturally acquired associative mappings that allow activation to spread from representations of facial appearance to representations of trait profiles. According to TIM, cultural instruments, including propaganda, illustrated storybooks, art and iconography, ritual, film and TV, expose many individuals within a community to common sources of correlated face-trait experience, yielding first impressions that are shared by many, but typically innacuate. We review emerging empirical findings, many of which accord with TIM, and argue that future work must distinguish first impressions based on invariant facial features (e.g., shape) from those based on facial behaviours (e.g., expressions).

Keywords:

Trait Inference Mapping; First impressions from faces; Face-trait inferences; Cultural learning

Where do first impressions from faces come from?

When we first encounter a stranger, we spontaneously attribute to them a wide variety of character traits based on their facial appearance [1, 2]. For example, physically attractive faces are judged to be more sociable and intelligent [3], while individuals with round faces and large eyes are judged to be naïve [4]. Males with wide faces are judged to be less trustworthy [5] and more aggressive [6, 7] than males with thin faces. Similarly, individuals shown smiling are judged to be warmer and more approachable than those who appear to scowl [8, 9]. These judgements are widely referred to as 'first impressions' from faces.

Our first impressions typically have little or no basis in reality – many of the traits we attribute to strangers show little correlation with their actual characteristics and behaviours [10, 11]. Nevertheless, these spontaneous trait inferences exert a strong influence on our behaviour. For example, first impressions based on facial appearance affect financial decisions [12], criminal sentencing [13], and the outcome of elections [14, 15].

For a long time, there was a dearth of interest in the origins of first impressions from faces. Historically, first impressions – in particular, so-called consensus impressions that are shared by many individuals – were often assumed to have an innate origin. They were attributed to an evolutionary adaptation for identifying trustworthy collaborators and good leaders [16-20]. However, a detailed evolutionary account has not been forthcoming. For example, the nature of the putative adaptation for trait inferences (i.e., the innate knowledge or innate mechanism) has not been clearly specified. Furthermore, it is unclear how or why knowledge that affords inaccurate trait judgements (i.e., that seemingly conveys no advantage on the observer) might become encoded genetically [21].

More recently, there has been a flurry of interest in the origins question. One catalyst behind this renewed attention was the publication of the Trait Inference Mapping (TIM) model in 2018 [22]. This account argues that first impressions from faces are the result of associative mappings, acquired within the lifetime of the individual, often through exposure to cultural instruments (e.g., propaganda, illustrated storybooks, art and iconography, ritual, film and TV,). These mappings allow activation to spread between representations of facial appearance and representations of the trait profiles that others may possess.

We begin our Opinion by outlining the key features of the TIM account. Next, we review recent findings that speak to the origins of first impressions. Although many findings accord with TIM, we suggest that an over-reliance on White and Western participants and White face stimuli may have obscured further evidence for the cultural learning account. Finally, we argue that findings which appear to contradict TIM are the product of different research groups implicitly endorsing different operational definitions of ‘first impressions’. To further advance our understanding of the origins of first impressions from faces, we suggest that future work must distinguish trait inferences based on facial appearance (i.e., invariant aspects of the face, such as shape, feature configuration, skin tone and texture) from those based on facial behaviours (e.g., expressions, gaze cues, head tilting).

Trait Inference Mapping

Contrary to the prevailing view at the time, TIM [21-24] proposed that automatic first impressions are the result of associative mappings acquired by individuals during the course of their lifetime that connect perceptual descriptions of facial appearance (points or regions in face space [Box 1]) with representations of the potential trait profiles that others may possess (locations in trait space [25-28]). These mappings are acquired via domain-general associative processes following exposure to correlated face-trait experience; i.e., learning episodes where certain facial features or feature configurations are predictive of particular trait profiles. Where the face of a stranger falls close to a mapped location in the observers’ face space, excitation is thought to propagate automatically from that perceptual description to the associated trait profile.

One source of correlated face-trait experience is our first-hand interactions with others. For example, where we encounter a teacher who is generous and kind, a mapping may form between the location of the teacher’s face in face space and the location in trait space that represents their particular trait profile. This kind of experience is thought to yield idiosyncratic mappings; i.e., face-trait associations that differ between people. However, this kind of learning cannot explain inaccurate consensus impressions (i.e., impressions that are widely shared within a culture, but that have little or no validity). It is unclear how most individuals within a community could learn the same erroneous face-trait mappings through first-hand interactions with others. Until quite recently, the prevailing view was therefore that learning plays a relatively limited role in the emergence of consensus impressions [16-20].

TIM resolves the apparent paradox whereby first impressions are inaccurate but widely shared by appealing to a second source of correlated face-trait experience - exposure to cultural instruments, including propaganda, illustrated storybooks, art and iconography, ritual, film and TV (Figure 1). For example, exposure to anti-Semitic propaganda may foster mappings that link faces with a pallid complexion and large noses (stereotypically associated with Jewish appearance) with negative traits such as greed and Machiavellianism [29]. Because these cultural devices expose many individuals within a community to a common source of correlated face-trait experience, they have the potential to yield inaccurate consensus impressions.

The development of certain face-trait mappings may be canalised by innate stimulus-response (S-R) behaviours. For example, 'infant schema' appears to elicit positive feelings and encourage nurturing behaviours [30, 31]. Similarly, some facial disfigurements may elicit aversion responses [32]. To some degree, these instinctive S-R behaviours may be conserved across species [30, 32]. These instinctive reactions may encourage the emergence of particular face-trait mappings (e.g., between infantile faces and trustworthy character). However, TIM argues that conserved S-R behaviours do not constitute trait inferences *per se*. Consistent with this view, instinctive S-R behaviours are also seen in animals that are not thought to attribute character traits to others [31, 33, 34].

TIM hypothesises a distinct 'face space' and 'trait space'. The dimensionality of both representation spaces is thought to be determined by the experience of the observer. As such, face space and trait space are likely to develop and change over time and may differ between individuals [27, 35]. The clear segregation of face space and trait space allows the same trait representation (e.g., trustworthy) to be excited independently by different types of sensory input. Thus, while the TIM framework was developed to explain first impressions from faces, the same architecture may be applied to understand other types of first impression, including those based on body shape [34] and vocal cues [36, 37].

The emerging evidence base

Next, we review evidence that speaks to the origin of first impressions from faces. We discuss evidence from several sources namely, studies of development, lab-based training, individual and cultural differences, and the speed and automaticity of impressions.

Developmental trajectory

In the context of the origins question, there has been much interest in the developmental trajectory of first impressions. Evidence that first impressions manifest in early infancy would be difficult to reconcile with a learning account, as young infants have little exposure to correlated face-trait experience. Conversely, gradual development across childhood would accord well with the cultural learning view.

To date, most developmental studies have focussed on attributions of trustworthiness [e.g., 38, 39-42]. A recent systematic review and meta-analysis found that consensus impressions of trustworthiness emerge around 3-5 years of age, and that trust impressions continue to develop throughout childhood, showing adult-like patterns between 10 and 13 years of age [43]. Attributions of competence and dominance appear to follow similar developmental trajectories [38].

It has been argued that the emergence of trait judgements at around 3-5 years is early enough to preclude a social learning account of their origin [38, 39]. Contrary to this view, however, the attribution of intelligence to those who wear glasses [44] also emerges at this point in development [45]. Importantly, glasses have been in existence for less than 800 years [46]. As such, this trait inference cannot possibly be a genetic adaptation; instead, it must be learned, either through exposure to cultural messages or via first-hand experience.

Seemingly in contradiction with TIM, it has been reported that infants as young as 7-months old attend to faces deemed trustworthy by U.S. adults in preference to faces deemed neutral and untrustworthy [40]. A follow-up study also found that 6-8 month-olds preferentially attended to trustworthy faces relative to untrustworthy faces, but only when faces were high in dominance – there was no effect of trustworthiness when faces were low in dominance [42]. However, it remains unclear whether these effects are products of sensitivity to facial trustworthiness *per se*, or cues to emotional expression. The facial stimuli used in these studies confound high and low trustworthiness with subtle expressions of happiness and anger respectively (Figure 2).

Training studies

There is now considerable evidence that people readily acquire new person knowledge during lab-based training procedures [47-49]. For example, training procedures might pair unfamiliar faces with positive (e.g., “Gave his balloon to a child who had let hers go”) or negative (e.g., “Stole money and jewellery from the relatives he was living with”) behaviours. At test, faces that have previously been paired with positive behaviours are judged to be more trustworthy than those paired with negative behaviours [47-49].

These judgements are not typically thought of as “first impressions” as they are decisions made about familiarised others informed by relevant evidence. Rather than “judging a book by its cover” the resulting judgements are akin to “judging a book by its first page”. Crucially, however, newly acquired face-trait associations generalise to novel faces of broadly similar appearance [50-53]. This suggests that our first impressions of strangers are likely influenced by our knowledge of familiar others and their traits.

More broadly, these findings confirm that the knowledge and mechanisms that underpin our first impressions of faces show some degree of plasticity – they can be modified through correlated face-trait experience. In particular, we can readily acquire new face-trait mappings through lab-based training. It is less clear, however, whether periods of lab-based training can "unteach" deeply engrained face-trait mappings. Recent attempts to reduce the effects of first impressions through training interventions have yielded mixed results [54, 55]. The study of renewal phenomena in the classical and evaluative conditioning literatures suggests a potential explanation. Specifically, new learning that contradicts old learning often manifests only in the context in which the new learning occurs [e.g., 56].

Individual and cultural differences

High-levels of inter-rater consensus within cultures combined with evidence of cultural universality would accord with the view that all humans are born with innate face-trait knowledge. Through its emphasis on cultural learning, TIM can explain the emergence of consensus impressions within cultures. However, it also predicts substantial and systematic individual differences as well as cross-cultural variation.

It is beyond doubt that some face-trait judgements exhibit high levels of inter-rater agreement within particular cultures, and are shared by observers in multiple cultures [4, 8, 19, 57, 58]. For example, people around the world infer naivety from babyfacedness [4] and warmth or approachability from smile cues [8, 19]. Similarly, the trait judgements seen in many cultures exhibit a broadly similar factor structure [59, 60]. We consider the meaning of these results below in the section "What are we trying to explain?"

Nevertheless, evidence of wide-spread individual differences continues to emerge. First impressions appear to vary as a function of observers' personality [61, 62], sex [61, 63], ethnicity [64-66], age [67-69], own appearance [70], and lay beliefs about how personality traits correlate [27, 28]. Recent findings from behavioural genetics have confirmed that idiosyncratic impressions of trustworthiness – those that differ across individuals – are mostly the products of the developmental environment [71].

Importantly, as predicted by TIM, we have also seen mounting evidence of cross-cultural differences in first impressions [19-21, 58, 60, 72-74], but see [75]. For example, Zebrowitz et al. [20] compared the first impressions of undergraduate observers from the U.S. with those of adults from the isolated Tsimane people in Bolivia. When rating White faces for dominance / respect and warmth / sociability, the US undergraduates exhibited extremely high inter-rater agreement. When judging the same faces, however, the ratings provided by the Tsimane people showed little or no consensus and inter-rater agreement failed to reach statistical significance [20].

It is likely that evidence of systematic individual and cultural differences in first impressions has been artificially restricted by the prevailing methodology in this field. Traditionally, first impressions research has been conducted in WEIRD cultures [21] and much of this work has focussed on the trait judgements made about White faces [24]. Some authors have also prevented people of colour from participating in studies as raters [76-78]. Unsurprisingly, however, there is growing evidence that racial stereotypes affect trait ratings [64-66]. Consequently, the focus on White face stimuli and White and WEIRD participants has likely inflated the apparent levels of inter-rater agreement [21, 24]. For detailed discussion of these issues, including suggestions for future research, see Cook and Over [24].

Speed and automaticity

First impressions can be based on a fleeting glimpse of a face. For example, observers form consistent impressions when to-be-judged faces are presented for 100 ms or less [57]. Similarly, first impressions of faces are thought to be automatic (hard to inhibit) [36, 79]. For example, impressions of faces influence the evaluation of voices even when participants have been asked to ignore them [36]. Some authors argue that the speed and automaticity of first impressions preclude an ontogenetic origin [16]. However, both of these features are compatible with a learning account. Reading is a cognitive process that is known to be learned, yet we do this quickly and automatically [80]. Consistent with this view, it has recently been shown that the attribution of intelligence to those who wear glasses [44] – an inference that must be learned – is also made when stimuli are presented for 100 ms and is hard to ignore [45].

Reinforcement and transmission

TIM suggests that face-trait mappings are acquired through exposure to correlated face-trait experience [21-24]. Importantly, however, TIM does not characterise individuals as passive recipients of first impressions from their environment. Rather, TIM suggests that individuals play an active role in the reinforcement of their first impressions and in the dissemination of innacurate face-trait knowledge to others [22, 81, 82].

Caregivers may intentionally or unintentionally scaffold the face-trait learning of children [22]. For example, when discussing images of strangers shown within a picture-book with their children, adults spontaneously reference the likely traits of the people depicted [81]. Through such discussions, caregivers may pass on their face-trait mappings to the next generation.

Similarly, in many cultures, children are encouraged to participate in rituals that foster particular face-trait mappings [21]. Examples of such rituals in the U.S. include Halloween (during which children learn about the appearance of witches and monsters) and beauty pageants (which reinforce the what-is-beautiful-is-good stereotype). These kinds of rituals not only teach children inaccurate face-trait knowledge, but also teach them a means to pass on that knowledge to future generations [21].

Recent findings also confirm that children can learn about the trustworthiness of faces through social referencing [82]. For example, target faces that elicit negative reactions from peers, quickly acquire negative valence and are judged to be untrustworthy by third-party observers. Once again, this kind of learning generalises to novel individuals who resemble the target faces encountered during the social referencing phase [82]. This accords with the suggestion that the reactions of caregivers and friends to strangers may inadvertently teach children the underlying face-trait mappings [22].

Once acquired, face-trait mappings may affect how we evaluate the behaviours of others. Should individuals exhibit signs of confirmation bias – for example, in their perception and recall of the social world – face-trait mappings may be self-reinforcing [22]. Consistent with this possibility, it has recently been shown that adults' impressions of children's facial trustworthiness influence their interpretation of ambiguous situations [83]. For example, children with trustworthy appearance were given the benefit of the doubt (e.g., that a negative outcome was accidental rather than deliberate) more often than those with untrustworthy appearance [83].

What are we trying to explain?

In light of the findings reviewed above, few authors – even those who hypothesise genetic adaptations for first impressions [84] – would deny that cultural learning likely plays a key role in the emergence of face-trait mappings. Nevertheless, evidence that some first impressions are seen in multiple cultures [4, 19, 58] and may emerge early in development [40, 42] leaves open the possibility of innate face-trait mappings. Where do we go from here? To further advance the literature on the origins of first impressions, we must consider carefully what kinds of phenomena we are trying to explain.

Impressions from appearance and behaviour

TIM [21-24] seeks to explain why we form spontaneous impressions about the likely traits of others based on their facial appearance (i.e., invariant aspects of the face including shape, feature configuration, skin tone and texture [85-87]). Trait attributions based solely on invariant features can be thought of as hypotheses about other people formed in the absence of relevant behavioural evidence and likened to "judging a book by its cover". Importantly, TIM does not seek to explain trait attributions from behaviour (e.g., the inference that someone shouting while wielding a gun is untrustworthy). This is a different kind of judgement. Trait attributions informed by relevant behavioural evidence can be likened to "judging a book by its content" (albeit just a few pages).

This distinction is particularly important when it comes to the treatment of trait inferences from facial expressions [Box 2]. Frequently, the facial stimuli used in first impressions research confound invariant features and expressive cues [8, 19, 40, 41, 88, 89]. For example, synthetic faces may be rendered more trustworthy by making them slimmer (a shape manipulation) or through the addition of a subtle smile (an expressive manipulation). Conversely, faces may be rendered less trustworthy by making them wider (a shape manipulation) or through the addition of a subtle scowl (an expressive manipulation) (see Figure 2).

Young infants show some basic understanding of expression valence [90, 91]. Thus, evidence that infants of this age also prefer to look at trustworthy faces that show subtle signs of positive affect [40, 42] comes as no great surprise. Similarly, people from different cultures around the world produce broadly similar emotional expressions and infer similar meaning from these displays [92, 93]. It is therefore unsurprising that people from these cultures also judge smiling faces to be nicer and more approachable than faces that appear to scowl [19, 58].

Whether these findings evidence an evolutionary adaption for the inference of character traits *per se* is far from clear. Adaptations that affect the production [94, 95] and recognition of expressions [96] potentially explain why trait inferences from expression cues emerge early in development [40, 42] and manifest cross-culturally [19, 58]. One does not need to posit innate mechanisms that have evolved specifically for the inference of character traits to explain these findings.

Crucially, evidence that one type of trait inference (from expression cues) manifests cross-culturally and early in development should not be used to argue that the other type of inference (from invariant features) has an innate origin. Because the existing evidence base confounds these two sources of variation so comprehensively, many important questions about the development and consistency of first impressions from invariant features remain entirely unresolved (see Outstanding questions).

Invariant features that resemble expression cues

Some people may have invariant facial features (e.g., narrow eyes; a mouth that naturally curves upwards at the corners) that cause observers to perceive expressive behaviour where none is intended [89]. People whose resting face shape resembles a scowl may be judged unfairly because interactants misattribute to them undesirable behaviours (scowling). Should these trait inferences be regarded as appearance- or behaviour-based?

If it were possible to monitor the muscles of a to-be-judged face or carefully examine how the face changes over time and in different situations, it would be possible to establish whether the person is scowling or whether they have an unusual facial shape. However, when viewing a photographic image of a stranger's face, study participants cannot establish the ground truth empirically. Instead, observers must 'guess' – or rather, their visual system must infer – the person's likely face shape and expression from the available perceptual evidence.

Importantly, when confronted with an image that depicts a person with an unusual face shape expressing no emotion, observers may well perceive a person with a statistically more likely face shape expressing emotion. For this reason, we believe the trait inferences in these ambiguous cases are likely based on expression even when the source of the facial variation is actually structural. When addressing questions of mechanism and origin, it makes little difference whether traits are inferred from veridical or mis-perceived expression cues. In both cases, the means by which participants infer traits is likely to be the same; in both cases, judgements are based on *perceived* expressive behaviours.

Other behavioural cues

To advance our understanding of the origin of first impressions, we believe it is necessary to distinguish trait inferences based on facial appearance from those based on facial behaviours. Facial expression is one type of behaviour that exerts a strong influence on first impressions. However, there are others. For example, faces with direct gaze tend to be judged more dominant than faces with averted gaze [97]. Similarly, individuals who tilt their head towards the observer are judged more dominant than those who do not [98].

When seeking to understand trait inferences from facial appearance (i.e., from invariant features), these behaviours are potential confounds. Note that gaze direction and head tilt exert strong consensus effects on trait inferences, particularly on impressions of dominance [97, 98]. Stimulus sets in which gaze direction and head tilt appear to vary may therefore afford higher levels of inter-rater agreement than stimulus sets in which gaze direction and head tilt are held constant.

Studying trait inferences from facial appearance

To study trait inferences from invariant facial features, researchers must prevent participants from basing their judgements on facial behaviours (e.g., expressions, gaze direction, head tilt) or otherwise account for their influence. One approach is to ensure that all target faces are depicted with a so-called “neutral” expression (i.e., impassive), direct gaze, and no head-tilt. This is not straightforward. In particular, naïve participants perceive facial emotion in a great many stimulus images described as “neutral” by their creators and the authors who use them [99-101]. Where authors seek face stimuli with neutral expressions, it is important that rigorous stimulus screening is employed.

Alternatively, researchers could present participants with multiple images of target faces exhibiting a set of facial expressions (e.g., happiness, sadness, anger, surprise). Provided each target face is shown making the same expressions, the nature and valence of the expressions cannot be used to infer the character traits of the individual. Showing observers how the appearance of a target face varies across different expressions may also help them form an accurate representation of its invariant properties. Exemplar variation is thought to facilitate perceptual learning about particular facial identities [102, 103]. One possibility is that the visual system identifies commonalities across the exemplars via averaging [104].

It is unclear whether inferences from invariant facial cues can be studied using ambient images. Under this approach, there is no attempt to control the expressions, head tilt, and gaze direction of the to-be-judged faces. To control for their influence statistically, the full range of facial behaviours (e.g., expressions, gaze direction, head tilt) present in each image must be quantified accurately and objectively.

Concluding remarks

The TIM account provides a framework for understanding the origins and development of first impressions from faces. In particular, it posits a key role for cultural learning in the emergence of consensus impressions. This account may ultimately be falsified or proved incomplete. Nevertheless, it has helped to focus attention on the origins of first impressions. While first impressions typically have little or no validity, they exert a pervasive influence on our lives [10]. It is important that we understand to what extent we ‘teach’ our children about the appearance of heroes and villains, jocks and geeks, those who are competent and incompetent. Understanding the role of learning in the emergence of first impressions may eventually inform efforts to protect against their worst consequences, for example, by modifying the nature of the correlated face-trait experience our children receive.

Outstanding questions

When do trait attributions from face structure emerge in development? To date, there has been little attempt to distinguish trait judgements based on face structure from those based on expression cues. Given that infants show precocious understanding of facial expressions, it comes as no surprise that young children form more positive impressions of those who appear to smile than of those who appear to scowl. If trait inferences from face structure are products of correlated face-trait experience, they may emerge later in development than those based on expression cues.

Do people from different cultures exhibit similar trait attributions from face structure? Given that people around the world exhibit broadly similar facial expressions and infer similar meaning from these displays, it is largely unsurprising that people from different cultures derive similar trait inferences from expression cues (e.g., smiles, scowls). Trait inferences from face structure may well show greater cross-cultural variability.

To what extent is susceptibility to consensus impressions determined by our environment? Recent findings from behavioural genetics indicate that idiosyncratic impressions are the product of individuals' developmental environment. It is important to determine whether observers' susceptibility to consensus impressions is also determined by environmental factors, akin to the formation of racial stereotypes. The answer may differ for consensus impressions based on face structure and those based on expression cues.

How will greater diversity in terms of study participants and the facial stimuli employed affect the origins debate? To date, the vast majority of first impressions research has been conducted with White and WEIRD participants using stimulus images that depict White faces. It is important that we understand how efforts to increase diversity affect estimates of inter-rater consensus and our understanding of the developmental trajectory of first impressions.

Box 1: The face space at the front end of the TIM architecture

Within the human visual system, faces are thought to be encoded as mean-relative points or vectors within a multi-dimensional representation space [35, 105-108]. Each dimension within this face space is thought to encode a particular source of facial variation (e.g., inter-ocular distance). A face's position on a given dimension is thought to be determined by the relative excitation of opponent neural populations with inverse tuning profiles [e.g., 107]; for example, one population might respond maximally to small inter-ocular distances, while the other might respond maximally to large inter-ocular distances. The respective winner in their 'tug of war' and the margin of victory, determines the encoding of inter-ocular distance.

The precise dimensionality of an observer's face space (e.g., the number of dimensions and the attributes encoded by each) is likely to be determined by the kinds of faces that they encounter in their environment - their particular 'diet of faces' [35, 105-108]. For example, someone who has spent their entire life in rural China may have a dimensionality optimised to encode the variation present within East Asian faces. As a result, this individual may lack the dimensionality necessary to fully encode the variation present within sets of White or Black faces [e.g., 35].

Each representation within face space is a 'best-guess' made by the visual system about a target individual's facial appearance; i.e., the most likely solution given the retinal input and available contextual information. Estimates of face shape, feature configuration, skin tone and texture, will be informed by our previous experience of facial appearance and expression, and based on numerous assumptions about light source (e.g., lit from above or below), surface reflectance properties, camera parameters, and the depicted individual's likely pose and position (e.g., are we looking up at or looking down on the target? Are they tilting their head?).

TIM assumes that, as we become increasingly familiar with a person, our best-guess about their likely appearance (e.g., their face shape) becomes more accurate [102-104]. In the absence of any person-specific perceptual learning, however, the representation of strangers' faces may be particularly error-prone [104]. This feature of the model provides an elegant account of why different images of the same unfamiliar face (e.g., with different poses, different lighting conditions) sometimes elicit different trait attributions [37, 109]. Different poses and different lighting conditions may produce different estimates of facial appearance and thereby excite different trait profiles.

Box 2: Trait inferences from invariant facial features and facial expression

When asked to evaluate the traits of people depicted in static stimulus images, participants can base their judgements on invariant face cues; permanent or semi-permanent aspects of facial appearance (e.g., shape, feature configuration, skin tone and texture). These are the same cues that support judgements of facial identity [85-87]. First impressions based on invariant cues include the inference of trustworthiness and aggression from facial width-to-height ratio [5-7] and the inference of naivety from round face shape [4].

Where available, participants can also base trait judgements on expression cues. Findings obtained with ambient images indicate that smiling faces are judged more confident and approachable [8, 9, 19]. Similarly, faces that are supposedly neutral in terms of their emotional expression are judged to be more or less trustworthy when participants detect subtle traces of happiness or anger, respectively [89, 110]. Relative to judgements based on invariant properties, judgements based on expression cues are more likely to vary across different images of the same person [37, 109].

Despite some superficial similarities, trait inferences from invariant properties and expression cues are qualitatively different. The expression cues present in a facial photograph can be thought of as a 'thin slice' of behaviour [111]. That someone shown scowling is judged less trustworthy than someone shown smiling is conceptually similar to the inference that someone shouting while wielding a gun is less trustworthy than someone singing while holding a coffee mug. In both cases, the likely traits are inferred from the person's behaviour rather than from their appearance.

Trait inferences from invariant facial properties and facial expressions are likely to be mediated by different neurocognitive mechanisms [85-87]. For example, findings from neuroimaging suggest that regions of fusiform and superior temporal cortex may contribute disproportionately to the processing of invariant facial features and facial expression, respectively [112, 113]. The interpretation of facial expressions – but not invariant properties – may also benefit from covert simulation within the action production network [114].

People exist who have problems interpreting invariant facial properties (e.g., they have difficulties identifying and discriminating faces) but not facial expression, and vice versa [115]. People with a relatively selective deficit of expression processing may show atypical trait inferences from expression but typical trait inferences from face shape. Those with a selective deficit that affects the processing of invariant features may show typical trait inferences from expression but atypical trait inferences from face shape.

Figures

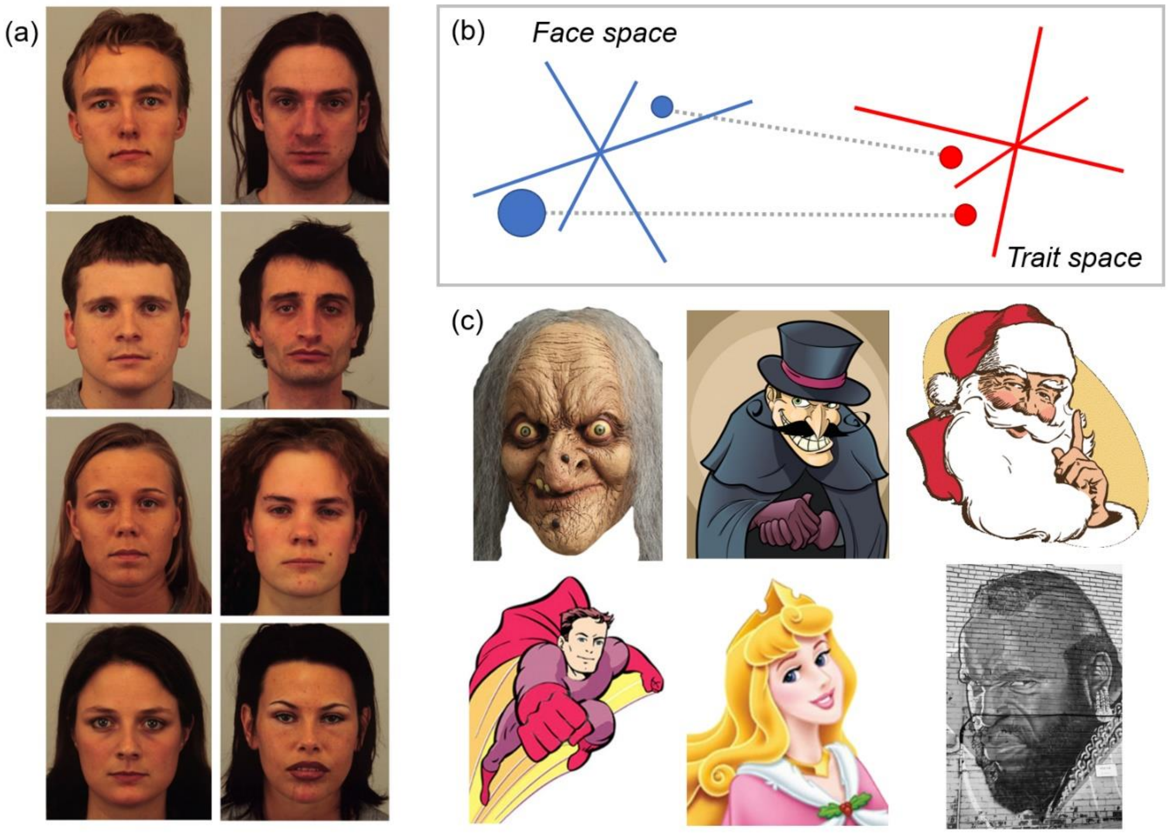

Figure 1. The Trait Inference Mapping (TIM) account of first impressions. (a) When we encounter strangers, we spontaneously attribute to them a wide variety of character traits based on their facial appearance. For example, young adults from the U.S. judged the faces on the left to be relatively trustworthy, while the faces on the right were judged to be relatively untrustworthy [57]. (b) According to TIM, first impressions of faces are products of associative mappings that allow excitation to propagate from representations of facial appearance (points in face space) to representations of the potential trait profiles that others may possess (points in trait space). These mappings are thought to be acquired ontogenetically through exposure to correlated face-trait experience. (c) The depictions of characters in illustrated storybooks, film and TV, art and iconography, and ritual, may help to canalise consensus impressions by exposing many individuals within a culture to shared sources of correlated face-trait experience.

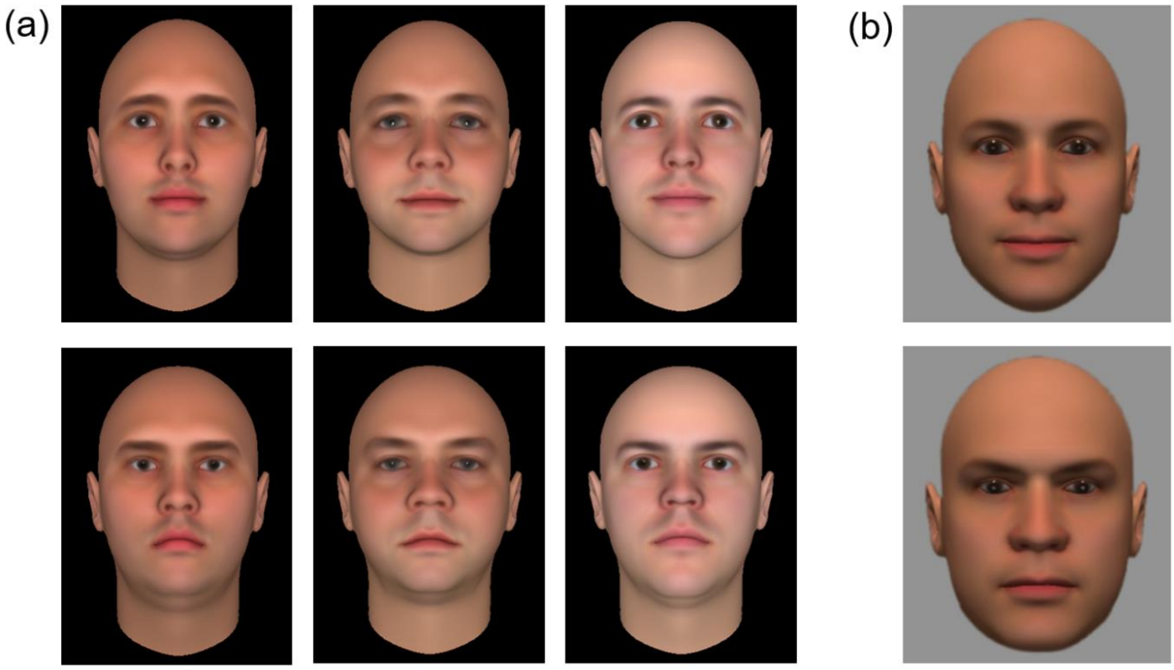

Figure 2. Stimulus images used in infant studies of first impressions. Trustworthy (top row) and untrustworthy (bottom row) facial stimuli employed by (a) Jessen and Grossmann [40] and (b) Sakuta and colleagues [42]. Although purportedly neutral, these stimuli are inherently ambiguous [89]. They can either be perceived as people with unusual face shapes expressing no emotion or people with more typical face shapes expressing subtle signs of happiness (top row) and anger (bottom row). Unsurprisingly, adults judge the trustworthy stimuli to be happier than the untrustworthy stimuli, while the untrustworthy stimuli are judged to be angrier than the trustworthy stimuli [100]. Thus, although infants (6-8 months old) attend preferentially to the trustworthy faces over the untrustworthy faces, this effect may simply reflect early sensitivity to facial affect.

Acknowledgments

This research was supported by awards from the European Research Council (ERC-STG-715824 to RC and ERC-STG-755719 to HO).

References

1. Todorov, A. et al. (2015) Social attributions from faces: Determinants, consequences, accuracy, and functional significance. Annual Review of Psychology 66, 519-545.

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

3. Eagly, A.H. et al. (1991) What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. Psychological Bulletin 110 (1), 109-128.

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

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

6. Geniole, S.N. et al. (2014) The facial width-to-height ratio shares stronger links with judgments of aggression than with judgments of trustworthiness. Journal of Experimental Psychology: Human Perception and Performance 40 (4), 1526-1541.

7. Summersby, S. et al. (2022) Tracking sexual dimorphism of facial width-to-height ratio across the lifespan: implications for perceived aggressiveness. Royal Society Open Science 9 (5), e211500.

8. Sutherland, C.A. et al. (2013) Social inferences from faces: Ambient images generate a three-dimensional model. Cognition 127, 105-118.

9. Vernon, R.J. et al. (2014) Modeling first impressions from highly variable facial images. Proceedings of the National Academy of Sciences of the USA 111 (32), E3353-E3361.

10. Olivola, C.Y. et al. (2014) Social attributions from faces bias human choices. Trends in Cognitive Sciences 18, 566-570.

11. Jaeger, B. et al. (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.

12. Duarte, J. et al. (2012) Trust and credit: The role of appearance in peer-to-peer lending. The Review of Financial Studies 25 (8), 2455-2484.

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

14. Ballew, C.C. and 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.

15. Todorov, A. et al. (2005) Inferences of competence from faces predict election outcomes. Science 308 (5728), 1623-1626.

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

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

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

19. Sutherland, C.A. et al. (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.

20. Zebrowitz, L.A. et al. (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.

21. Over, H. et al. (2020) Ritual and the origins of first impressions. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 375 (1805), e20190435.

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

23. Cook, R. and 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.

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

25. Stolier, R.M. et al. (2018) A dynamic structure of social trait space. Trends in Cognitive Sciences 22 (3), 197-200.

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

27. Stolier, R.M. et al. (2018) The conceptual structure of face impressions. Proceedings of the National Academy of Sciences of the United States of America 115 (37), 9210-9215.

28. Stolier, R.M. et al. (2020) Trait knowledge forms a common structure across social cognition. Nature Human Behaviour 4 (4), 361-371.

29. Keen, S. (1991) Faces of the enemy: Reflections of the hostile imagination, Harper San Francisco.

30. Glocker, M.L. et al. (2009) Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults. Ethology 115 (3), 257-263.

31. Lorenz, K. (1971) Studies in animal and human behaviour: II, Harvard University Press.

32. Ryan, S. et al. (2012) Facial disfigurement is treated like an infectious disease. Evolution and Human Behavior 3 (6), 639-646.

33. Shibasaki, M. and Kawai, N. (2009) Rapid detection of snakes by Japanese monkeys (Macaca fuscata): an evolutionarily predisposed visual system. Journal of Comparative Psychology 132 (2), 131-135.

34. Hu, Y. et al. (2018) First impressions of personality traits from body shapes. Psychological Science 29 (12), 1969-1983.

35. Furl, N. et al. (2002) Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis. Cognitive Science 26 (6), 797-815.

36. Mileva, M. et al. (2018) Audiovisual integration in social evaluation. Journal of Experimental Psychology: Human Perception and Performance 44 (1), 128-138.

37. Lavan, N. et al. (2021) Trait evaluations of faces and voices: Comparing within-and between-person variability. Journal of Experimental Psychology: General 150 (9), 1854–1869.

38. Cogsdill, E.J. et al. (2014) Inferring character from faces: A developmental study. Psychological Science 25 (5), 1132-1139.

39. Ewing, L. et al. (2019) Children show adult-like facial appearance biases when trusting others. Developmental Psychology 55 (8), 1694–1701.

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

41. Jessen, S. and Grossmann, T. (2019) Neural evidence for the subliminal processing of facial trustworthiness in infancy. Neuropsychologia 126, 46-53.

42. Sakuta, Y. et al. (2018) Infants prefer a trustworthy person: An early sign of social cognition in infants. PloS One 13 (9), e0203541.

43. Siddique, S. et al. (2022) Development of face-based trustworthiness impressions in childhood: A systematic review and metaanalysis. Cognitive Development 61, e101131.

44. Fleischmann, A. et al. (2019) You can leave your glasses on: Glasses can increase electoral success. Social Psychology 50 (1), 38–52.

45. Eggleston, A. et al. (2021) Culturally learned first impressions occur rapidly and automatically and emerge early in development. Developmental Science 24 (2), e13021.

46. Ilardi, V. (2007) Renaissance vision from spectacles to telescopes, American Philosophical Society.

47. Falvello, V.B. et al. (2015) The robustness of learning about the trustworthiness of other people. Social Cognition 33, 368-386.

48. Todorov, A. and Uleman, J.S. (2002) Spontaneous trait inferences are bound to actors' faces: evidence from a false recognition paradigm. Journal of Personality and Social Psychology 83 (5), 1051-1065.

49. Cone, J. et al. (2021) The long-term effects of new evidence on implicit impressions of other people. Psychological Science 32 (2), 173-188.

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

51. FeldmanHall, O. et al. (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.

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

53. Lee, R. et al. (2021) Spontaneous first impressions emerge from brief training. Scientific Reports 11 (1), e15024.

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

55. Jaeger, B. et al. (2020) Can we reduce facial biases? Persistent effects of facial trustworthiness on sentencing decisions. Journal of Experimental Social Psychology 90, e104004.

56. Rydell, R.J. and 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.

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

58. Walker, M. et al. (2011) Universals and cultural differences in forming personality trait judgments from faces. Social Psychological and Personality Science 2, 609-617.

59. Lin, C. et al. (2021) Four dimensions characterize attributions from faces using a representative set of English trait words. Nature Communications 12 (1), e5168.

60. Jones, B. et al. (2021) To which world regions does the valence-dominance model of social perception apply? Nature Human Behaviour 5 (1), 159-169.

61. Mattarozzi, K. et al. (2015) Effects of gender and personality on first impression. PloS One 10 (9), e0135529.

62. Meng, X. et al. (2022) Self-construal and generalized trust predict first impressions from faces. Personality and Individual Differences 191, e111537.

63. Mileva, M. et al. (2019) Social evaluation of faces across gender and familiarity. Perception 48 (6), 471-486.

64. Xie, S.Y. et al. (2021) Facial impressions are predicted by the structure of group stereotypes. Psychological Science 32 (12), 1979-1993.

65. Stanley, D.A. et al. (2011) Implicit race attitudes predict trustworthiness judgments and economic trust decisions. Proceedings of the National Academy of Sciences of the United States of America 108 (19), 7710-7715.

66. Trent, J. and Ferguson, Y. (2021) How ethnicity, expression, and study design influence first impressions of approachability. Psychological Reports 124 (2), 862-895.

67. Zebrowitz, L.A. et al. (2013) Older and younger adults' first impressions from faces: similar in agreement but different in positivity. Psychology and Aging 28 (1), 202-212.

68. Li, Y. et al. (2021) Age differences in facial trustworthiness judgement based on multiple facial cues. British Journal of Psychology 112 (2), 474-492.

69. Twele, A.C. and Mondloch, C.J. (2022) The dimensions underlying first impressions of older adult faces are similar, but not identical, for young and older adult perceivers. British Journal of Psychology. https://doi.org/10.1111/bjop.12568.

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

71. Sutherland, C.A. et al. (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.

72. Lakshmi, A. et al. (2021) The India Face Set: International and Cultural Boundaries Impact Face Impressions and Perceptions of Category Membership. Frontiers in Psychology 12, e627678.

73. Chen, F.F. et al. (2016) Culture matters: The looks of a leader are not all the same. Social Psychological and Personality Science 7 (6), 570-578.

74. Fiala, V. et al. (2022) Africans and Europeans differ in their facial perception of dominance and sex-typicality: a multidimensional Bayesian approach. Scientific Reports 12, e6821.

75. Hester, N. et al. (2021) Little Between-Region and Between-Country Variance When People Form Impressions of Others. Psychological Science 32 (12), 1907-1917.

76. Sutherland, C.A. et al. (2016) Integrating social and facial models of person perception: Converging and diverging dimensions. Cognition 157, 257-267.

77. Talamas, S.N. et al. (2016) Eyelid-openness and mouth curvature influence perceived intelligence beyond attractiveness. Journal of Experimental Psychology: General 145 (5), 603-620.

78. Collova, J.R. et al. (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.

79. Thierry, S.M. et al. (2021) Mandatory first impressions: Happy expressions increase trustworthiness ratings of subsequent neutral images. Perception 50 (2), 103-115.

80. MacLeod, C.M. (1991) Half a century of research on the Stroop effect: an integrative review. Psychological Bulletin 109 (2), 163-203.

81. Eggleston, A. et al. (2021) Parents reinforce the formation of first impressions in conversation with their children. Plos One 16 (8), e0256118.

82. Eggleston, A. et al. (2021) Young children learn first impressions of faces through social referencing. Scientific Reports 11 (1), e14744.

83. Thierry, S.M. and 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.

84. Sutherland, C.A. et al. (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.

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

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

87. Haxby, J.V. et al. (2000) The distributed human neural system for face perception. Trends in Cognitive Sciences 4, 223-233.

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

89. Said, C.P. et al. (2009) Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. Emotion 9 (2), 260-264.

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

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

92. Jack, R.E. et al. (2016) Four not six: Revealing culturally common facial expressions of emotion. Journal of Experimental Psychology: General 145, 708-730.

93. Cowen, A.S. et al. (2020) Sixteen facial expressions occur in similar contexts worldwide. Nature 589 (7841), 251-257.

94. Matsumoto, D. and Willingham, B. (2009) Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. Journal of Personality and Social Psychology 96 (1), 1-10.

95. Tracy, J.L. and Matsumoto, D. (2008) The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. Proceedings of the National Academy of Sciences of the United States of America 105 (33), 11655-11660.

96. Johnson, M.H. (2005) Subcortical face processing. Nature Reviews Neuroscience 6 (10), 766-774.

97. Main, J.C. et al. (2009) Integrating gaze direction and sexual dimorphism of face shape when perceiving the dominance of others. Perception 38 (9), 1275-1283.

98. Witkower, Z. et al. (2022) Is a downwards head tilt a cross-cultural signal of dominance? Evidence for a universal visual illusion. Scientific Reports 12, e365.

99. Tsantani, M. et al. (2022) How does the presence of a surgical face mask impair the perceived intensity of facial emotions? PLoS One 17 (1), e0262344.

100. Eggleston, A. et al. (2022) Why do infants preferentially attend to trustworthy faces? Exploring the role of facial affect. https://psyarxiv.com/57pc3.

101. Gray, K.L. et al. (2017) Modulation of the composite face effect by unintended emotion cues. Royal Society Open Science 4 (4), e160867.

102. Burton, A.M. et al. (2016) Identity from variation: Representations of faces derived from multiple instances. Cognitive Science 40 (1), 202-223.

103. Murphy, J. et al. (2015) Exemplar variance supports robust learning of facial identity. Journal of Experimental Psychology: Human Perception and Performance 41 (3), 577-581.

104. Burton, A.M. et al. (2005) Robust representations for face recognition: The power of averages. Cognitive Psychology 51 (3), 256-284.

105. 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.

106. Valentine, T. et al. (2016) Face-space: A unifying concept in face recognition research. The Quarterly Journal of Experimental Psychology 69 (10), 1996-2019.

107. O'Toole, A.J. et al. (2018) Face space representations in deep convolutional neural networks. Trends in Cognitive Sciences 22 (9), 794-809.

108. Leopold, D.A. et al. (2001) Prototype-referenced shape encoding revealed by high-level aftereffects. Nature Neuroscience 4 (1), 89-94.

109. Mileva, M. et al. (2019) Understanding facial impressions between and within identities. Cognition 190, 184-198.

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

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

112. Andrews, T.J. and Ewbank, M.P. (2004) Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. Neuroimage 23 (3), 905-913.

113. Winston, J.S. et al. (2004) fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. Journal of Neurophysiology 92 (3), 1830-1839.

114. Niedenthal, P.M. et al. (2010) The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. Behavioral and Brain Sciences 33 (6), 417-480.

115. Bate, S. and Bennetts, R. (2015) The independence of expression and identity in face-processing: evidence from neuropsychological case studies. Frontiers in Psychology 6, e770.