Cognition 170 (2018) 190–200

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Where do spontaneous first impressions of faces come from?

Harriet Over[a,*], Richard Cook[b]

[a] Department of Psychology, University of York, United Kingdom
[b] Department of Psychological Sciences; Birkbeck, University of London, United Kingdom

ARTICLE INFO

Keywords:
First impressions
Faces
Traits
Cultural learning
Trustworthiness
Child development

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 behaviour (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, Said, Engell, & Oosterhof, 2008; Todorov et al., 2015).

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., 2008, 2015). 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

* Corresponding author at: Department of Psychology, University of York, York YO10 5DD, United Kingdom.
E-mail address: harriet.over@york.ac.uk (H. Over).

http://dx.doi.org/10.1016/j.cognition.2017.10.002
Received 18 April 2017; Received in revised form 3 October 2017; Accepted 4 October 2017

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’ (Bonnefon, Hopfensitz, & De Neys, 2015), others appear unrelated to the true behavioural tendencies of the people being judged (Oosterhof & Todorov, 2008; Todorov et al., 2008).

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 & Reierson, 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;