

Running head: *The origin of first impressions*

## Where do spontaneous first impressions of faces come from?

Harriet Over<sup>1\*</sup> and Richard Cook<sup>2</sup>

<sup>1</sup>Department of Psychology; University of York

<sup>2</sup>Department of Psychological Sciences; Birkbeck, University of London

**\*Corresponding author:**

[harriet.over@york.ac.uk](mailto:harriet.over@york.ac.uk)

Department of Psychology

University of York

York

YO10 5DD

Tel: +44 1904 322906

Fax: +44 1904 323181

**Acknowledgements:** We would like to thank Tim Andrews, Rebecca Brewer, Malinda Carpenter, Cecilia Heyes, Cade McCall, David Over, Steven Tipper, and Alex Todorov for valuable comments on this work. Harriet Over was funded by an ESRC grant (reference number: ES/K006702/1).

## **Abstract**

Humans spontaneously attribute a wide range of traits to strangers based solely on their facial features. These first impressions are known to exert striking effects on our choices and behaviours. In this paper, we provide a theoretical account of the origins of these spontaneous trait inferences. We describe a novel framework ('Trait Inference Mapping') in which trait inferences are products of mappings between locations in 'face space' and 'trait space'. These mappings are acquired during ontogeny and allow excitation of face representations to propagate automatically to associated trait representations. This conceptualization provides a framework within which the relative contribution of ontogenetic experience and genetic inheritance can be considered. Contrary to many existing ideas about the origins of trait inferences, we propose only a limited role for innate mechanisms and natural selection. Instead, our model explains inter-observer consistency by appealing to cultural learning and physiological responses that facilitate or 'canalise' particular face-trait mappings. Our TIM framework has both theoretical and substantive implications, and can be extended to trait inferences from non-facial cues to provide a unified account of first impressions.

## **1. Introduction**

Humans spontaneously attribute a wide variety of traits to strangers based solely on their facial appearance. These first impressions include judgements about trustworthiness, honesty, competence, intelligence, dominance, and likeability (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Todorov, 2017; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015; Zebrowitz & Montepare, 2008). These judgements are formed quickly and exert a striking influence over our behavior (Olivola, Funk, & Todorov, 2014). For example, spontaneous impressions from faces are thought to affect employment opportunities (Olivola, Eubanks, & Lovelace, 2014), voting patterns (Ballew & Todorov, 2007), and sentencing decisions (Funk & Todorov, 2013; Wilson & Rule, 2015). Previous research has done much to elucidate the cues on which these judgements are based and to describe their social consequences (Olivola, Funk et al., 2014; Todorov et al., 2015; Todorov, Said, Engell, & Oosterhof, 2008).

In the present paper, we consider where these spontaneous trait inferences come from and present a novel theoretical account of their origins: Trait Inference Mapping (TIM). We propose that spontaneous trait inferences can be understood as mappings between locations in ‘face space’ and locations in ‘trait space’. This conceptualization provides a framework within which the relative contribution of ontogenetic experience and genetic inheritance can be considered. First, we contend that face space and trait space – requisites for face-trait mappings – are themselves subject to considerable environmental control. In other words, our face perception ability and our knowledge of other people’s traits are both shaped by experience. We go on to assert that facial representations become associated with particular traits as a consequence of correlated face-trait experience. This learning may be heavily influenced by cultural factors including systematic differences in the ways in which individuals with different character traits are depicted in literature, film and other forms of media. Innate physiological responses play a limited and specific role in our model whereby they facilitate or ‘canalize’ (Waddington, 1942) the emergence of particular face-trait mappings. Before outlining our argument in detail, we briefly review the literature on spontaneous first impressions (for comprehensive reviews see: Olivola, Funk et al., 2014; Todorov, 2017; Todorov et al., 2015; Todorov et al., 2008). We then consider existing ideas about the origins of spontaneous trait judgements, before presenting our framework and discussing its theoretical and substantive implications.

## **2. Making trait inferences from faces**

When viewing photographs of unfamiliar faces, adults spontaneously attribute a range of characteristics to the person depicted. While a wealth of spontaneous judgements have been studied, many first impressions appear to load on two principal dimensions often described as ‘trustworthiness’ and ‘dominance’ (Oosterhof & Todorov, 2008; Sutherland et al., 2013). Adults make stable trait judgements when faces are viewed for only 100 ms (Willis & Todorov, 2006), and inter-rater agreement is above chance when faces are presented for only 33 ms (Todorov, Pakrashi, & Oosterhof, 2009). Developmental research has shown that by the age of 3, children are capable of making explicit judgements about how ‘nice’ and ‘strong’ a person appears to be (Cogsdill, Todorov, Spelke, & Banaji, 2014). Indeed, from 7 months of age, infants preferentially orient towards faces deemed trustworthy by adults rather than those deemed untrustworthy (Jessen & Grossmann, 2016). Interestingly, although some first impressions may be based on ‘a kernel of truth’ (Bonnenfon, Hopfensitz, & De Neys, 2015), others appear unrelated to the true behavioural tendencies of the people being judged (Oosterhof & Todorov, 2008; Todorov et al., 2008).

### Figure-1

Considerable research has sought to identify the facial cues on which observers base trait inferences. Permanent facial features which resemble subtle facial emotions may provoke inferences in line with those provoked by the corresponding emotional expression (Montepare & Dobish, 2003; Said, Sebe, & Todorov, 2009). For example, lower eyebrows, a common feature of facial displays of anger, may cause an individual with naturally low eyebrows to be perceived as dominant (Keating, Mazur, & Segall, 1981). Faces with infantile features are often judged to be trustworthy (Zebrowitz, 2004; Zebrowitz, Franklin, & Boshyan, 2015; Zebrowitz & Zhang, 2011). Attractive faces, associated with facial symmetry, averageness, and sexual dimorphism (Rhodes, 2006; Thornhill & Gangestad, 1999), elicit positive evaluation on a number of dimensions, including competence and trustworthiness (Dion, Berscheid, & Walster, 1972; Talamas, Mavor, & Perrett, 2016; Verhulst, Lodge, & Lavine, 2010; Wilson & Eckel, 2006). Masculine features increase judgements of dominance when judging male faces (Batres, Re, & Perrett, 2015; Swaddle & Reiersen, 2002). Facial adiposity (fatty tissue) affects perceived leadership ability (Re & Perrett, 2014). Eye-lid openness and mouth curvature influence perceived intelligence (Talamas, Mavor, Axelsson, Sundelin, & Perrett, 2016) and males with wider faces may be

perceived as more dominant and less trustworthy (Stirrat & Perrett, 2010; Valentine, Li, Penke, & Perrett, 2014). This list, while not exhaustive, illustrates the range of cues and attributions that have been studied.

### **3. Previous accounts of the origins of trait inferences**

To date, there has been a dearth of detailed discussion of the origins of trait inferences. Within this embryonic literature, theorising has tended to fall back on evolutionary explanations. Judgements of trustworthiness are thought to have emerged from a selection pressure to distinguish friends from foe (Oosterhof & Todorov, 2008; Zebrowitz, 2004; Zebrowitz & Zhang, 2011) and judgements of dominance are thought to have emerged from a need to distinguish potential leaders from followers (Van Vugt & Grabo, 2015). In both cases, however, it has been suggested that the ability to make trait judgements conferred an advantage on our ancestors; spontaneous impressions were important for survival and successful social interaction in our evolutionary past. As a result, the cognitive mechanisms for making trait judgements are often characterised as inherited products of gene-based natural selection. Such evolutionary explanations have overshadowed potential learning accounts. Authors have typically ascribed a limited role to learning, citing environmental factors as a potential source of idiosyncratic differences between observers' trait inferences (Todorov, 2017).

Claims of innateness have been justified in a number of different ways. Zebrowitz and Zhang (2011) argue that the speed and automaticity of trait judgements demands a nativist explanation. However, this logic does not withstand scrutiny. Speed and automaticity alone do not provide strong evidence for innateness. For example the classic Stroop Effect demonstrates that reading – a prototypical example of a learned skill – occurs quickly and automatically (Stroop, 1935). Other researchers have pointed to the fact that trait inferences appear early in development (Cogsdill et al., 2014; Jessen & Grossmann, 2016). However, manifestation early in development alone need not imply a strong innate basis either. Consider, for example, that newborn infants prefer to hear their native language over a foreign language, an effect that must be a product of the environment (Kinzler, Dupoux, & Spelke, 2007; Moon, Panneton-Cooper, & Fifer, 1993).

The specific architecture hypothesised by innate accounts has not been articulated and it is unclear on which mechanism or process natural selection is thought to have acted. For

example, models fail to state clearly which aspects of face perception and personality understanding are thought to be innately specified, or how visual inputs come to excite trait representations. A strong nativist account – implied in much of the discussion cited above – would hold that inferences of the type ‘this person appears trustworthy’ rely on a genetically inherited mechanism of some description. To maintain this position would appear to require that the two requisite skills underlying spontaneous first impressions from faces - perception of faces and understanding of others’ traits - must also have an innate basis. Although the rudiments of these abilities may well be innately specified, development continues throughout childhood and beyond. Any convincing framework for understanding spontaneous first impressions must incorporate these developmental trajectories.

#### **4. Trait inference mapping (TIM)**

We propose a Trait Inference Mapping (TIM) framework in which, rather than being viewed as judgments in a unitary face-trait space (e.g., Sutherland et al., 2013), trait inferences are understood as the products of mappings between locations in two distinct spaces: ‘face space’ and ‘trait space’. We argue that these mappings are acquired ontogenetically as a consequence of correlated face-trait experience. Mappings allow excitation to spread automatically from face representations to trait representations, and thereby give rise to spontaneous trait inferences. In the following subsections, we outline the components of our model in more detail, first explaining how face space and trait space are both heavily influenced by experience. Having outlined evidence for these two claims, we then describe how correlated face-trait experience may induce mappings between these spaces.

##### *4.1 Face space*

Face space comprises a multitude of dimensions that each describes a mode of facial variation. Together, these dimensions may be thought of as a multidimensional space within which the visual system represents the faces it encounters (Rhodes & Leopold, 2011; Valentine, 1991; Valentine & Endo, 1992; Webster & MacLeod, 2011). Attributes are thought to be represented through opponent coding, whereby loadings on a dimension are determined by the relative excitation of two neural populations with complementary receptive fields (e.g., Susilo, McKone, & Edwards, 2010). TIM posits a face space in which i) many opponent pools exhibit view invariant responses, and ii) transient (expression) and structural (face shape) sources of variation are coded by broadly separate dimensions (Calder, Burton,

Miller, Young, & Akamatsu, 2001; Calder & Young, 2005). Transient and structural dimensions may approximate subspaces that allow identity-invariant description of expression and expression-invariant description of facial shape, respectively.

While we recognise that there may be differences between the visual processing of familiar and unfamiliar faces (Burton & Jenkins, 2011; Hancock, Bruce, & Burton, 2000), TIM assumes that all faces, both familiar and unfamiliar, are encoded as vectors within a common face space. Representations of unfamiliar faces – particularly those depicted in facial photographs – may be imperfect, given the inherent ambiguity present when inferring 3D structure from a single 2D image (Todd, 2004). However, as observers acquire familiarity with a face, the associated vector may become more stable (Burton & Jenkins, 2011; Jenkins & Burton, 2011; Murphy, Ipser, Gaigg, & Cook, 2015).

Some aspects of face perception may be present from birth. For example, new born infants famously orient to face-like patterns (Johnson, 2005; Johnson, Dziurawiec, Ellis, & Morton, 1991). However, face space is thought to emerge ontogenetically (Rhodes & Leopold, 2011; Valentine, 1991; Valentine & Endo, 1992; Webster & MacLeod, 2011). This accords with the observation that structural development of the face processing network continues well into adulthood (Gomez et al., 2017), and with findings that observers' face recognition ability peaks in their early thirties (Germine, Duchaine, & Nakayama, 2011). A popular view, inspired by data reduction algorithms (e.g. principal components analysis), is that the visual system extracts important modes of variation from the population of faces encountered by the individual (Calder & Young, 2005). Where an individual encounters many faces of a particular type, the resulting dimensionality may be optimized to describe this variation (Furl, Phillips, & O'Toole, 2002). The lack of tailored dimensionality may explain culture-specific recognition abilities; for example, why many Western Caucasians experience difficulties recognising East Asian faces, and vice versa (Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005).

The visual description of a face, captured by its position in face space, represents the input into our model. Factors and manipulations that influence how that face is represented in face space will therefore directly determine which mappings are activated and consequently which trait space representations gain excitation. It has previously been reported that visual illusions that alter the appearance of faces – including facial aftereffects (Buckingham et al., 2006;

Engell, Todorov, & Haxby, 2010) and the composite face illusion (Todorov, Loehr, & Oosterhof, 2010) – also alter the traits that observers infer from those faces. These effects likely result from changing the input into the mechanism for deriving trait inferences, rather than modulating the mechanism itself. It has also been reported that the traits inferred from a particular face can vary depending on the distribution of recently encountered faces (Dotsch, Hassin, & Todorov, 2016). We speculate that this may be a conceptually similar example of input modulation.

#### *4.2 Trait space*

Our model assumes that conceptual knowledge of others' traits is accumulated throughout development. For the sake of exposition, we refer to this body of knowledge as a 'trait-space'. This description accords with the dimensional characterisation of personality and intelligence constructs (Digman, 1990; Mackintosh, 1998), and the use of a dimensional space to represent the trait variation of others (Conway, Catmur, & Bird, 2017; Fiske, Cuddy, & Glick, 2007; Oosterhof & Todorov, 2008; Sutherland et al., 2013; Todorov et al., 2008). We assume that trait space dissociates from observers' knowledge of, and ability to perceive, faces ('face space') but we make no claims about the neural instantiation of this knowledge space (e.g. whether traits are represented through opponent coding). In our framework, trait representations describe stable characteristics of an individual that predict their likely responses and behaviours across a range of situations. Crucially, traits are latent constructs that must be inferred (e.g. personality, intelligence, abilities and talents), not surface attributes (facial identity, gender, 'babyfacedness', physical attractiveness) available from a superficial analysis of visible features (also see discussion of appearance qualities; e.g., Zebrowitz et al., 2012).

The behaviour of others is frequently unpredictable, ambiguous, and context dependent. Identifying latent trait dimensions, and learning to represent others within this dimensionality, therefore poses a formidable challenge. In many cases, theory of mind (Premack & Woodruff, 1978) may be a necessary prerequisite. For example, understanding traits like 'honesty' requires that individuals first recognise the possibility of deception (Sodian, 1991). Even making simple attributions of 'nice' and 'not nice', often requires some understanding of an actor's intentions, as well as the actual outcome of their actions (Cushman, Sheketoff, Wharton, & Carey, 2013; Vaish, Carpenter, & Tomasello, 2010). The



acquisition of trait-related vocabulary is also likely to scaffold the development of some trait concepts, and vice versa (Bretherton & Beeghly, 1982). For example, learning subtle distinctions between traits such as ‘likeable’, ‘trustworthy’ and ‘approachable’ may benefit from linguistic constructs. As a result, mature trait concepts may emerge relatively late in childhood. Indeed, individuals may continue to develop and refine their unique trait space throughout their lifetime.

Nevertheless, rudiments of personality understanding may appear early in development (Baillargeon, Scott, & Bian, 2016; Hamlin, Wynn, & Bloom, 2007). For example, in relation to *trustworthiness*, Hamlin and colleagues (Hamlin et al., 2007) have shown that even in the first year of life, infants appear to prefer individuals who are helpful over individuals who cause harm. In relation to *dominance*, 10- to 13-month-old infants (but not 8-month-old infants) expect the larger of two agents to prevail in a conflict (Thomsen, Frankenhuus, Ingold-Smith, & Carey, 2011) (see also Mascaro & Csibra, 2014). In relation to *competence*, 14 month old infants prefer to learn from individuals who have been reliable in the past than from individuals who have previously been unreliable (Zmyj, Buttelman, Carpenter, & Daum, 2010).

We propose that the emergence of mature trait constructs may be subject to considerable environmental influence; that each individual’s trait space reflects their own particular developmental experience. For example, individuals raised in challenging social environments may develop complex dimensionality for representing the negative traits of others and may broadly expect strangers to be untrustworthy (Dodge, 1980). Exposure to prototypical characters in fairy tales and children’s stories (Lee et al., 2014), the extent to which mothers use trait terms in conversation (Taumoepeau & Ruffman, 2006), as well as training in particular social activities such as pretend play (Dockett, 1998) and deception (Heyes & Frith, 2014; Lohmann & Tomasello, 2003) may influence children’s developing abilities in this domain.

Evidence that trait space is at least partially under environmental control comes from cross-cultural research. For example, some rural groups value knowledge relating to plants, animals, weather, soils, and diseases, whereas the Western concept of intelligence is abstract and less relevant (Reyes-García et al., 2010; Zebrowitz et al., 2012). More generally, there

are cultural differences in the extent to which individuals use trait constructs, as opposed to inferences about the situation, to explain the behaviour of others. In interdependent cultures, traits do not necessarily form the basis of social judgements; rather, individuals within these cultures are more likely to focus on situational factors as potential determinants of an individual's actions (Norenzayan, Choi, & Nisbett, 2002).

#### *4.3 Two routes to trait inferences*

Brief inspection of a stranger's facial features and head shape allows the observer to represent the stranger's face shape in face space. But how do we represent strangers in trait space? We speculate that observers can represent individuals in trait space via two routes. Much of the existing literature assumes that an 'automatic route' is responsible for spontaneous trait inferences about unfamiliar faces (Todorov, Mandisodza, Goren, & Hall, 2005; Willis & Todorov, 2006). The elucidation of this automatic route is the focus of the present article. It is important to recognize, however, that trait inferences need not be automatic. In line with dual process accounts of the mind more generally (Chaiken & Trope, 1999; Kahneman, 2003), we propose that there must also be a second 'controlled route' by which a stranger can be placed in trait space via explicit reasoning about their appearance and behaviour.

Explicit reasoning about the likely traits of a newly encountered individual may be based on evidence accumulated from various sources. Innumerable behaviours provide clues about the traits of others, ranging from complex, protracted displays of altruism or treachery, to 'thin slices' of behaviour (Ambady & Rosenthal, 1992), such as facial expressions. It is often possible to directly observe these behaviours (first hand evidence). However, we can also learn about strangers' behaviours from others (second hand evidence). For example, our estimation of a stranger's traits may be informed by the reactions of known others to that individual. We can also learn about the previous interactions and behaviours of strangers through written or verbal communication (e.g., 'gossip'; Feinberg, Willer, & Schultz, 2014).

While the cognitive processes underlying the controlled and automatic inference of others' traits likely dissociate, the controlled route may play a key role in the development of the automatic route. By allowing observers to simultaneously represent an individual's traits as well as their facial appearance, the controlled route ensures that the conditions for associative learning are met (see Section 4.4). The two routes may also interact in other ways. For

example, explicit reasoning may override activation of automatic face-trait mappings, analogous to the way in which top-down control processes can override learned stereotypes (Devine, 1989). Similarly, automatic face-trait mappings may bias explicit reasoning about a stranger's traits. For example, first impressions may sometimes shift the criterion for positive evaluation such that people with an untrustworthy appearance need to work harder to be considered kind.

#### 4.4 Face-trait mappings

How might automatic mappings between face space and trait space be formed? Once an observer is able to simultaneously represent an individual's facial appearance in face space and some aspect of their character in trait space, a face-trait mapping may be established. According to the TIM framework, mappings between points or regions in face space and points in trait space are learned through correlated face-trait experience. Several learning traditions have converged on the idea that covariation fosters association or grouping (Aslin & Newport, 2012; Pearce & Bouton, 2001; Rescorla & Wagner, 1972). Consistent with this view, we propose that face-trait mappings emerge when a predictive relationship (contingency) exists between representations in face-space and trait-space. Put simply, when one repeatedly encounters individuals with a particular face shape or facial feature who subsequently exhibit a particular trait, a mapping forms between the corresponding face and trait representations. Once acquired, these mappings mediate spontaneous trait judgements from faces: where the face of a stranger falls close to a mapped location in face space, excitation automatically propagates to the associated representation in trait space (Figure 2). Individuals may begin to acquire mappings during infancy, before mature face and trait representation spaces have developed, and may continue to acquire and refine mappings throughout their lives.

#### Figure-2

The specificity of the mapping may reflect the nature of the face-trait covariation. Specific face-trait contingencies between a single facial identity and particular trait signature might yield mappings between particular *points* in face and trait space. Conversely, non-specific contingencies between a broadly defined "face type" and a nebulous cluster of traits might yield mappings between corresponding *regions* in face and trait space. The latter may

correspond to intergroup judgements where observers acquire mappings between a more general group identity and particular traits (e.g., taking the form of stereotypes). In addition, we speculate that some mappings may be bidirectional, whereby active trait representations can also excite face representations.

### **5. Evidence for ontogenetic mapping**

Three lines of evidence suggest that face-trait mappings emerge as a result of learning. First, lab-based empirical research suggests that, when learning about particular individuals, brief periods of face-trait training can affect the inference of traits from faces. Falvello, Vinson, Ferrari and Todorov (2015) presented adult participants with unfamiliar faces for 4.5 seconds each. These faces were paired 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 intended to reveal features of the person’s character. At test, participants were presented with a subset of these faces without their descriptors and asked how trustworthy they appeared to be. Faces that had previously been paired with positive information were judged to be more trustworthy. Related to this, we note that adults have been shown to generalise information that they have learned about the relative trustworthiness of one individual to novel individuals with similar facial appearance (Verosky & Todorov, 2010).

Second, in relation to spontaneous trait inferences about strangers, preliminary cross-cultural work suggests some variability in the trait inferences made by different cultural groups (Na & Kitayama, 2011; Sofer et al., 2017; Walker, Jiang, Vetter, & Sczesny, 2011; Zebrowitz et al., 2012). For example, Zebrowitz and colleagues compared the trait inferences made by US undergraduates and the culturally-isolated Tsimane people who live in the Bolivian Rainforest (Zebrowitz et al., 2012). When rating Caucasian 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. Similarly, the cues used by Western observers to infer aggressiveness, extroversion, social skills, and trustworthiness, elicit weaker trait judgements in East Asian observers (Walker et al., 2011). Interestingly, East Asian observers are also slower to make trait inferences from faces, perhaps because trait representations are a less salient aspect of personality understanding in these cultures (Walker et al., 2011).

Third, we appear to base trait inferences on aspects of facial appearance originating in recent human history. For example, research suggests that individuals who wear glasses are perceived to be more intelligent, industrious, dependable, and trustworthy (Hellström & Tekle, 1994; Manz & Lueck, 1968; Thornton, 1944). This connection has at times taken on enormous cultural significance. For example, the Khmer Rouge targeted individuals with glasses as they were believed to be a sign of intellectualism (Ung & McElroy, 2011). Similarly, individuals who have tattoos on their face or neck are perceived as untrustworthy (Timming & Perrett, 2016) and more likely to be judged guilty in criminal justice settings (Funk & Todorov, 2013; Todorov, 2017), and individuals with facial piercings are seen as less employable, particularly in client facing roles (Timming, Nickson, Re, & Perrett, 2015). As recent cultural creations, natural selection has not had an opportunity to alter the perceptual processing of glasses, tattoos, or piercings, or act on associated trait inferences. These mappings, therefore, must be acquired culturally.

## **6. Resolving the consistency paradox**

Veridical relationships have been found between facial attributes and some trait-related behaviours (Bond, Berry, & Omar, 1994; Bonnefon et al., 2015; De Neys, Hopfensitz, & Bonnefon, 2015; Stirrat & Perrett, 2010; Zebrowitz, Voinescu, & Collins, 1996). Where veridical relationships exist – even ones that are relatively weak – consistent face-trait mappings will emerge provided children have sufficient exposure to these contingencies (e.g., Rescorla & Wagner, 1972). For other face-trait mappings, however, researchers have found no evidence of a veridical basis in reality (Olivola & Todorov, 2010; Todorov et al., 2015). At first glance, this appears to pose a challenge for our learning model because we have argued that face-trait mappings emerge only when a predictive relationship exists between representations in face space and trait space. Where facial appearance and trait-related behaviours are unrelated in the general population, consistent trait inferences appear to argue against our hypothesis. One might reasonably expect children raised in these environments to acquire no face-trait mappings, or acquire highly idiosyncratic mappings reflecting their unique face-trait experiences (Honekopp, 2006; Todorov, 2017). To the extent that there is consistency in these trait inferences, it appears paradoxical: How can a permissive learning framework yield (some) consistent face-trait mappings in environments that provide inconsistent face-trait experience?

We suggest that two factors may resolve this apparent paradox. First, we argue that cultural learning supports the emergence of consistent face-trait mappings. Second, we argue that innate physiological responses to certain facial cues may canalise (Waddington, 1942) the emergence of particular face-trait mappings.

### *6.1 Cultural induction and reinforcement*

Our day-to-day interactions with others may be a relatively poor source of correlated face-trait experience. However, various cultural instruments expose us to strong contingencies between facial appearance and character traits (Figure 3a). For many centuries, fairy tales have presented archetypal characters (Bottigheimer, 2009; Warner, 1995) that may help children develop a rudimentary trait space. Since industrialization, these character descriptions have frequently been accompanied by visual illustrations of the characters' appearance. Regularities in these illustrations may effectively 'teach' children that people with particular traits (heroism, bravery) have a certain visual appearance (square jaw, blemish-free skin). More recently, the stereotypical depiction of heroes, heroines and villains in animated films (e.g., Disney and Pixar) and in comic book art has also been widely noted (England, Descartes, & Collier-Meek, 2011; Gillam & Wooden, 2008; Reynolds, 1992). Stereotypical trait illustrations of character are not restricted to children's stories. For example, the depiction of Santa Claus may contribute to the impression that portly, round faced men with white facial hair are jovial and kind in several Westerns cultures (Belk, 1987). Similarly, the typecasting of feature-film actors may serve as a rich source of correlated face-trait experience (Zuckerman, Kim, Ukanwa, & Von Rittmann, 2003).

Some of the cultural devices that transmit face-trait mappings are relatively modern (e.g., animations, television, and film). Others, however, have been around much longer. For example, there are interesting differences between the depiction of the virtuous and the unworthy in many forms of religious iconography (Morgan, 1997). In other cultural contexts, the use of masks to depict facial traits is an age-old tradition (Pernet, 1992). In the masks made by the Dan people of West Africa, for example, narrow eyes and an oval face are used to depict mild mannered traits, whereas large and protruding eyes are used to depict vigorous and aggressive individuals (Fischer, 1967; Jędrej, 1986). Interestingly, these cultural instruments potentially reinforce different face-trait mappings to Western cultures in which

large eyes are often used to depict trustworthiness (Zebrowitz, Fellous, Mignault, & Andreoletti, 2003).

While some cultural devices may reinforce trait stereotypes inadvertently, other devices are explicitly designed to foster particular face-trait mappings. Propaganda provides an example of this type of cultural device. For instance, fear of cheap immigrant labour from China prompted a spate of anti-Chinese propaganda in California in the 1870s (Figure 3b). These images intentionally foster mappings between East Asian facial features and negative traits (e.g. aggression, deceitfulness).

### Figure-3

Once particular face-trait mappings are established, they can easily spread both within and between generations through cultural learning. We hypothesise that because adults are reliable in their judgements of who appears trustworthy and untrustworthy, children will be exposed to consistent differences in how caregivers and other individuals respond to unfamiliar individuals who vary in their facial appearance. Caregivers may explicitly warn their children to avoid individuals with particular facial features; for example, making reference to apparent untrustworthiness.

Caregivers' responses to particular facial features may also leak into their non-verbal behaviour in a way that children can detect and learn from (Weisbuch, Pauker, & Ambady, 2009). Research on social referencing has shown that young children are extremely sensitive to other people's non-verbal behaviour and that they use their nonverbal responses to learn vicariously about the world. For example, infants are reluctant to approach a stranger unless their mother has interacted with them positively (Fein, 1975). We speculate that this form of learning may, therefore, produce face-trait mappings quite early in development. This tendency to rely on others' nonverbal responses persists throughout childhood. Preschool children acquire negative attitudes towards strangers after observing people interact with them in a negative way (Castelli, De Dea, & Nesdale, 2008; Skinner, Meltzoff, & Olson, 2016). Other research has shown that from the age of 4, children are sensitive to non-verbal displays of dominance and status (Over & Carpenter, 2015). To the extent that adults respond to individuals who appear dominant with consistent non-verbal responses (Tiedens &

Fragale, 2003), observing third party interaction will provide children with an opportunity to learn face-trait mappings about dominance.

### *6.2 Canalisation through physiological states*

It is widely thought that certain facial features elicit physiological states in observers. For example, so-called ‘infant schema’ appears to elicit positive feelings and may encourage nurturing and approach-related behaviours (Glocker et al., 2009). Similarly, facial disfigurements elicit aversion responses (Ryan, Oaten, Stevenson, & Case, 2012), and attractive faces may elicit sexual arousal (Rhodes, 2006). To some degree these stimulus-response (S-R) mappings may be hardwired and conserved across species (Glocker et al., 2009; Rhodes, 2006; Ryan et al., 2012). Crucially, we argue that conserved S-R behaviours do not constitute a trait inference *per se*. Conceptually similar “fight or flight” responses can be elicited by other types of visual stimulus – including spiders (Rakison & Derringer, 2008) and snakes (LoBue & DeLoache, 2008) – to which we do not typically ascribe traits. Moreover, innate S-R behaviours can be seen in a host of species seemingly incapable of meaningful trait inferences (Lorenz, 1943; Shibasaki & Kawai, 2009).

Nevertheless, there is overwhelming evidence that physiological states (e.g. arousal, emotion) alter human decision making (Lerner, Li, Valdesolo, & Kassam, 2015; Phelps, Lempert, & Sokol-Hessner, 2014). We speculate that these states bias explicit reasoning about strangers’ traits (see Section 4.3). For example, when locating strangers in trait space via the controlled route, the induction of positive affect in observers may shift the criterion for positive evaluation. Frequently when observing the behaviours of strangers, their intentions and motivations are ambiguous. Beautiful faces and infantile facial features may elicit positive feelings in the observer that promote a favourable interpretation of ambiguous behaviours – in these situations we may be inclined to give strangers the benefit of the doubt. Consequently, observers may more likely to map beautiful and infantile faces to positively valenced traits (e.g. trustworthiness).

### *6.3 The importance of studying developmental trajectories*



The vast majority of research on spontaneous first impressions has been conducted with adults. More recently, however, developmental psychologists have become interested in this topic (Cogsdill et al., 2014; Jessen & Grossmann, 2016, 2017). We believe that this work may prove to be revealing with respect to the mechanisms underlying consistent trait inferences. Specifically, we hypothesize that the face-trait contingencies we experience early in development might be particularly influential. Correlated face-trait experiences in childhood (including caregivers' reactions to strangers and cultural devices designed to teach children about heroes and villains) may induce consistent face-trait mappings that are hard to 'unlearn' later in life.

The importance of studying trait inferences across the lifespan can be illustrated with work from the associative learning tradition. In renewal paradigms, participants first learn that a stimulus predicts one outcome (e.g. a reward). They are then transferred to another context where they learn that the same stimulus no longer predicts that outcome (e.g., it may instead predict no outcome or the delivery of an aversive stimulus). Crucially, however, when returned to the original context or transferred to a novel context, subjects once again expect the stimulus to signal reward (Bouton & King, 1983; Nelson, Sanjuan, Vadillo-Ruiz, Pérez, & León, 2011; Peck & Bouton, 1990). Results such as these suggest that second-learned contingencies do not overwrite the associations first acquired by a stimulus; rather, second-learned contingencies exist in parallel, but only manifest in particular contexts (also see 'spontaneous recovery'; e.g., Rescorla, 2004). Insights from this area have previously been used to explain behaviour in a range of contexts including relapse of addicts (Bouton, 1994, 2002). In relation to first impressions, this work suggests that first-learned face-trait mappings acquired early in development may endure and continue to generalise to novel contexts throughout the individual's lifetime. In contrast, face-trait mappings acquired later in life may tend to affect behaviour in particular contexts.

More generally, development is an inherently recursive process in which early experiences can set in motion patterns of learning that affect later outcomes (Heckman, 2006; Master & Walton, 2013). Previous developmental research on the origins of intergroup divisions has shown that expectations about other people are self-reinforcing in at least two important respects. First, expectations can bias the interpretation of ambiguous situations such that the behaviour of disliked individuals is more likely to be interpreted in a negative way (Sagar &

Schofield, 1980). Second, memory for social information is biased such that behaviour that conforms to initial expectations is more likely to be remembered (Baron & Dunham, 2015; Liben & Signorella, 1980). Thus, we predict that children may be more likely to remember behaviours in which individuals who appear to be untrustworthy act in an untrustworthy way, reinforcing early face-trait mappings.

## **7. The role of facial emotion cues**

Perceived emotion cues are thought to contribute to spontaneous trait inferences, most notably trustworthiness (Montepare & Dobish, 2003; Todorov et al., 2015; Todorov et al., 2008). When meeting a stranger, it is not always easy to distinguish their permanent facial features from their transient facial expressions; whether someone is scowling or simply has narrow eyes. According to the emotion overgeneralisation hypothesis (e.g., Todorov, 2017; Zebrowitz & Zhang, 2011), permanent facial features which resemble subtle facial emotions may provoke inferences in line with those provoked by the corresponding emotional expression. Thus, someone with narrow eyes might be judged untrustworthy because misperceived facial anger is threatening cue. Consistent with this view, observers with impaired emotion recognition show atypical judgements of trustworthiness (Brewer, Collins, Cook, & Bird, 2015; Sprengelmeyer et al., 2016). Explaining the consistency between observers in these judgements is relatively easy. Insofar as the ability to recognise and interpret so-called ‘basic emotions’ is widespread across the globe (Elfenbein & Ambady, 2002; Russell, 1994), one would expect observers in many cultures to derive similar judgements from the presence of perceived emotion.

While the emotion overgeneralisation hypothesis offers an account of some of the cues used to infer the traits of others – the input into the system – the nature of the inferential process remains opaque. The TIM framework offers a means by which to delineate the potential mechanisms. First, trait judgements derived from facial emotion may be products of explicit reasoning (the ‘controlled route’). When asked about the *traits* of the to-be-judged individual, observers may extrapolate from the current perceived emotional *state* of the target. In the absence of any other information, one might reason that someone who is smiling in the current stimulus image is likely to be cheerful or kind in general (Sutherland et al., 2013). Alternatively, trait judgements derived from facial emotion could be mediated by mappings between face and trait space (the ‘automatic route’). Countless cultural devices

teach us that traits and facial expression co-vary; for example, picture books frequently use facial expressions to illustrate the character of protagonists (Nikolajeva & Scott, 2006). Exposure to material of this nature may encourage mappings between regions of face space encompassing expression variation and points in trait space. Moreover, the observation of facial affect may induce physiological states in observers that helps canalise the emergence of particular face-trait mappings (see Section 6.2). For example, the detection of facial anger may bias how observers interpret the behaviour of strangers, and thereby encourage mappings between facial anger and negative traits (e.g., untrustworthiness).

It is important to note that evidence suggesting trustworthiness judgements are reliable after brief stimulus presentations (Todorov et al., 2009; Willis & Todorov, 2006) does not show the automaticity of the trait inference *per se*. Rather, it may indicate that the cues on which judgements are based are available quickly. It is well known that facial emotion competes effectively for early visual processing resources; for example, emotional faces are harder to mask from conscious awareness, relative to emotionally neutral faces (Jiang & He, 2006; Yang, Zald, & Blake, 2007).

## **8. Predictions and falsification**

To be of value, psychological theories must make novel, testable predictions and be falsifiable. As we have discussed throughout this paper, the TIM framework makes a number of interesting empirical predictions for future research. Some of these predictions we have also highlighted in the table. In the remainder of Section 8, we wish to consider in more detail three complementary empirical approaches that have the potential to falsify the key tenet of our model; that the face-trait mappings responsible for automatic trait inferences are products of the environments in which we are raised.

### Table

Developmental data demonstrating that young infants make behavioural predictions about individuals whose faces vary in apparent trustworthiness, dominance, competence, etc., would challenge the view that all mappings between face space and trait space are acquired through experience. Such a finding would be particularly striking if it were in a sample of infants too young to have had extensive opportunities to learn about other individuals through

social referencing. In order to provide a convincing falsification of our framework, it would be necessary to distinguish behavioural products of nascent trait inferences from simple reactions to expressions of emotion. Several previous studies have reported precocious inference of facial trustworthiness by infants and young children (Cogsdill et al., 2014; Jessen & Grossmann, 2016, 2017). In each of these studies, however, facial trustworthiness is confounded with subtle yet perceptible differences in emotional expression. Consequently, apparent evidence of precocious inference of trustworthiness may simply reflect the fact that infants can detect and recognise subtle facial emotions. For example, the finding that infants preferentially look towards a face that adults would label as trustworthy (Jessen & Grossmann, 2016) may simply indicate that infants can detect subtle smile cues and experience feelings of reward as a result (Kuchuk, Vibbert, & Bornstein, 1986). In this context, we note that animals not typically thought to engage in trait inferences (sheep) show reliable preferences for smiling human faces over angry faces (Tate, Fischer, Leigh, & Kendrick, 2006).

Twin studies offer a second means by which the central claim of the TIM framework could be falsified. TIM predicts that there ought to be similar concordance between the face-trait mappings of identical (MZ) and fraternal (DZ) twins where both siblings have been raised in the same environment. If, however, the cognitive architecture necessary for making trait inferences from faces is genetically specified, then the trait inferences of MZ twins should show greater similarity to each other than do those of DZ twins. Offering some support for our position, we note that related empirical work has shown that judgments of facial attractiveness are influenced more strongly by the environment than by genes (Germine et al., 2015). Importantly, to demonstrate that face-trait mappings have an innate basis, authors would need to demonstrate that greater concordance between MZ twins, where observed, was not the result of genetic differences in the processing of faces – including the perception of facial emotion cues – or traits more generally. Face perception ability is known to be a heritable trait (Shakeshaft & Plomin, 2015; Wilmer et al., 2010; Zhu et al., 2010). In a similar way, genetic influences likely affect individuals' ability to learn about the traits of others. By shaping the development of face space and trait space independently, genetic factors may conceivably increase the concordance in facial trait judgements of MZ twins without acting on face-trait mappings *per se*.

Finally, cross-cultural research will provide an important test of the TIM framework. Many trait inferences may be broadly similar across cultures. For example, one would expect observers in many cultures to derive similar judgements from explicit reasoning about perceived emotion cues (see Section 7). Similarly, the physiological states induced by facial beauty, facial disfigurement, and facial affect, may canalise similar face-trait mappings in observers around the world (see Section 6.2). Crucially, however, TIM predicts that many of the spontaneous trait inferences made by observers *will* differ across cultures (e.g., intelligence, competence). Furthermore, TIM makes a specific prediction about the nature of these cross-cultural differences; namely that, where differences are observed, these differences ought to be closely related the available cultural input. To offer one example, in cultures where Santa Claus is a prominent cultural figure, cues such as a white beard, a round face and an elderly appearance ought to be more strongly associated with traits such as trustworthiness and generosity than in cultures where this cultural association is less common. Evidence that trait inferences are insensitive to variability in environmental input would therefore challenge a key prediction of the TIM framework.

## **9. Implications and concluding thoughts**

In the present paper, we have outlined a novel account of the origins of first impressions from faces. Our TIM framework asserts that trait inferences from faces result from face-trait mappings, acquired during ontogeny, that permit excitation of face representations to propagate automatically to trait representations. In describing our model, we presented evidence that the development of face space (Section 4.1) and trait space (Section 4.2) are both subject to considerable environmental control. We then argued that mappings between locations within these representation spaces emerge as a result of correlated face-trait experience (Section 4.3), and cite preliminary findings consistent with this view (Section 5). Contrary to existing ideas about the origins of trait judgements (Section 3), we propose a limited role for innate factors and natural selection in the induction of face-trait mappings. Our model explains inter-observer consistency by appealing to cultural learning (Section 6.1) and by citing physiological responses elicited by particular types of stimuli that canalise certain face-trait mappings (Section 6.2). Finally, we discussed how emotion cues may support trait inferences within the TIM framework (Section 7), before considering the predictions and falsification of the model (Section 8).

To date, much of the literature on first impressions has focussed on the traits inferred from faces. Indeed, within this literature authors frequently adopt a narrow definition of “facial cue” focussing on the internal features of faces, and disregarding the contribution of other cues (e.g., hairstyle, facial hair, hairstyle, glasses, tattoos, jewellery, and piercings). An important feature of the TIM framework is its ability to explain a wide range of trait inferences. If trait inferences from faces are learned, then any aspect of facial appearance can, in principle, become associated with particular trait representation. For example, our model can easily be extended to explain why wearing glasses is associated with inferences about intelligence (Hellström & Tekle, 1994; Manz & Lueck, 1968; Thornton, 1944) and why facial tattoos are associated with inferences about criminality (Funk & Todorov, 2013; Timming & Perrett, 2016). Indeed, our framework could be developed further to explain trait inferences from body shape (Teachman & Brownell, 2001; Vartanian & Silverstein, 2013), walking gait (Koppensteiner, Stephan, & Jäschke, 2016; Thoresen, Vuong, & Atkinson, 2012), and vocal cues (McAleer, Todorov, & Belin, 2014; Rezlescu et al., 2015). In each of these domains, we speculate that visual representations may come to excite locations in a common trait space by virtue of acquired mappings comparable to those hypothesised for faces.

TIM offers a single unifying framework within which automatic trait inferences from social perception cues may be understood. In contrast, the view that trait inferences from faces are mediated by an innate mechanism that conferred an evolutionary advantage on our ancestors (Oosterhof & Todorov, 2008; Van Vugt & Grabo, 2015; Zebrowitz, 2004; Zebrowitz & Zhang, 2011) cannot explain trait inferences from modern artefacts (glasses, piercings) and fashions (tattoos, contemporary hairstyles) as these cues simply emerged too recently in human history to have been influenced by natural selection. The fact that observers make consistent trait inferences from features not in the evolutionary environment (e.g. glasses, make-up, tattoos) underscores the plausibility of a learning account; these examples are positive evidence for the role of learning in the acquisition of face-trait mappings. Such observations represent a proof of sufficiency; evidently it is possible to explain all trait inferences – both idiosyncratic and normative – by appealing to learning. Once one accepts that some of the normative inferences are products of learning, the question becomes: is it necessary to posit any innate face-trait mappings? These findings shift the onus back to

proponents of more strongly nativist accounts to defend their assertion that some mappings are innate.

Finally, TIM has important practical implications. The influence of first impressions on our choices and behaviours has been well-documented; for example, research has shown that face-trait mappings influence employment opportunities (Olivola, Eubanks et al., 2014), voting patterns (Ballew & Todorov, 2007), and sentencing decisions (Funk & Todorov, 2013; Wilson & Rule, 2015). As a result, associations between appearance and perceived traits may contribute to systematic societal injustice (Devine, 1989). If these judgements are inflexible, then the scope for intervention is limited; relatively little may be done beyond increasing awareness and informing people about the dangers of relying first impressions (Olivola, Funk et al., 2014). However, if trait inferences are shaped by cultural learning, then it may be possible to mitigate widely held but deleterious societal beliefs through modifying the available cultural input.

## References

- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*(2), 256-274.
- Aslin, R. N., & Newport, E. L. (2012). Statistical learning from acquiring specific items to forming general rules. *Current Directions in Psychological Science, 21*(3), 170-176.
- Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology, 67*, 159-186.
- Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the USA, 104*, 17948-17953.
- Baron, A. S., & Dunham, Y. (2015). Representing "Us" and "Them": building blocks of intergroup cognition. *Journal of Cognition and Development, 16*(5), 780-801.
- Batres, C., Re, D. E., & Perrett, D. I. (2015). Influence of perceived height, masculinity, and age on each other and on perceptions of dominance in male faces. *Perception, 44*, 1293-1309.
- Belk, R. W. (1987). A child's Christmas in America: Santa Claus as deity, consumption as religion. *The Journal of American Culture, 10*(1), 87-100.
- Bond, C. F., Berry, D. S., & Omar, A. (1994). The kernel of truth in judgments of deceptiveness. *Basic and Applied Social Psychology, 15*(4), 523-534.
- Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2015). Face-ism and kernels of truth in facial inferences. *Trends in Cognitive Sciences, 19*, 421-422.
- Bottigheimer, R. B. (2009). *Fairy Tales: A New History*. Albany: State University of New York Press.



- 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(10), 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(3), 248-265.
- Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology, Public Policy, and Law*, 18, 906-921.
- Brewer, R., Collins, F., Cook, R., & Bird, G. (2015). Atypical trait inferences from facial cues in alexithymia. *Emotion*, 15, 637-643.
- Buckingham, G., DeBruine, L. M., Little, A. C., Welling, L., Conway, C. A., Tiddeman, B. P., et al. (2006). Visual adaptation to masculine and feminine faces influences generalized preferences and perceptions of trustworthiness. *Evolution and Human Behavior*, 27(5), 381-389.
- Burton, A. M., & Jenkins, R. (2011). Unfamiliar face perception. In A. J. Calder, G. Rhodes, J. V. Haxby & M. H. Johnson (Eds.), *The Oxford Handbook of Face Perception* Oxford: Oxford University Press.
- Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, 41, 1179-1208.
- Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6, 641-651.
- Castelli, L., De Dea, C., & Nesdale, D. (2008). Learning social attitudes: Children's sensitivity to the nonverbal behaviors of adult models during interracial interactions. *Personality and Social Psychology Bulletin*, 34(11), 1504-1513.

- Chaiken, S., & Trope, Y. (1999). *Dual-Process Theories in Social Psychology*: Guilford Press.
- Cogsdill, E. J., Todorov, A. T., Spelke, E. S., & Banaji, M. R. (2014). Inferring character from faces a developmental study. *Psychological Science, 25*, 1132-1139
- Conway, J., Catmur, C., & Bird, G. (2017, September, 25). What's wrong with social cognition research? Paper presented at A Penny for Your Thoughts: A Workshop on Social Cognition. London, UK.
- Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition, 127*(1), 6-21.
- De Neys, W., Hopfensitz, A., & Bonnefon, J. F. (2015). Adolescents gradually improve at detecting trustworthiness from the facial features of unknown adults. *Journal of Economic Psychology, 47*, 17-22.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417-440.
- Dion, K. K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*, 285-290.
- Dockett, S. (1998). Constructing understandings through play in the early years. *International Journal of Early Years Education, 6*(1), 105-116.
- Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development, 51*, 162-170.
- Dotsch, R., Hassin, R. R., & Todorov, A. (2016). Statistical learning shapes face evaluation. *Nature Human Behaviour, 1*, 0001.

- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin*, *128*(2), 203-235.
- Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Common neural mechanisms for the evaluation of facial trustworthiness and emotional expressions as revealed by behavioral adaptation. *Perception*, *39*(7), 931-941.
- England, D. E., Descartes, L., & Collier-Meek, M. A. (2011). Gender role portrayal and the Disney princesses. *Sex Roles*, *64*(7-8), 555-567.
- 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.
- Fein, G. G. (1975). Children's sensitivity to social contexts at 18 months of age. *Developmental Psychology*, *11*, 853-854.
- Feinberg, M., Willer, R., & Schultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychological Science*, *25*(3), 656-664.
- Fischer, E. (1967). Erzahlgut der Westlichen Dan in Liberia. *Anthropos*, *62*, 686-744.
- 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.
- Funk, F., & Todorov, A. (2013). Criminal stereotypes in the courtroom: Facial tattoos affect guilt and punishment differently. *Psychology, Public Policy, and Law*, *19*(4), 466-478.
- 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*, 797-815.
- Germine, L., Russell, R., Bronstad, P. M., Blokland, G. A., Smoller, J. W., Kwok, H., et al. (2015). Individual aesthetic preferences for faces are shaped mostly by environments, not genes. *Current Biology*, *25*(20), 2684-2689.

- Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: face learning ability peaks after age 30. *Cognition*, *118*(2), 201-210.
- Gillam, K., & Wooden, S. R. (2008). Post-princess models of gender: The new man in Disney/Pixar. *Journal of Popular Film and Television*, *36*(1), 2-8.
- Glocker, M. L., Langleben, D. D., Ruparel, K., Loughhead, 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.
- Gomez, J., Barnett, M. A., Natu, V., Mezer, A., Palomero-Gallagher, N., Weiner, K. S., et al. (2017). Microstructural proliferation in human cortex is coupled with the development of face processing. *Science*, *355*(6320), 68-71.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557-559.
- Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*, 330-337.
- Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900-1902.
- Hellström, Å., & Tekle, J. (1994). Person perception through facial photographs: Effects of glasses, hair, and beard on judgments of occupation and personal qualities. *European Journal of Social Psychology*, *24*(6), 693-705.
- Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, *344*(6190), 1243091.
- Honekopp, J. (2006). Once more: is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 199-209.

- Jędrej, M. C. (1986). Dan and Mende masks: A structural comparison. *Journal of the International African Institute*, 56, 71-80.
- Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society of London B*, 366, 1671-1683.
- 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. (2017). Neural evidence for the subliminal processing of facial trustworthiness in infancy. *Neuropsychologia*.
- Jiang, Y., & He, S. (2006). Cortical responses to invisible faces: dissociating subsystems for facial-information processing. *Current Biology*, 16(20), 2023-2029.
- Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience*, 6, 766-774.
- Johnson, M. H., Dziurawiec, S., Ellis, H. D., & Morton, J. (1991). Newborns' preferential tracking of faces and its subsequent decline. *Cognition*, 38, 1-19.
- Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, 58(9), 697-720.
- Keating, C. F., Mazur, A., & Segall, M. H. (1981). A cross-cultural exploration of physiognomic traits of dominance and happiness. *Ethology and Sociobiology*, 2, 41-48.
- Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the USA*, 104, 12577-12580.
- Koppensteiner, M., Stephan, P., & Jäschke, J. P. M. (2016). Moving speeches: Dominance, trustworthiness and competence in body motion. *Personality and Individual Differences*, 94, 101-106.

- Kuchuk, A., Vibbert, M., & Bornstein, M. H. (1986). The perception of smiling and its experiential correlates in three-month-old infants. *Child Development, 57*, 1054-1061.
- Lee, K., Talwar, V., McCarthy, A., Ross, I., Evans, A., & Arruda, C. (2014). Can classic moral stories promote honesty in children? *Psychological Science, 25*(8), 1-7.
- Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology, 66*, 799-823.
- Liben, L. S., & Signorella, M. L. (1980). Gender-related schemata and constructive memory in children. *Child Development, 51*, 11-18.
- LoBue, V., & DeLoache, J. S. (2008). Detecting the snake in the grass: Attention to fear-relevant stimuli by adults and young children. *Psychological Science, 19*(3), 284-289.
- Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development, 74*(4), 1130-1144.
- Lorenz, K. (1943). Die angeborenen formen möglicher erfahrung. *Ethology, 5*(2), 235-409.
- Mackintosh, N. J. (1998). *IQ and Human Intelligence*. Oxford: Oxford University Press.
- Manz, W., & Lueck, H. E. (1968). Influence of wearing glasses on personality ratings: crosscultural validation of an old experiment. *Perceptual and Motor Skills, 27*(3), 704.
- Mascaro, O., & Csibra, G. (2014). Human infants' learning of social structures the case of dominance hierarchy. *Psychological Science, 25*(1), 250-255.
- Master, A., & Walton, G. M. (2013). Membership in a minimal group increases motivation and learning in young children. *Child Development, 84*, 737-751.
- McAlear, P., Todorov, A., & Belin, P. (2014). How do you say "Hello"? Personality impressions from brief novel voices. *PloS One, 9*(3), e90779.

- Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior, 27*, 237-254.
- Moon, C., Panneton-Cooper, R., & Fifer, W. E. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development, 16*, 495-500.
- Morgan, D. (1997). *Visual Piety: A History and Theory of Popular Religious Images*. London: University of California Press.
- Murphy, J., Ipser, I., Gaigg, S. B., & Cook, R. (2015). Exemplar variance supports robust learning of facial identity. *Journal of Experimental Psychology: Human Perception and Performance, 41*, 577-581.
- Na, J., & Kitayama, S. (2011). Spontaneous trait inference is culture-specific: behavioral and neural evidence. *Psychological Science, 22*, 1025-1032.
- Nelson, J. B., Sanjuan, M. D. C., Vadillo-Ruiz, S., Pérez, J., & León, S. P. (2011). Experimental renewal in human participants. *Journal of Experimental Psychology: Animal Behavior Processes, 37*(1), 58-70.
- Nikolajeva, M., & Scott, C. (2006). *How Picturebooks Work*. New York: Routledge.
- Norenzayan, A., Choi, I., & Nisbett, R. E. (2002). Cultural similarities and differences in social inference: Evidence from behavioral predictions and lay theories of behavior. *Personality and Social Psychology Bulletin, 28*(1), 109-120.
- Olivola, C. Y., Eubanks, D. L., & Lovelace, J. B. (2014). The many (distinctive) faces of leadership: Inferring leadership domain from facial appearance. *The Leadership Quarterly, 25*(5), 817-834.
- 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., & Carpenter, M. (2015). Children infer affiliative and status relations from watching others imitate. *Developmental Science, 18*(6), 917-925.
- Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology, 52*, 111-139.
- Peck, C. A., & Bouton, M. E. (1990). Context and performance in aversive-to-appetitive and appetitive-to-aversive transfer. *Learning and Motivation, 21*(1), 1-31.
- Pernet, H. (1992). *Ritual Masks: Deceptions and Revelations*: Univeristy of South California Press.
- Phelps, E. A., Lempert, K. M., & Sokol-Hessner, P. (2014). Emotion and decision making: multiple modulatory neural circuits. *Annual Review of Neuroscience, 37*, 263-287.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(04), 515-526.
- Rakison, D. H., & Derringer, J. (2008). Do infants possess an evolved spider-detection mechanism? *Cognition, 107*(1), 381-393.
- Re, D. E., & Perrett, D. I. (2014). The effects of facial adiposity on attractiveness and perceived leadership ability. *Quarterly Journal of Experimental Psychology, 67*, 676-686.
- Rescorla, R., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F.



- Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.
- Rescorla, R. A. (2004). Spontaneous recovery. *Learning & Memory, 11*(5), 501-509.
- Reyes-García, V., Kightley, E., Ruiz-Mallén, I., Fuentes-Pelaez, N., Demps, K., Huanca, T., et al. (2010). Schooling and local environmental knowledge: Do they complement or substitute each other? *International Journal of Educational Development, 30*(3), 305-313.
- Reynolds, R. (1992). *Super Heroes: A Modern Mythology*. Jackson: University Press of Mississippi.
- Rezlescu, C., Penton, T., Walsh, V., Tsujimura, H., Scott, S. K., & Banissy, M. J. (2015). Dominant voices and attractive faces: The contribution of visual and auditory information to integrated person impressions. *Journal of Nonverbal Behavior, 39*(4), 355-370.
- Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57*, 199-226.
- Rhodes, G., & Leopold, D. A. (2011). Adaptive norm-based coding of face identity. In A. J. Calder, G. Rhodes, J. V. Haxby & M. H. Johnson (Eds.), *The Oxford Handbook of Face Perception* (pp. 263-286). Oxford: Oxford University Press.
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin, 115*(1), 102-141.
- Ryan, S., Oaten, M., Stevenson, R. J., & Case, T. I. (2012). Facial disfigurement is treated like an infectious disease. *Evolution and Human Behavior, 33*(6), 639-646.
- Sagar, H. A., & Schofield, J. W. (1980). Racial and behavioral cues in Black and White children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology, 39*(4), 590-598.

- Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion, 9*, 260-264.
- 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*, 440-444.
- Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences of the USA, 112*(41), 12887-12892.
- Shibasaki, M., & Kawai, N. (2009). Rapid detection of snakes by Japanese monkeys (*Macaca fuscata*): an evolutionarily predisposed visual system. *Journal of Comparative Psychology, 123*(2), 131-135.
- Skinner, A. L., Meltzoff, A. N., & Olson, K. R. (2016). "Catching" Social Bias: Exposure to Biased Nonverbal Signals Creates Social Biases in Preschool Children. *Psychological Science, 19*(18), 1543-1548.
- Sodian, B. (1991). The development of deception in young children. *British Journal of Developmental Psychology, 9*(1), 173-188.
- Sofer, C., Dotsch, R., Oikawa, M., Oikawa, H., Wigboldus, D. H. J., & Todorov, A. (2017). For your local eyes only: Culture-specific face typicality influences perceptions of trustworthiness. *Perception, 46*(8), 914-928.
- Sprenghelmeyer, R., Young, A. W., Baldas, E. M., Ratheiser, I., Sutherland, C. A. M., Müller, H. P., et al. (2016). The neuropsychology of first impressions: Evidence from Huntington's disease. *Cortex, 85*, 100-115.
- Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust male facial width and trustworthiness. *Psychological Science, 21*, 349-354.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662.

- Susilo, T., McKone, E., & Edwards, M. (2010). What shape are the neural response functions underlying opponent coding in face space? A psychophysical investigation. *Vision Research*, *50*, 300-314.
- Sutherland, C., 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.
- Swaddle, J. P., & Reiersen, G. W. (2002). Testosterone increases perceived dominance but not attractiveness in human males. *Proceedings of the Royal Society of London B: Biological Sciences*, *269*, 2285-2289.
- 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*, 603-620.
- Talamas, S. N., Mavor, K. I., & Perrett, D. I. (2016). Blinded by beauty: Attractiveness bias and accurate perceptions of academic performance. *PloS One*, *11*, e0148284.
- Tate, A. J., Fischer, H., Leigh, A. E., & Kendrick, K. M. (2006). Behavioural and neurophysiological evidence for face identity and face emotion processing in animals. *Philosophical Transactions of the Royal Society of London B*, *361*(1476), 2155-2172.
- Taumoepeau, M., & Ruffman, T. (2006). Mother and infant talk about mental states relates to desire language and emotion understanding. *Child Development*, *77*(2), 465-481.
- Teachman, B. A., & Brownell, K. D. (2001). Implicit anti-fat bias among health professionals: is anyone immune? *International Journal of Obesity*, *25*(10), 1525-1531.
- Thomsen, L., Frankenhuys, W. E., Ingold-Smith, M., & Carey, S. (2011). Big and mighty: Preverbal infants mentally represent social dominance. *Science*, *331*(6016), 477-480.
- Thoresen, J. C., Vuong, Q. C., & Atkinson, A. P. (2012). First impressions: Gait cues drive reliable trait judgements. *Cognition*, *124*(3), 261-271.

- Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences*, 3, 452-460.
- Thornton, G. R. (1944). The effect of wearing glasses upon judgments of personality traits of persons seen briefly. *Journal of Applied Psychology*, 28(3), 203-207.
- Tiedens, L. Z., & Fragale, A. R. (2003). Power moves: Complementarity in submissive and dominant nonverbal behavior. *Journal of Personality and Social Psychology*, 84, 558-568.
- Timming, A. R., Nickson, D., Re, D. E., & Perrett, D. (2015). What do you think of my ink? Assessing the effects of body art on employment chances. *Human Resource Management*, 56(1), 133-149.
- Timming, A. R., & Perrett, D. (2016). Trust and mixed signals: A study of religion, tattoos and cognitive dissonance. *Personality and Individual Differences*, 97, 234-238.
- Todd, J. T. (2004). The visual perception of 3D shape. *Trends in Cognitive Sciences*, 8, 115-121.
- Todorov, A. (2017). *Face Value: The Irresistible Influence of First Impressions*. Princeton: Princeton University Press.
- Todorov, A., Loehr, V., & Oosterhof, N. N. (2010). The obligatory nature of holistic processing of faces in social judgments. *Perception*, 39(4), 514-532.
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308, 1623-1626.
- Todorov, A., Olivola, C. Y., 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., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*, 455-460.
- Ung, S. K., & McElroy, T. (2011). *I Survived the Killing Fields*. Seattle: S & T Publishing.
- Vaish, A., Carpenter, M., & Tomasello, M. (2010). Young children selectively avoid helping people with harmful intentions. *Child Development*, *81*(6), 1661-1669.
- Valentine, K. A., Li, N. P., Penke, L., & Perrett, D. I. (2014). Judging a man by the width of his face the role of facial ratios and dominance in mate choice at speed-dating events. *Psychological Science*, *25*, 806-811.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology*, *43*, 161-204.
- Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology*, *44*, 671-703.
- Van Vugt, M., & Grabo, A. E. (2015). The many faces of leadership: an evolutionary-psychology approach. *Current Directions in Psychological Science*, *24*(6), 484-489.
- Vartanian, L. R., & Silverstein, K. M. (2013). Obesity as a status cue: perceived social status and the stereotypes of obese individuals. *Journal of Applied Social Psychology*, *43*(S2), E319-E328.
- Verhulst, B., Lodge, M., & Lavine, H. (2010). The attractiveness halo: Why some candidates are perceived more favorably than others. *Journal of Nonverbal Behavior*, *34*, 111-117.
- Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science*, *21*(6), 779-785.
- Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature*, *150*(3811), 563-565.

- 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(6), 609-617.
- Warner, M. (1995). *From the Beast to the Blonde: On Fairy Tales and Their Tellers*. London: Vintage.
- Webster, M. A., & MacLeod, D. I. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society of London B*, 366, 1702-1725.
- Weisbuch, Z., Pauker, K., & Ambady, N. (2009). The subtle transmission of race bias via televised nonverbal behavior. *Science*, 326(5960), 1711-1714.
- Willis, J., & Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17, 592-598.
- Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., et al. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the USA*, 107(11), 5238-5241.
- Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, 26, 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, 189-202.
- Yang, E., Zald, D. H., & Blake, R. (2007). Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion*, 7(4), 882-886.
- Zebrowitz, L. A. (2004). The origins of first impressions. *Journal of Cultural and Evolutionary Psychology*, 2, 93-108.
- 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, 194-215.

- Zebrowitz, L. A., Franklin, R. G., & Boshyan, J. (2015). Face shape and behavior: Implications of similarities in infants and adults. *Personality and Individual Differences, 86*, 312-317.
- Zebrowitz, L. A., & Montepare, J. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*(3), 1497-1517.
- Zebrowitz, L. A., Voinescu, L., & Collins, M. A. (1996). "Wide-eyed" and "crooked-faced": Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin, 22*(12), 1258-1269.
- Zebrowitz, L. A., Wang, R., Bronstad, P. M., Eisenberg, D., Undurraga, E., Reyes-García, V., 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.
- 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.
- Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., et al. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology, 20*(2), 137-142.
- Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. M. (2010). The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology, 106*(4), 208-220.
- Zuckerman, E. W., Kim, T. Y., Ukanwa, K., & Von Rittmann, J. (2003). Robust identities or nonentities? Typecasting in the feature-film labor market. *American Journal of Sociology, 108*(5), 1018-1074.

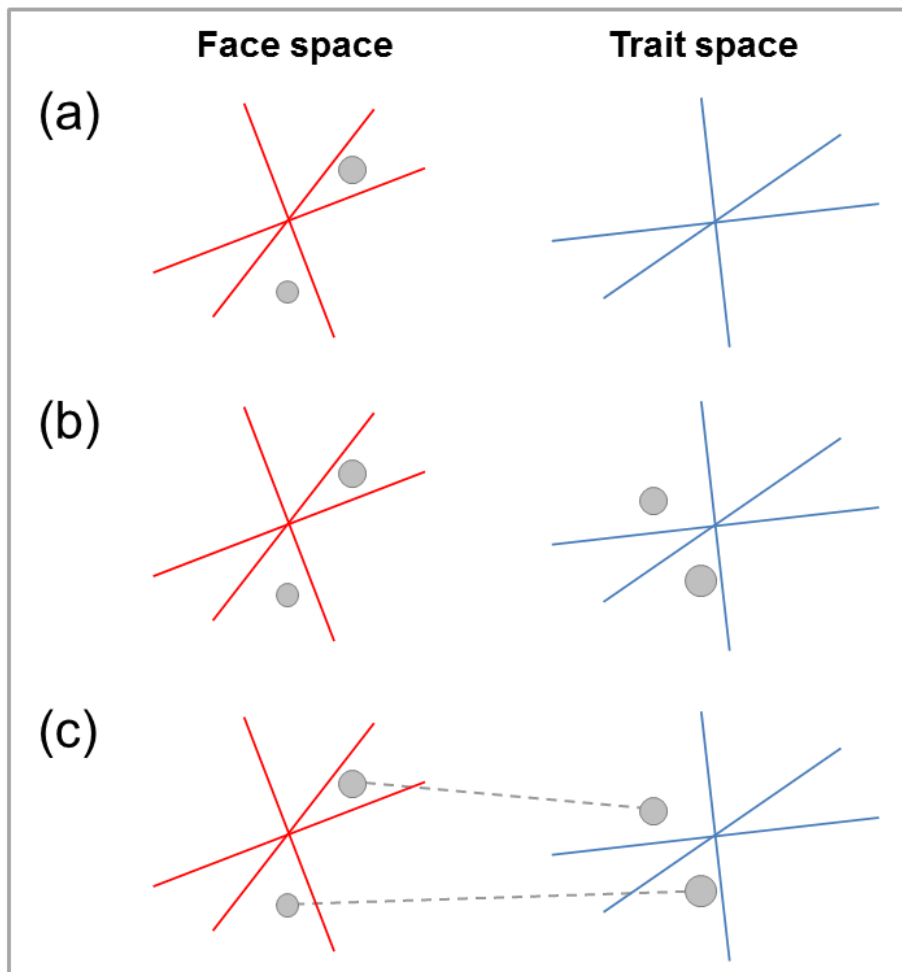
**Figure 1**



**Figure 1:** Unfamiliar faces can vary significantly in the extent to which they are perceived trustworthy (top) or untrustworthy (bottom).



**Figure 2**



**Figure 2:** Learning trait mappings *de novo*. (a) When a stranger is encountered, the observer is able to represent their face as a vector in face space. (b) Having learned about the stranger's behaviours, the observer can reason about the stranger's traits, thereby placing them in trait space. (c) Where excitation of representations in face space predicts the nature of representations in trait space over time, a face-trait mapping emerges.

**Figure 3**



**Figure 3:** Cultural induction of face-trait mappings. (a) Several cultural devices may foster face-trait mapping inadvertently. (b) Other cultural devices may foster particular face-trait mappings intentionally. For example, Anti-Chinese propaganda produced in California during the 1870s deliberately fostered mappings between East Asian facial features and negative traits.

Table: Directions for future empirical studies suggested by the TIM framework

Claim made by TIM	Prediction for future empirical research
First impressions are learned culturally (Section 5; Section 6.1; Section 8; Section 9).	<p>Some spontaneous trait inferences will differ across cultures. The specific nature of observed cultural differences will be predictable from the nature of the available cultural input.</p> <p>First impressions may change and evolve over time. For example, pale skin became associated with high status and sophistication in Elizabethan England.</p> <p>It will be possible to train novel face-trait mappings within lab-based situations.</p> <p>Caregivers will encourage and reinforce particular face-trait mappings (whether consciously or unconsciously) in interactions with their children.</p>
Face-trait mappings are products of domain-general associative processes, such as those revealed through the study of conditioning phenomena (Section 4.4; Section 6.3).	<p>The rate of associative learning is known to be determined by both temporal contiguity (the number of times A and B are paired together) and contingency (the strength of the predictive relationship between A and B). The acquisition of face-trait mappings will, therefore, be related to the number of times we encounter face-trait pairings, but also how often we encounter exceptions to the to-be-learned relationship.</p> <p>The study of conditioning phenomena has revealed that learned responses can be subject to contextual control; associations can manifest in one context but not in others. Should the acquisition of face-trait mappings conform to the same principles of associative learning, then the same face may produce different trait inferences when encountered in different contexts.</p>
Trait space is independent of face space (Section 4.2; Section 9).	Facial cues and other appearances cues (such as body shape, walking gait, hairstyle clothing) will activate representations in a common trait space. It may be possible to identify these overlapping neural codes with neuroimaging.
There are two routes (automatic, controlled) for deriving trait judgements from faces (Section 4.3).	Individuals may exist who are capable of making one type of inference, but not the other. For example, neuropsychological patients may exist who are incapable of flexible reasoning about the traits of others from thin slices of behaviour but who continue to exhibit automatic inferences acquired before their injury.