Evolution and Human Behavior 39 (2018) 19–29

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehbonline.org

Original Article

The influence of shape and colour cue classes on facial health perception

Alex L. Jones

Department of Psychology, Swansea University, Vivian Tower, Singleton Park, Swansea SA2 8PP, United Kingdom.

ARTICLE INFO

Article history:
Initial receipt 29 July 2017
Final revision received 26 September 2017

Keywords:
Faces
Health
Colouration
Averageness
Cues

ABSTRACT

Facial appearance signals information about an individual, and one trait in particular is vitally important for social interaction and mate choice decisions: physical health. Facial cues to health can be divided into two broad classes - facial shape, which is linked to previous health and is relatively fixed; and facial colouration, which changes over the short-term, reflecting current health. These cue classes in themselves give insight into the kinds of health condition valued by human observers when making social inferences. Here, using novel and generalizable methods, the influence of these cue classes on health perception and their link to a measure of general health are examined. Study One employs a Brunswik lens model approach, finding that observers utilise exclusively shape cues to judge health, and that of these shape cues, only averageness is related to a measure of self-reported general health. Study Two shows that when averageness and carotenoid colouration are varied together, both make separable contributions to perceived health, but that averageness explains a larger proportion of variance. Taken together, these results indicate that humans may have evolved to favour cues to previous condition when judging health, because they are more valid. However, the findings also suggest that the role of facial appearance in perceiving health is more complex than previously thought, with different cues potentially reflecting specific aspects of physiological health.

1. Introduction

The belief that our faces convey information about our health is not a new one, existing across cultures and through history (Bridges, 2012). Recently, a growing body of literature investigating these associations has supported early ideas about the relationship between facial appearance and health. Humans can readily discriminate healthy and unhealthy looking faces (Jones, Kramer, & Ward, 2012), unconsciously associate positive attributes to healthy-looking faces (Grandfield, Thompson, & Turpin, 2005), and show strong agreement on which faces appear healthy or not (Jones, Porcheron, Sweda, Morizot, & Russell, 2016). An evolutionary approach to understanding these perceptions indicates that some facial traits may be honest cues to health, and that we evolved a sensitivity to these cues to find mates with good health (Rhodes, 2006; Scott, Clark, Boothroyd, & Penton-Voak, 2013; Thornhill & Gangestad, 2006). What are these facial traits that we use to make judgments about health, and are they accurate?

Early work on the link between health and facial appearance focused on aspects of facial shape as cues to health. Facial symmetry has long been considered to indicate good health and genetic stability (Møller & Swaddle, 1997), with symmetrically developing individuals being able to better resist environmental insults that cause deviations from perfect symmetry. In agreement with this, more symmetrical individuals report better physical and mental health (Shackelford & Larsen, 1997; Thornhill & Gangestad, 2006). Concordantly, individuals are perceived as healthier with more symmetrical faces (Rhodes et al., 2001), suggesting an evolved ability to detect physical health from facial appearance alone. Sexual dimorphism in faces, or their femininity and masculinity, also seems linked to health. For example, the immunocompetence hypothesis suggests that more masculine-looking men, whose faces developed under the influence of greater amounts of immunosuppressing testosterone, should have better health (Thornhill & Gangestad, 2006). Women with higher facial femininity also seem to report better physical health, with fewer infections per year (Gray & Boothroyd, 2012), and have greater reproductive health (Law Smith et al., 2006). For both females and males, facial averageness (the closer an individual face is to the population average) is linked with both perceived health and actual health (Rhodes et al., 2001), and seems to share a relationship with actual genetic diversity, and thus greater disease resistance (Lie, Rhodes, & Simmons, 2008). Finally, facial adiposity, or the amount of weight carried in the face, is a good predictor of perceived health (Henderson, Holzleitner, Talamas, & Perrett, 2016). Given its relationship with body mass index, adiposity is also clearly related to actual health, with underweight and overweight individuals having poorer health outcomes (Coetzee, Perrett, & Stephen, 2009).

More recent research has examined the role facial colouration plays in health perception. Skin lightness, redness, and yellowness all seem to be preferred in faces when making judgments of health (Henderson et al., 2016; Stephen, Coetzee, Law Smith, & Perrett, 2009a; Stephen, Law

E-mail address: alex.l.jones@swansea.ac.uk.

https://doi.org/10.1016/j.evolhumbehav.2017.09.005

Smith, Stirrat, & Perrett, 2009b). Carotenoid colouration, a specific combination of those colours, is strongly preferred (Lefevre & Perrett, 2015; Stephen, Coetzee, & Perrett, 2011; Tan, Graf, Mitra, & Stephen, 2017) and reflects the deposition of dietary carotenoids in the skin. This seems to be an honest cue to health, as carotenoids improve both immune function (Alexander, Newmark, & Miller, 1985) and reproductive health (Dowling & Simmons, 2009). Observers also rely on colouration in specific areas of faces when judging health. Darker skin under the eyes significantly lowers ratings of apparent health (Jones et al., 2016), as does the yellow-blue contrast around the lips (Russell et al., 2016; Stephen & McKeegan, 2010) and eyes (Russell, Sweda, Porcheron, & Mauger, 2014). These aspects of colouration also seem tied to actual health, in that sleep deprivation can cause dark circles and reddened sclera (Sundelin et al., 2013) and accelerate skin aging (Oyetakin-White et al., 2015), while poor circulation can affect lip colouration (Ponsonby, Dwyer, & Couper, 1997). Other cues to health from skin alone are related to texture homogeneity, with a smooth, even appearance seen as healthy and attractive (Matts, Fink, Grammer, & Burquest, 2007) and a cue to the absence of infectious disease (Samson, Fink, & Matts, 2010).

Physical health, then, seems manifest in the human face in the form of colouration and shape cues, and observers readily use these cues to infer health from facial appearance. These cues can be broadly separated into two 'classes', both of which seem related to distinct aspects of health. This segregation is apparent in the literature, but has not yet been widely argued for (Getty, 2002; Scott et al., 2013; Scott, Pound, Stephen, Clark, & Penton-Voak, 2010; Smith, Jones, DeBruine, & Little, 2009; Tybur & Gangestad, 2011). For example, consider that facial shape is almost entirely fixed and unchangeable by the individual after puberty (with the exception of adiposity, which takes concerted effort over time to produce a perceptual change; Re & Rule, 2016), and so is a 'class' of cues suited to reflecting long-term health and developmental stability. Indeed, the symmetry of older adults' faces reflects their childhood socioeconomic status – regardless of their health in later life, early life experiences seem to permanently shape the face (Hope et al., 2013), and this change can be seen from a young age (Özener & Fink, 2010). Averageness too is related to childhood and adolescent health (Rhodes et al., 2001; Zebrowitz & Rhodes, 2004), and twin studies suggest variance in facial averageness has a large environmental component (Lee et al., 2016). Conversely, colour cues from facial skin are a class of cues ideally suited to indicating current health, rather than previous condition. Colour cues can and do change along with health. For example, dietary increases in carotenoids alter skin colour to appear healthier (Stephen et al., 2011; Tan et al., 2017; Whitehead, Re, Xiao, Ozakinci, & Perrett, 2012), smoking cessation lightens skin (Cho et al., 2012), and acute infection causes facial skin to become paler and greener (Henderson et al., 2017). Even simply missing several hours of sleep causes colour changes in facial areas related to a healthy appearance (Axelsson et al., 2010; Jones et al., 2016; Sundelin et al., 2013), and reduces observers' desire to interact with individuals who look this way (Sundelin, Lekander, Sorjonen, & Axelsson, 2017).

While both shape and colour cues indicate health, they seem related to different aspects of it. Undoubtedly, both classes of cues are important and carry consequences for mate choice. Long term health, as indexed by shape, may cue indirect benefits to health such as disease resistance or developmental health (Lie et al., 2008), while colour may cue more direct benefits in the form of being free from disease and not currently infectious (Henderson et al., 2017; Sundelin et al., 2017). But which is more important? Evolutionary arguments have tended to focus on the former set of cues (Scheib, Gangestad, & Thornhill, 1999; Thornhill & Gangestad, 2006) because of their possible association with genetic quality, a fundamental tenet of reproductive success. However, recent mathematical models indicate that paying attention to cues to current condition can bestow fitness advantages, but there are diminishing returns to attending to cues to past health, particularly in species who have shorter lifespans and higher parasite loads (Adamo & Spiteri, 2005, 2009). Is there evidence from human behavior that might indicate a reliance on one cue class over another?

The results of studies addressing this question have been equivocal. Observers seem to rely more strongly on colouration (current condition) than masculinity (past condition) when judging male attractiveness (Scott et al., 2013, 2010), and the intermediary shape cue of adiposity is a better predictor of health than masculinity (Rantala et al., 2013). However, skin yellowness seems to better predict perceived health than adiposity (Henderson et al., 2016). There are also interactions between cues to current and previous condition, with more feminine and masculine facial shapes being more attractive when those faces appear healthy (Smith et al., 2009). Colour cues also seem to be relied on more for perceiving health when shape adiposity is very low or high (Fisher, Hahn, DeBruine, & Jones, 2014). Recent work, using novel conjoint analysis techniques, has shown even more complex interactions between cue classes, with sexual dimorphism being preferred over colour for male faces, but symmetry and colour preferred for female faces (Mogilski & Welling, 2017). Conversely, more evidence suggests colour is not utilised when judging female health, with observers relying solely on femininity, which itself shows no relationship with actual health (Foo, Simmons, & Rhodes, 2017). The cues used by observers to make judgments of health are established, but a relationship with actual health and any interactions between cue classes are unclear.

The current studies aim to reveal what facial cues observers rely on when making judgements about female health, whether those cues are valid, and what interactions there are between cue classes. In Study One, a large set of female faces is measured for femininity, averageness, and symmetry, as well as skin luminance, redness, and yellowness. These cues are selected as they represent well-researched, distinct 'classes', indicating both long-term (shape factors) and current (colour factors) health. Participants also provided information on their general health condition and health over the previous four weeks using the 12-Item Short Form Health Survey (SF-12; Ware, Kosinski, & Keller, 1996), which provides an estimate of general and recent functioning in daily life that may be related to both long and short term health.

The relationships between perceived health, a measure of general health, and attributes of facial appearance are best understood using a Brunswik lens model (Brunswik, 1956), which is applied in Study One. This analytical approach has roots in early models of decision-making (Brunswik, 1955), stemming from the idea that humans operate in rich environments where things share variance. When making a decision about something, an individual can use cues (or 'lenses') correlated with the required decision to guide their judgment. The lens model thus describes two sets of relationships. The first are known as *cue utilisation*s, which are the correlations between an observer's decision or judgment, and the lenses themselves. That is, an individual judgment of health from an observer may show a correlation with facial colouration and facial shape, one or the other, or none at all. Regardless, the model first highlights the kinds of cues that are used by an observer when making a judgment. The second set of described relationships is known as *cue validities*, or the correlations between the underlying variable and the cues themselves. These illustrate whether the cues themselves are good indicators of a directly unobservable trait – for example, does facial luminance share variance with a measure of health in daily living? In full, the model describes the cues observers utilise to make a decision, and whether the cues are valid or not. The lens model has had great influence in personality perception research (Gosling, Ko, Mannarelli, & Morris, 2002; Naumann, Vazire, Rentfrow, & Gosling, 2009; Vazire, Naumann, Rentfrow, & Gosling, 2008), but is not broadly used in evolutionary models of face perception, where it may be informative (but see Zebrowitz et al., 2014; Zebrowitz & Rhodes, 2004). Finally, Study Two examines and confirms the relative reliance on shape and colour cues based on the results of the lens model analysis, using facial manipulation techniques.

2. Study One

The Brunswik lens model is uniquely suited to analysing what classes of cues are important for perceiving health, and whether those cues are tied to an estimate of general and current health. Here, aspects of facial shape (averageness, femininity, symmetry) and colour (skin luminance, redness, yellowness) are extracted from faces, which were additionally rated for perceived health. Each participant also provided data on their health in daily life using the SF-12 (Ware et al., 1996), which captures both current and recent health within the past month, arguably indexing aspects of both stable and current health condition.

Both classes of cues may interact, since they cue different types of condition (Getty, 2002), and so there are some general predictions that can be made. First, all cues should show some evidence of *cue utilisation* – that is, perceptions of health should correlate with all cues - given the body of previous work (Rhodes et al., 2001; Stephen et al., 2009b), though it is difficult to predict what weightings may occur. Second, there will be evidence of *cue validity* for some cues. Specifically, averageness is a good candidate for being related to actual health – it is influenced by the environment, reflecting developmental stability (Lee et al., 2016), is linked to childhood and adolescent health in numerous studies (Rhodes et al., 2001; Zebrowitz & Rhodes, 2004), and shows relationships with immunocompetence (Lie et al., 2008). Femininity may also show cue validity, as it has some links with actual health (Gray & Boothroyd, 2012). Finally, symmetry may show cue validity (Shackelford & Larsen, 1997), but equally may be an invalid cue (Pound et al., 2014). It is unclear whether colouration will show cue validity, given recent equivocal findings (Foo et al., 2017; Whitehead et al., 2012).

2.1. Method

2.1.1. Participants and photo capture

One hundred and eight females (age $M = 20.56$, $SD = 2.28$) posed as models, photographed with a Canon EOS 5D Mark II camera and a Canon EF 24-105 mm lens. Participants removed all traces of cosmetics, facial jewellery (such as nose piercings), and tied their hair back from their face. Participants were instructed to maintain a neutral expression and look directly into the camera, and were compensated with £6 for their participation.

Photographs were taken with a constant shutter speed (1/100), ISO rating (100), lens aperture (F8.0), and focal length (65 mm). White balance was set to manual to avoid artefacts in lighting an automatic balancing can introduce. Photographs were taken against a white background. Between each shot, the white balance and RGGB values were kept constant (WB: 2433, RGGB: 1024, 1024, 1601). These values determine how pixels are assigned values through the de-mosaicing algorithms in modern camera sensors, which ultimately determine the overall colour temperature of images (for technical details, see Zhen & Stevenson, 2015), so their consistency is important.

It is important to note that the images here were not subject to a formal colour calibration as is commonly carried out when examining the role of colour in social perceptions. Rather, all images were captured under carefully controlled and constant camera settings. Nevertheless, in the absence of a full colour calibration, an analysis of the constant background colour present in all images was carried out to test for the presence of significant variability in lighting. This indicated a very low level of lighting variability, suggesting skin colour measurements are appropriate and accurate. The full analysis and discussion can be found in Appendix A: Supplementary data.

2.1.1.1. Design sensitivity. For the majority of analyses used in this study, the photographed participants were used as the unit of analysis. G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) was used to estimate the sensitivity of the design – that is, the likely effect size that can be reliably detected. Using correlations with $n = 108$ participants and $\alpha = 0.05$, the design is sufficiently sensitive to detect effect sizes of $r = 0.26$ with 80% power, a medium effect size (Cohen, 1992). Predictors of perceived health are frequently in that category or greater, such as skin yellowness (Henderson et al., 2016), facial femininity (Gray & Boothroyd, 2012), averageness, (Zebrowitz & Rhodes, 2004) or facial adiposity (Coetzee et al., 2009).

2.1.2. Measures

2.1.2.1. Physical health. Participants completed the SF-12 (Ware et al., 1996) to provide a composite measure of their general health and recent health over the last four weeks. The SF-12 is designed to measure general health and well-being in daily life (McCallum, 1995), and generates two scores, a physical component (PCS) and mental component (MCS). Only PCS scores were examined in this study. Respondents give an appraisal of how good their health is in general, describe whether their current health interferes with physical activities (like climbing stairs or chores), indicate whether physical pain has affected their work, and whether their health has prevented them from achieving as much as they would like. These health measures are considered from an immediate perspective and over the past four weeks. It is worth noting that the SF-12 does not discriminate clearly between long-term health, such as that experienced during childhood or adolescence, and current condition, in the form of infection, for example. However, the SF-12 is well-validated, and produces lower PCS scores in respondents with minor ailments such as chesty cough or acid indigestion, as well as more chronic health problems (Ware et al., 1996). In the absence of disease, the SF-12 also produces lower PCS scores for individuals with lower socioeconomic status (Younsi & Chakroun, 2014), a variable that is related to facial shape cues and long term health (Hope et al., 2013; Özener & Fink, 2010). Other research has shown the health constructs measured by the SF-12 seem to be reflected in facial appearance. Composite faces produced from individuals high and low on PCS scores appear different, and are discriminated with above chance accuracy (Jones et al., 2012; Kramer & Ward, 2010). However, health is a multidimensional construct and not monolithic (Eberst, 1984), and studies showing accuracy in health perception have assessed health in multiple ways with varying outcomes (Foo et al., 2017; Zebrowitz et al., 2014; Zebrowitz & Rhodes, 2004). As the SF-12 produces a composite measure of previous and current condition, the following results should be interpreted in this light and considered against the wider literature and methods of health assessment. Five participants did not respond to the full set of questions, meaning their PCS could not be computed, leaving data for 103 participants (range 22.14–64.09, mean PCS = 51.60, $SD = 7.56$). Given the sample of participants was young, and therefore might be expected to have little variation in health, a direct test of a ceiling effect was performed. PCS scores are out of a maximum of 100, and a one-sample t-test against this value confirmed no ceiling effect was present, $t(102) = 64.94$, $p < 0.001$, $d = 6.40$.

2.1.2.2. Averageness. The photographs of all 108 participants were delineated using JPsychomorph (Tiddeman, Burt, & Perrett, 2001), using a set of 160 landmarks to outline facial appearance. An anthropometric approach (Farkas, 1994) was used to calculate facial averageness, following the procedure described by Kościński (2012). From the full set of 160 landmarks, custom Python software was used to extract a set of 14 Euclidean distances, based on clearly perceived facial proportions that have been used in previous studies (Baudouin & Tiberghien, 2004; Burriss, Little, & Nelson, 2007; Cunningham, 1986). To standardise these measurements between faces, each distance was divided by the interpupillary distance (Baudouin & Tiberghien, 2004). Distances that involve measured bilateral traits were averaged pair-wise. An additional four ratio measurements were calculated, which have been linked to facial attractiveness in computational studies, but are not commonly studied in the literature (Chang & Chou, 2009). To compute averageness, the 18 measurements were Z-scored across participants to provide a standardised distribution for each measurement. Then, for each participant, absolute *Z*-scores were summed across all 18 measurements to arrive at the averageness index. These scores were multiplied by −1, such that lower scores indicate a more average face. This metric provides a summed score of how each face differs from the average – faces that diverge more from the average across measurements will have a greater score. Table 1 outlines the distances measured, and Fig. 1 delineates them on a facial image.

2.1.2.3. Femininity. In order to compute facial femininity, a subset of nine Euclidean distances was taken from the full set used to compute averageness. These measurements were found to differ significantly between female and male faces (Kościński, 2012), indicating sexual dimorphism. The femininity index was computed by summing the *Z*-scores for each participant across the nine sexually dimorphic distances, and multiplying by −1. Table 1 describes the measurements used for the femininity index.

2.1.2.4. Symmetry. Horizontal facial symmetry was calculated according to validated measures used in previous research (Little et al., 2008; Penton-Voak et al., 2001; Scheib et al., 1999). The midpoint between seven pairs of bilateral points (described in Table 1 and Fig. 1) was calculated by averaging the coordinates of each pair of bilateral points. If a face is perfectly symmetrical, all points should fall along the same midline. The absolute difference between all possible combinations of midpoints was then summed to provide a measure of symmetry for each face, with higher numbers reflecting greater asymmetry.

2.1.2.5. Colouration. Using MATLAB, facial photographs were converted from RGB representation to CIEL*a*b*, a perceptually uniform colour space approximating how the human eye perceives colour (Weatherall & Coombs, 1992). Using shape information provided by the 160 landmarks, the entirety of facial skin regions, excluding the eyebrows, eyes, and lips, was extracted and split into luminance (light to dark; L^*), alpha (red to green; a^*) and beta (yellow to blue; b^*) channels. The pixel values of these three regions, consisting solely of facial skin, were then averaged to provide a measure of skin luminance, redness, and yellowness.

The Python and MATLAB code used to extract the shape and colour information, as well as associated template files, are available from the OSF.

Table 1
Facial measurements and shape indices they comprise.

Measurement	Measurement used to calculate indices		
Interpupillary distance (0)			Symmetry
Inner interocular distance (1)	Averageness		Symmetry
Outer interocular distance (2)			Symmetry
Eye height (3)	Averageness	Femininity	
Brow height (4)	Averageness		
Nose length (5)	Averageness		
Nose width (6)	Averageness	Femininity	Symmetry
Mouth width (7)	Averageness		Symmetry
Upper lip thickness (8)	Averageness		
Lower lip thickness (9)	Averageness		
Mouth height (10) - *Sum of* (8, 9)	Averageness		
Bizygomatic width (11)	Averageness		Symmetry
Jaw width (12)	Averageness	Femininity	Symmetry
Nose-mouth distance (13)	Averageness	Femininity	
Chin height (14)	Averageness	Femininity	
Facial height (15)	Averageness	Femininity	
Mouth location - *Ratio of* (13, 14)	Averageness		
Upper lip to lower lip - *Ratio of* (8, 9)	Averageness		
Mouth width to jaw width - *Ratio of* (7, 12)	Averageness	Femininity	
Nose width to mouth width - *Ratio of* (6, 7)	Averageness	Femininity	

Note. Numbers in parentheses denote the measurement numbers in Fig. 1.

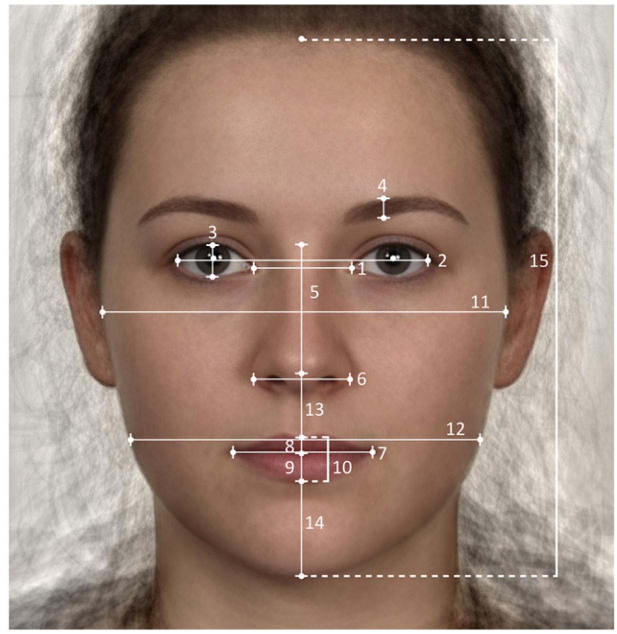

Fig. 1. The Euclidean distances used to compute shape information from faces. Interpupillary distance is not shown, but measures the distance between points on both pupils.

2.1.3. Observers

Twenty-two observers (age $M = 23.14$, $SD = 6.96$, 15 females) rated all 108 faces for perceived health through an online testing platform (www.testable.org). Faces appeared in a random order, and were cropped to just below the chin and above the hairline, and to the widest point of the face. Observers were asked 'how healthy is this face?', and indicated their responses on a one (*very unhealthy*) to seven (*very healthy*) scale, via a mouse click. Observers were recruited through social media websites.

2.1.4. Procedure

After participants completed the SF-12 questionnaire and were photographed, their images were delineated and subjected to shape (averageness, femininity, and symmetry) and colour (CIEL*a*b*) analysis. Observer ratings of perceived health were averaged across observers to form a perceptual measure of health for each face. The above variables were then subjected to a Brunswik lens model (Brunswik, 1956) analysis. First, relationships between perceived health and shape and colour variables were examined (*cue utilisation*) – that is, what facial properties are linked to observer ratings of health? Second, relationships between facial properties and measured general health from the SF-12 (*cue validity*) were also tested, revealing whether certain facial properties are linked to actual measures of health. This model describes how an observer might look for an environmental cue (such as skin colour) to infer a trait that is not directly perceptible (such as health), using the cue to make a judgment about an underlying trait. Observers are accurate judges if they utilise cues that are valid indicators, and ignore invalid cues that are misleading. This model has been used successfully to examine cues related to perception of personality from static images (Naumann et al., 2009), health in faces from expression and attractiveness (Zebrowitz et al., 2014), and an occupants personality from a room's environment (Gosling et al., 2002). For evolutionary approaches to facial health perception, this model is uniquely suited to demonstrate what long-term and short-term cues are both utilised, as well as valid.

2.2. Results

The ratings assigned by observers showed good internal consistency, with Cronbach's $\alpha = 0.89$, 95% CI [0.86, 0.92]. Ratings were averaged

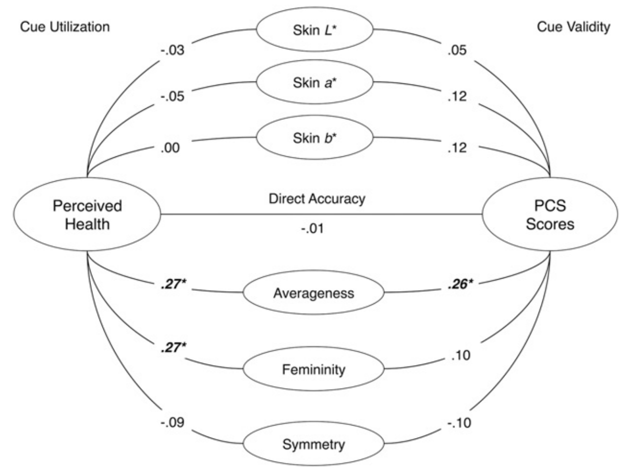

Fig. 2. Brunswik lens model describing the relationship between perceived health, underlying PCS scores, and short-term (skin colour) and long-term (facial shape) cues. Pathways on the left describe cue utilisation, and cue validity on the right. Values represent Pearson's *r*. Significant correlations are in bold. All *ps* < 0.008. Utilisation $df = 106$, Validity $df = 101$.

across observers to provide a mean score for each face. This composite score represents a measure that is independent of idiosyncrasies in individual ratings, providing a reliable measure of what health information can be detected from an unacquainted face. There was no direct relationship with perceived health and underlying PCS scores, $p = 0.909$, indicating that health, as represented by PCS scores, could not be directly observed. However, the short and long term health cues provided further information, and indicate the usefulness of a lens model approach here. The full Brunswik lens analysis, with utilisation and validity pathways, is shown in Fig. 2.

2.2.1. Cue utilisation

Pathways on the left of Fig. 2 illustrate cue utilisation. That is, how does the average rating of health assigned by observers correlate with the different cues? For this analysis, all 108 faces were used, given that shape, colour, and perceived health ratings were available. Short-term health cues in the form of skin colouration showed no utilisation, with very weak correlations between perceived health and all three colour channels, all $ps > 0.636$. For long-term health cues in the form of shape variables, both averageness, $r(106) = 0.27$ [0.09, 0.44], $p = 0.008$, and femininity, $r(106) = 0.27$ [0.08, 0.43], $p = 0.005$, showed a relationship, indicating healthier looking female faces tend to be more average and feminine to a similar degree.

2.2.2. Single observer utilisation

The above utilisation correlations represent the relationships between shape and colour cues in faces, and the average health rating assigned by observers. Here, an alternative analysis is conducted at the level of the individual observer, to examine what cues the average observer might use when judging health, and whether they are accurate. This is an important question, since correlations with averaged ratings, which treat the sample averages as population estimates (as above), can often diverge greatly from individual responses (Monin & Oppenheimer, 2005). They also do not speak to psychological processes that might underpin how the average observer uses facial cues to judge health, or their accuracy (Zebrowitz et al., 2014). Both kinds of analysis are typical when examining cue utilisation in lens models (Naumann et al., 2009; Zebrowitz et al., 2014). To do this, the ratings provided by each individual observer were correlated with the colour and shape cues, as well as the PCS scores, for each face. Each observer then had a set of correlations, indicating how much their ratings of health shared variance with the available cues. These correlations were then transformed using Fisher's *r*-to-*z* formula, and analysed using a one-sample *t*-test against zero, determining whether the cue was utilised beyond what would be expected by chance.

Across observers, the average correlation between health ratings and PCS scores was not different from zero ($M = -0.01$ [-0.04, 0.02]), $t(21) = 0.56$, $p = 0.583$, $d = 0.11$, reflecting the full analysis shown in Fig. 2. For colour cues, there was also no evidence of utilisation, for luminance ($M = 0.00$ [-0.05, 0.05]), $t(21) = 0.06$, $p = 0.950$, $d = 0.01$, redness ($M = -0.02$ [-0.05, 0.00]), $t(21) = 1.98$, $p = 0.060$, $d = 0.42$, or yellowness ($M = 0.00$ [-0.04, 0.04]), $t(21) = 0.05$, $p = 0.961$, $d = 0.01$. For shape cues, observers showed evidence of significant utilisation for averageness ($M = 0.17$ [0.13, 0.21]), $t(21) = 8.92$, $p < 0.001$, $d = 1.90$, as well as femininity ($M = 0.16$ [0.11, 0.21]), $t(21) = 6.48$, $p < 0.001$, $d = 1.38$. Interestingly, there was also weaker evidence of utilisation of symmetry ($M = -0.06$ [-0.10, -0.01]), $t(21) = 2.78$, $p = 0.011$, $d = 0.59$, indicating observers assigned higher ratings of health to more symmetrical faces.

2.2.3. Cue validity

Though cue validity correlations were generally larger than utilisation correlations, only averageness showed a significant relationship with PCS scores, $r(101) = 0.26$ [0.07, 0.43], $p = 0.008$, indicating individuals with more average faces had better health in their daily life currently, and in the past month.

2.3. Discussion

A Brunswik lens model analysis is able to describe the utilisation and validity of cues when making a judgment about underlying traits. Here, results suggested utilisation of facial appearance when judging health is restricted to a single class of cues: facial shape. Faces with higher femininity and averageness were perceived as healthier, while individual observer utilisation showed an additional though weaker effect of symmetry utilisation. These results offered partial support of initial hypotheses. Colour cues showed no utilisation, which was not predicted given previous work (Henderson et al., 2016). Were observers accurate in their judgments? While no evidence of direct accuracy was found, there was evidence for validity for averageness only. That is, more average faces *look* healthier, and individuals with more average faces report better health in general, less physical pain, and are less likely to have their health interfere with daily tasks and goals both at the time of responding and during the previous month.

These results support established findings regarding the utilisation and validity of facial averageness as a cue to certain measures of health (Rhodes et al., 2001; Zebrowitz & Rhodes, 2004). They also support the finding that femininity is used as a cue to health (Foo et al., 2017), but that evidence for an actual link to many health measures is mixed (Foo et al., 2017; Gray & Boothroyd, 2012). Femininity may appear healthy because it is intrinsically linked to female attractiveness, which itself has been shown to impede perceptions of health (Kalick, Zebrowitz, Langlois, & Johnson, 1998). In addition, symmetry showed only very weak utilisation at the observer level, and no validity (Pound et al., 2014).

That colour cues showed no utilisation or validity is surprising, but also supports newer evidence indicating the role of skin colour in perceiving health and attractiveness may be much smaller than previously thought (Foo et al., 2017; Mogilski & Welling, 2017). However, validity correlations between colour and actual health were not as small as utilisation correlations. Therefore, there may be a link between colour and actual health (which manipulation studies readily reveal), but it requires a sample size that is significantly larger to detect using natural faces. For example, from the above data, there is an average cue validity correlation of $r_{Mean} = 0.09$ between colour cues and health. To detect an effect present at this size would require $n = 779$, assuming standard parameters of power and alpha. Conversely, slight differences in methodology may explain the absence of colour utilisation here. When whole face colour is manipulated, it seems to produce reliable changes in perceptions of health (Lefevre & Perrett, 2015; Stephen et al., 2011; Stephen et al., 2009b, 2009). However, in studies where it is measured and related to perceived health, it is often the average of a series of cropped skin patches, typically derived from the cheeks and forehead (Foo et al., 2017; Henderson et al., 2016). Recent research has found that different colouration, in different regions of the face (particularly the cheeks), communicates separate and dissociable variance to health judgments (Jones et al., 2016). It may be that while a manipulation of colour affects health judgments, it does so because it alters colour in the specific areas that contribute to perceived health. Second, studies showing a correlation between skin colouration and health may have done so because they measured colour in the specific areas responsible for signalling aspects of health. The method here measured skin colour from the entirety of facial skin, in line with the claims of the current literature, but showed no utilisation. Future research is needed to disentangle whether whole face colouration, or just specific regions, is responsible for a face looking healthy.

The results from this study suggest that perceptions of female facial health utilise indices of long-term health condition in the form of facial shape, and that one of the utilised cues is related to aspects of health measured by the SF-12 – facial averageness.

3. Study Two

The results of Study One indicate that facial averageness is a valid and utilised cue to health, falling within a class of cues related to long-term health condition. Conversely, facial colouration seems to show little validity and almost no utilisation, counter to previous research. One reason for this pattern of findings is that colour channels were examined in isolation. Work on perceived health and skin colour has often shown the specific combination of luminance, redness, and yellowness associated with dietary carotenoid intake reliably increases perceived health (Lefevre & Perrett, 2015; Tan et al., 2017; Whitehead et al., 2012), and the increased carotenoids themselves should confer current health benefits (Rao & Rao, 2007). It may be that it is the specific combination of colours that is important for cue utilisation and validity, and not the separate levels of colour themselves. Alternatively, the measure of health used in Study One (PCS from the SF-12) is perhaps unable to capture health variance that is related to behaviours directly related to skin colour – for example, a very recent infection, or a respondent missing a night's sleep before completing the questionnaire (Henderson et al., 2017; Sundelin et al., 2017).

Study Two aims to address these shortcomings. Here, faces are manipulated to both increase and decrease their averageness, as well as their carotenoid colouration. With these manipulations, it is possible to estimate how much each cue class contributes to judgments of health by comparing ratings of faces within different levels of each cue class. Given that these cue classes represent different kinds of health status (long-term health in the form of averageness; current health in the form of carotenoid colouration Getty, 2002), it is predicted that there may be an interaction between both cue classes. For example, a face with low averageness and high carotenoid colouration may be seen as equally healthy as a face with high averageness and low carotenoid colouration, given the balance between current and long-term health condition. This echoes previous findings where adiposity and colour interact (Fisher et al., 2014), where undesirable higher or lower levels of adiposity are 'offset' by healthy colouration.

3.1. Method

3.1.1. Participants

The same facial images of the participants used in Study 1 were used here.

3.1.1.1. Design sensitivity. For a repeated measures ANOVA with 108 participants, $\alpha = 0.05$, and 80% power, the design is sufficiently sensitive to detect an eta squared of $\eta^2 = 0.01$, a very small effect size.

3.1.2. Stimuli

Using JPsychomorph, the faces of all participants were manipulated along both a shape and colour axes to produce four versions of each face. Faces were increased and decreased along averageness and carotenoid colouration dimensions so that short-term and long-term cues to health were matched (e.g. high averageness, high carotenoid) or crossed (e.g. high averageness, low carotenoid). An example is shown in Fig. 3.

3.1.2.1. Averageness. To manipulate averageness, the faces of all 108 models were aggregated to produce the average facial shape. Then, the face of each participant was transformed along the linear shape difference between the participant and the average face by ±50%, i.e. applying half of the difference between the participant and the average both towards *and* away from the average. This produced a version of each face with a more average and a more distinctive facial appearance.

3.1.2.2. Carotenoid colouration. To alter carotenoid colouration, the averageness-manipulated faces were altered along all three CIEL*a*b* axes. High carotenoid colouration involved subtracting -1.1 L^* units, and adding $+1.4$ a^* units, and $+4.35$ b^* units to faces, as used elsewhere (Lefevre & Perrett, 2015). Low carotenoid colouration was simulated using the reverse alteration. This change was achieved by creating two colour patches in MATLAB representing the above values (± 1.1 L^*; ± 1.4 a^*; ± 4.35 b^*). These patches were then conformed to the average

Fig. 3. An example of the manipulations used. Rows indicate carotenoid colouration manipulation (high on left), and columns indicate averageness manipulation (more average on top). To protect individual identity, the example stimulus is composite of two faces from the set of participants.

face shape, and had eyebrows, eyes, and lips removed so as not to be affected by the colour change. That is, those features were held constant. JPsychomorph was used to transform the colour of each face ±100% of the linear difference between the patches, creating the four versions of each model. The MATLAB code used to create these colour transforms is available from OSF.

3.1.3. Observers

One hundred and ninety one (age $M = 23.80$, $SD = 11.09$, 157 females) observers rated faces for perceived health through an online testing platform (www.testable.org). Observers were divided randomly into four groups ($ns = 49, 45, 50$, and 47) who saw different 'line ups' of the manipulated faces, detailed below. Faces appeared in a random order and were cropped in the same way as before. Observers were asked 'how healthy is this face?', and indicated their responses on a one (*very unhealthy*) to seven (*very healthy*) scale, via a mouse click. Observers were recruited through a mixture of social media and for course credit. A subsample of those observers ($n = 32$) completed their ratings on a lab computer with a monitor, to check if variances in monitor appearances may affect results, though this has not been demonstrated elsewhere (Foo et al., 2017; Lefevre, Ewbank, Calder, Hagen, & Perrett, 2013; Lefevre & Perrett, 2015).

3.1.4. Procedure

Previous research examining the interaction between cues in faces has typically involved fully within-subjects designs with a relatively small sample of faces. For example, observers may rate all manipulations of a small set of identities in a single session (Fisher et al., 2014; Mogilski & Welling, 2017), be presented with individual identities with different levels of manipulation in forced choice designs (Smith et al., 2009), or utilise continuum manipulations allowing observers to select an optimal appearance from the full range of cue interactions (Carrito et al., 2016). While informative, designs such as this can suffer from carryover effects. An observer viewing a face with an unhealthy colouration, when seeing the same face with healthy colouration some trials later, may offer an inflated increase in perceived health (for discussion of these issues, see (Jones & Kramer, 2015; Morrison, Morris, & Bard, 2013). To avoid this, four line-ups were created that contained all 108 participants, but under different cue combinations. For example, in line up A, participant 1 appeared with a high carotenoid, high averageness appearance, in line up B, with a high carotenoid, low averageness appearance, and so on. Across the four line-ups, the proportion of participants appearing with each combination of cues was balanced, with 27 participants in each combination. Using this method, there are no possibilities of carryover effects, and each participant receives a rating that is dependent entirely on the cue combination they present with. This is a more conservative, but ecologically valid approach, appropriating how faces are judged in the real world – once, and with a fixed combination of cue classes.

After data collection, ratings were averaged across observers and line-ups for each participant, resulting in each receiving four mean scores representing their perceived health under the combinations of shape and colour cues.

3.2. Results

The four scores per participant were submitted to a 2 (Colouration: High carotenoid, low carotenoid) × 2 (Averageness: high, low) repeated measures ANOVA, using participants as the unit of analysis. Eta squared was used as a measure of effect size, directly explaining the variance in perceived health attributable to each factor. Faces were perceived as healthier with high averageness ($M = 4.26$ [4.14, 4.38]) compared to low averageness ($M = 3.37$ [3.27, 3.48]), $F(1, 107) = 605.47$, $p < 0.001$, $\eta^2 = 0.34$. Similarly, faces received higher health ratings with high carotenoid colouration ($M = 3.94$ [3.83, 4.06]) compared to low ($M = 3.69$ [3.58, 3.80]), $F(1, 107) = 52.87$, $p < 0.001$, $\eta^2 = 0.04$. However, there was no evidence of any integration between averageness and colouration, with no interaction $F(1, 107) = 0.20$, $p = 0.652$, $\eta^2 = 0.00$.

3.2.1. Subsample analysis

One reason for the low variance explained by colour could be down to observers completing the experiment across different monitors with different colour profiles. This could lead to colour cues being 'washed out' by this variation. As such, the above analysis was repeated, but only using ratings obtained from observers that had completed the experiment using a single monitor, holding colour profile constant. The same pattern of results emerged, with main effects of Shape, $F(1, 107) = 101.82$, $p < 0.001$, $\eta^2 = 0.19$, Colouration, $F(1, 107) = 8.72$, $p = 0.004$, $\eta^2 = 0.01$, and no interaction, $F(1, 107) = 0.00$, $p = 0.979$, $\eta^2 = 0.00$, confirming previous results and that facial colouration research is suitable for online testing (Lefevre et al., 2013).

3.2.2. Variance confidence estimates

Given that shape explained a much larger proportion of variance than colour, it is important to establish confidence intervals around this somewhat surprising finding. If the study were repeated, how much variance would likely be attributable to both shape and colour? In order to estimate this, *R* was used to conduct a bootstrap resampling procedure. The original data was randomly sampled with replacement 5000 times, and the repeated measures ANOVA calculated at every iteration. Eta squared was derived for both main effects of and the interaction. The bootstrap procedure revealed that confidence intervals differed widely between Shape ($\eta^2_{Mean} = 0.34$ [0.28, 0.40]) and Colour ($\eta^2_{Mean} = 0.04$ [0.02, 0.06]) factors, but that the effect size estimates were highly accurate for the main analysis. The effect size of the interaction remained extremely close to zero across resamples ($\eta^2_{Mean} = 0.00$ [0.00, 0.00]). The *R* code used for this procedure is available from the OSF.

3.3. Discussion

Study Two contrasted the contributions of facial averageness and carotenoid colouration to perceived health. Surprisingly, there was no interaction between the cues, but two main effects of shape and colour. These results support the distinct contribution cue classes make to health perception (Getty, 2002; Smith et al., 2009), but also suggest that there is no interaction between these cue classes, which run counter to previous research (Fisher et al., 2014; Smith et al., 2009). However, it is worth noting that previous work has examined adiposity (Fisher et al., 2014) which is not truly a fixed cue to developmental health as it can change over time, and femininity (Smith et al., 2009), which does not seem a valid cue to health, as indicated here and elsewhere (Foo et al., 2017). Conversely, averageness *is* related to actual health, and does not seem to interact with carotenoid colouration. This is surprising, given that a clear prediction would be that current condition could offset a cue to poor previous condition, or the converse. Alternatively, observers may have found faces with low averageness and low carotenoid colouration particularly aversive, or faces with high levels of both cues as extremely healthy. It is also worth noting that carotenoid colouration seems to be preferred only up to a point, at least in Asian faces (Tan et al., 2017), and that this amount of carotenoid change was on average less than the more standard manipulation used here and elsewhere (Lefevre & Perrett, 2015; Whitehead et al., 2012). Though no optimal level of data exists in Caucasian samples currently, it may be the smaller effect size observed here is due to carotenoid levels being increased to levels beyond what is optimal.

The main finding from this study suggests that the contributions these cue classes make are separate, but that averageness explains a significantly higher proportion of variance in health perception compared to carotenoid colouration, which is relatively small (8.5 times smaller). These findings indicate that health perception is weighted towards cues to long-term condition, rather than current condition, and complement the findings of Study One, which showed that averageness is a valid and utilised cue to a measure of health beyond colouration.

4. General discussion

Using an generalizable and ecologically-valid model, Study One demonstrated that observers exclusively utilise shape cue classes to make judgments of health, and that at least one of these cues, facial averageness, is a valid cue to a measure that assesses several aspects of health in daily living. Study Two demonstrated that when averageness is varied along with a colour cue reliably linked to perceived health - carotenoid colouration - facial averageness explains a far larger proportion of variance in perceived health than carotenoid colouration. Moreover, the two cue classes showed no evidence of an interaction, indicating reliably separable contributions to perceived health.

These results have interesting theoretical contributions, as they suggest that health perception is tied to a class of cues that are linked to indicators of previous health, but rely only very little on cue classes tied to current condition. Why might this be? A tentative suggestion comes from a cross-examination of modern and historical datasets that seem to point to evolved mechanisms that weight health perception to cue classes of previous condition. Consider that facial averageness seems to have a link to genetic diversity and immune response (Lie et al., 2008), but also seems to have a significant environmental cause (Lee et al., 2016). Averageness also seems related to childhood health (Rhodes et al., 2001; Zebrowitz & Rhodes, 2004), when perturbations in appearance due to the environment begin to appear (Özener & Fink, 2010). Also consider that historically, a human lifespan has been significantly shorter than is typical today. Data from pre-industrial populations in Europe (Finch & Crimmins, 2004; Oeppen & Vaupel, 2002), and hunter-gatherer societies in the 20th century (Gurven & Kaplan, 2007), indicate that mortality in childhood was approximately 10%–30%, and of those who survived past 20, the average life expectancy was around 40 years of age (Finch, 2010). Unsurprisingly, the leading cause of death was infectious disease in these populations (Gurven, Kaplan, & Supa, 2007). These datasets give clues to the kinds of highly infectious environments early humans evolved under, but also suggest that survivors of these dangerous early infections suffered earlier immunosenescence – the depletion of memory T-cells that allow effective immune response to new diseases (Caspari & Lee, 2006; Crimmins & Finch, 2006; Finch, 2010). Moreover, the increased inflammatory loads they experienced likely promoted chronic illnesses, like cardiovascular disease (Finch, 2010). Crucially, the propensity for these diseases are passed on as a direct consequence of the infection load mothers experienced (Mazumder, Almond, Park, Crimmins, & Finch, 2010). A final and important point from these historical population data is the suggestion that the average age of reproduction was around 20, the age at which those who had experienced earlier illness had survived (Kaplan, Hill, Lancaster, & Hurtado, 2000). Facial health cues are vitally important in attractiveness judgments (Rhodes et al., 2007), and it may be that ancestral generations did well to attend to cues to previous condition, given that these can have direct consequence for offspring. Additionally, an individual with a high inflammatory load could still be 'currently' healthy, but biological damage from early infection would still be present and affect later life or future offspring. This mirrors the findings of Study Two, where shape and colour both influenced health, but shape was a stronger predictor.

The field of psychology is moving towards a new way of describing results, with a focus on meta-analytic thinking and an understanding of the size of effects, rather than their significance (Cumming, 2011). These findings take a step in this direction with a focus on the magnitude of effects, obtained from more ecologically valid designs. In addition, they also complement other findings that examine the effect size of surface texture features in judging attractiveness. Facial cosmetics increase facial contrast (Jones, Russell, & Ward, 2015), a colour cue to health (Russell et al., 2016), age (Porcheron, Mauger, & Russell, 2013), and attractiveness (Russell, 2003). However, despite well documented effects of cosmetics on increased facial attractiveness (Etcoff, Stock, Haley, Vickery, & House, 2011; Nash, Fieldman, Hussey, Lévêque, & Pineau, 2006), the overall effect size is very small, for both an everyday sample of women (Jones & Kramer, 2015), and even professional models (Jones & Kramer, 2016). Taken in context, these findings suggest that even a dramatic change to colour cues does not account for a large proportion of variance in judgments relevant to mate choice.

This does not indicate that colour cues are unimportant, or affect perceptions so little they are not worth future study. The effects of carotenoid colouration, as a cue to current condition, seems reliably preferred across different ethnicities (Stephen et al., 2011; Tan et al., 2017) even in cultures where melanin and tanning colouration are highly regarded (Pezdirc et al., 2017). Rather, there may be other contexts that drive the importance of current-condition cues. Individuals who are high in perceived vulnerability to disease demonstrate a stronger preference towards faces that appear healthy (Welling, Conway, Debruine, & Jones, 2007). While little is known about the conditions which cause individuals to perceive themselves as vulnerable (Tybur & Gangestad, 2011), these individuals may value current condition cue classes more than others.

There may also be cultural factors that influence the weighting given to certain cue classes. For example, African observers seem to be more sensitive to cues to colouration than shape when making judgments of attractiveness and health, whereas European observers seem to rely more on shape cues (Coetzee, Greeff, Stephen, & Perrett, 2014; Kleisner et al., in press). There are two explanations for this – one might be simply that perceptual experience with faces outside of what is seen typically means health cues in the shape of colour are not readily available until more experience with colouration is acquired. Alternatively, it may reflect a reliance on certain cue classes depending on the pathogenic load of the environment – in places where current infection is more dangerous, it may pay observers to attend closely to cues signalling that information. However, these findings are at odds with research on masculinity preferences, a putative long-term health cue, that seems to increase with decreasing national health outcomes (DeBruine, Jones, Crawford, Welling, & Little, 2010). Given this, it is likely that complex interactions between environmental conditions and observer reliance on various cue classes exist.

Additional evidence suggests that priming individuals with infection and disease relevant stimuli can cause individuals to value cues to health more strongly, a domain in which current condition, signalled via colour, would be well-suited (Little, DeBruine, & Jones, 2011). Finally, the kind of mate an individual desires could affect the cue class they rely on. Individuals seeking a short term encounter surprisingly seem to place more weight on shape cues, rather than colour (Mogilski & Welling, 2017), perhaps willing to accept a possible infection for the indirect benefits of a more robust mate. Conversely, those desiring a long-term mate seem to prefer both sexual dimorphism as well as healthy colour (Mogilski & Welling, 2017), perhaps reflecting a desire to acquire a mate with good current health and the propensity to be healthy (Tybur & Gangestad, 2011). These kinds of contexts were not considered in the data presented here, but may well reveal more nuanced utilisation of cues that are tied to valid aspects of health.

While the methods employed in this set of studies set out to have greater ecological validity than previous studies, it is important to note the generalizability of the results is limited to a sample of young adults, who are generally healthy. This is an important limitation, because accurate perception of health may change with face age and reflect a different utility for individuals outside of reproductive age. For example, in adults over the age of 40, facial appearance might reflect long and short-term health in a variety of ways, as genetic health and lifestyle factors have become 'visible' on the face. The sample used here has very little variance in age. After the age of 40, individuals with the same chronological age can look on average six years younger or older than one another, and this difference in aging speed is reflected by health indicators in the blood that are related to genetic disposition and environmental choices, such as cholesterol (Chen et al., 2015). Additionally, identical twins who have unhealthy lifestyle habits (such as smoking or excessive sun exposure) look older than their siblings and suffer worse health (Guyuron et al., 2009). Given this, it is likely that health cues in older faces are not so easily cast into short or long term cue classes, but present as a complex interaction of genetic and environmental factors. Indeed, accuracy is higher for older faces than younger faces and, where a lens model is applied, cues only show validity in older adults (Zebrowitz et al., 2014). However, the cue classes observers utilise, and which are valid, are not clear for older adults beyond attractiveness and expression (Zebrowitz et al., 2014).

However, it is worth mentioning that some specific aspects of facial appearance, such as nasion-chin distance and lip thickness, seem to change with age and are related to health indicators in blood (Chen et al., 2015). These measures and others are both captured in the measure of averageness and femininity used here. In addition, individuals predisposed to faster aging through genetic factors also appear older even in their youth (Belsky et al., 2015), and so there is some evidence to suggest the cues examined here may be relevant still in older adults. However, the current findings can only be clearly generalised to younger adult faces.

There is now a growing body of work that examines the links between facial appearance and health, and the accuracy of these judgments (Foo et al., 2017; Gray & Boothroyd, 2012; Pound et al., 2014; Rantala et al., 2013; Rhodes et al., 2001; Zebrowitz et al., 2014; Zebrowitz & Rhodes, 2004). However, the relationships between cues and underlying health are not consistently clear. This is important, since many assumptions of mate choice in human sexual selection rest on significance of validity correlations between facial appearance and underlying health. Part of the difficulty in examining this relationship is measuring the multidimensional construct of health (Eberst, 1984). What is clear is that facial appearance is related to some dimensions of health, but not others. For example, physician assessments of health do correlate with some aspects of facial appearance in either adulthood or adolescence (Zebrowitz & Rhodes, 2004), but evidence is limited in older adults (Zebrowitz et al., 2014). When health is assessed in terms of immune function or oxidative stress, there appears to be no relationship between cues like averageness and colouration (Foo et al., 2017). The measure of health used here, derived from a medical questionnaire designed to assess general and previous functioning over the past month, did show a positive relationship with a set of facial traits. The utilisation of those traits was reflected in the findings of Study Two. However, the results of this study should be considered against the wider literature and the kinds of ways health shows an association with facial appearance. Indeed, the ways in which appearance and health are both related and *not* related, and the evolutionary significance of those relationships, is a vital question. More recent evidence suggests attributes like a positive facial expression can account for large amounts of variation in perceived health, but relationships with underlying health are likely extremely complex (Henderson et al., 2016; Jones et al., 2017).

In summary, this work shows that aspects of facial appearance are valid cues to health, and that observers utilise them to make judgments of health. Several lines of evidence to converge to suggest that human health perception seems calibrated to the valid use of cues to long-term and previous health in the form of facial shape, which may reflect the highly infectious environments humans evolved under. The current data also suggests the utilisation of colour cues, as a proxy for immediate, current health, is more complex than previously thought, and also suggests the kinds of health measures that are used may impact any relationship between appearance and health.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest

None.

Data availability

The data associated with this research, including the scripts for analysis and stimulus generation, are available at the OSF: https://osf.io/azhcb/.

Acknowledgements

I would like to thank Kate Sullivan, Lauren O'Dwyer, Molly Walker, and Ashley Holloway for their hard work during data collection, and two anonymous reviewers and the editor Professor Ben Jones for their constructive comments that have improved the work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2017.09.005.

References

Adamo, S. A., & Spiteri, R. J. (2005). Female choice for male immunocompetence: when is it worth it? *Behavioral Ecology*, *16*(5), 871–879. https://doi.org/10.1093/beheco/ari068.

Adamo, S. A., & Spiteri, R. J. (2009). He's healthy, but will he survive the plague? Possible constraints on mate choice for disease resistance. *Animal Behaviour*, *77*(1), 67–78. https://doi.org/10.1016/j.anbehav.2008.09.011.

Alexander, M., Newmark, H., & Miller, R. G. (1985). Oral beta-carotene can increase the number of OKT4+ cells in human blood. *Immunology Letters*, *9*(4), 221–224.

Axelsson, J., Sundelin, T., Ingre, M., Van Someren, E. J. W., Olsson, A., & Lekander, M. (2010). Beauty sleep: experimental study on the perceived health and attractiveness of sleep deprived people. *BMJ (Clinical Research Ed.)*, *341*, c6614.

Baudouin, J. -Y., & Tiberghien, G. (2004). Symmetry, averageness, and feature size in the facial attractiveness of women. *Acta Psychologica*, *117*(3), 313–332. https://doi.org/10.1016/j.actpsy.2004.07.002.

Belsky, D. W., Caspi, A., Houts, R., Cohen, H. J., Corcoran, D. L., Danese, A., ... Moffitt, T. E. (2015). Quantification of biological aging in young adults. *Proceedings of the National Academy of Sciences*, *112*(30), E4104–E4110. https://doi.org/10.1073/pnas.1506264112.

Bridges, L. (2012). *Face Reading in Chinese Medicine* (2 edition). Edinburgh: Churchill Livingstone.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed). Berkeley: University of California Press.

Burriss, R. P., Little, A. C., & Nelson, E. C. (2007). 2D:4D and sexually dimorphic facial characteristics. *Archives of Sexual Behavior*, *36*(3), 377–384. https://doi.org/10.1007/s10508-006-9136-1.

Carrito, M. d. L., Santos, I. M. B. d., Lefevre, C. E., Whitehead, R. D., Silva, C. F. d., & Perrett, D. I. (2016). The role of sexually dimorphic skin colour and shape in attractiveness of male faces. *Evolution and Human Behavior*, *37*(2), 125–133. https://doi.org/10.1016/j.evolhumbehav.2015.09.006.

Caspari, R., & Lee, S. -H. (2006). Is human longevity a consequence of cultural change or modern biology? *American Journal of Physical Anthropology*, *129*(4), 512–517. https://doi.org/10.1002/ajpa.20360.

Chang, F., & Chou, C. -H. (2009). A bi-prototype theory of facial attractiveness. *Neural Computation*, *21*(3), 890–910. https://doi.org/10.1162/neco.2008.07-07-566.

Chen, W., Qian, W., Wu, G., Chen, W., Xian, B., Chen, X., ... Han, J. -D. J. (2015). Three-dimensional human facial morphologies as robust aging markers. *Cell Research*, *25*(5), 574–587. https://doi.org/10.1038/cr.2015.36.

Cho, Y. H., Jeong, D. W., Seo, S. H., Lee, S. Y., Choi, E. J., Kim, Y. J., ... Chin, H. W. (2012). Changes in skin color after smoking cessation. *Korean Journal of Family Medicine*, *33*(2), 105–109. https://doi.org/10.4082/kjfm.2012.33.2.105.

Coetzee, V., Perrett, D. I., & Stephen, I. D. (2009). Facial adiposity: A cue to health? *Perception*, *38*(11), 1700–1711. https://doi.org/10.1068/p6423.

Coetzee, V., Greeff, J. M., Stephen, I. D., & Perrett, D. I. (2014). Cross-cultural agreement in facial attractiveness preferences: The role of ethnicity and gender. *PLoS One*, *9*(7), e99629. https://doi.org/10.1371/journal.pone.0099629.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Crimmins, E. M., & Finch, C. E. (2006). Infection, inflammation, height, and longevity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 498–503. https://doi.org/10.1073/pnas.0501470103.

Cumming, G. (2011). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Cunningham, M. R. (1986). Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. *Journal of Personality and Social Psychology*, *50*(5), 925–935. https://doi.org/10.1037/0022-3514.50.5.925.

DeBruine, L. M., Jones, B. C., Crawford, J. R., Welling, L. L. M., & Little, A. C. (2010). The health of a nation predicts their mate preferences: cross-cultural variation in women's preferences for masculinized male faces. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1692), 2405–2410. https://doi.org/10.1098/rspb.2009.2184.

Dowling, D. K., & Simmons, L. W. (2009). Reactive oxygen species as universal constraints in life-history evolution. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1663), 1737–1745. https://doi.org/10.1098/rspb.2008.1791.

Eberst, R. M. (1984). Defining health: a multidimensional model. *The Journal of School Health*, *54*(3), 99–104.

Etcoff, N., Stock, S., Haley, L. E., Vickery, S. A., & House, D. M. (2011). Cosmetics as a feature of the extended human phenotype: modulation of the perception of biologically important facial signals. *PLoS One*, *6*(10), e25656. https://doi.org/10.1371/journal.pone.0025656.

Farkas, L. G. (1994). Examination. In L. G. Farkas (Ed.), *Anthropometry of the head and face* (pp. 3–56). New York: Raven Press.

Faul, F., Erdfelder, E., Lang, A. -G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146.

Finch, C. E. (2010). Evolution of the human lifespan and diseases of aging: Roles of infection, inflammation, and nutrition. *Proceedings of the National Academy of Sciences*, *107*(Suppl. 1), 1718–1724. https://doi.org/10.1073/pnas.0909606106.

Finch, C. E., & Crimmins, E. M. (2004). Inflammatory exposure and historical changes in human life-spans. *Science*, *305*(5691), 1736–1739. https://doi.org/10.1126/science.1092556.

Fisher, C. I., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2014). Integrating shape cues of adiposity and color information when judging facial health and attractiveness. *Perception*, *43*(6), 499–508. https://doi.org/10.1068/p7728.

Foo, Y. Z., Simmons, L. W., & Rhodes, G. (2017). Predictors of facial attractiveness and health in humans. *Scientific Reports*, *7*. https://doi.org/10.1038/srep39731.

Getty, T. (2002). Signaling health versus parasites. *The American Naturalist*, *159*(4), 363–371. https://doi.org/10.1086/338992.

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, *82*(3), 379–398.

Grandfield, T. A., Thompson, A. R., & Turpin, G. (2005). An attitudinal study of responses to a range of dermatological conditions using the implicit association test. *Journal of Health Psychology*, *10*(6), 821–829. https://doi.org/10.1177/1359105305057316.

Gray, A. W., & Boothroyd, L. G. (2012). Female facial appearance and health. *Evolutionary Psychology*, *10*(1), 66–77. https://doi.org/10.1177/147470491201000108.

Gurven, M., & Kaplan, H. (2007). Longevity among hunter-gatherers: A cross-cultural examination. *Population and Development Review*, *33*(2), 321–365. https://doi.org/10.1111/j.1728-4457.2007.00171.x.

Gurven, M., Kaplan, H., & Supa, A. Z. (2007). Mortality experience of Tsimane Amerindians of Bolivia: Regional variation and temporal trends. *American Journal of Human Biology*, *19*(3), 376–398. https://doi.org/10.1002/ajhb.20600.

Guyuron, B., Rowe, D. J., Weinfeld, A. B., Eshraghi, Y., Fathi, A., & Iamphongsai, S. (2009). Factors contributing to the facial aging of identical twins. *Plastic and Reconstructive Surgery*, *123*(4), 1321–1331. https://doi.org/10.1097/PRS.0b013e31819c4d42.

Henderson, A. J., Holzleitner, I. J., Talamas, S. N., & Perrett, D. I. (2016). Perception of health from facial cues. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *371*(1693), 20150380. https://doi.org/10.1098/rstb.2015.0380.

Henderson, A. J., Lasselin, J., Lekander, M., Olsson, M. J., Powis, S. J., Axelsson, J., & Perrett, D. I. (2017). Skin colour changes during experimentally-induced sickness. *Brain, Behavior, and Immunity*, *60*, 312–318. https://doi.org/10.1016/j.bbi.2016.11.008.

Hope, D., Bates, T., Penke, L., Gow, A. J., Starr, J. M., & Deary, I. J. (2013). Symmetry of the face in old age reflects childhood social status. *Economics and Human Biology*, *11*(2), 236–244. https://doi.org/10.1016/j.ehb.2011.06.006.

Jones, A. L., & Kramer, R. S. S. (2015). Facial cosmetics have little effect on attractiveness judgments compared with identity. *Perception*, *44*(1), 79–86. https://doi.org/10.1068/p7904.

Jones, A. L., & Kramer, R. S. S. (2016). Facial Cosmetics and attractiveness: Comparing the effect sizes of professionally-applied cosmetics and identity. *PLoS One*, *11*(10), e0164218. https://doi.org/10.1371/journal.pone.0164218.

Jones, A. L., Kramer, R. S. S., & Ward, R. (2012). Signals of personality and health: the contributions of facial shape, skin texture, and viewing angle. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(6), 1353–1361. https://doi.org/10.1037/a0027078.

Jones, A. L., Russell, R., & Ward, R. (2015). Cosmetics alter biologically-based factors of beauty: Evidence from facial contrast. *Evolutionary Psychology*, *13*(1), 210–229. https://doi.org/10.1177/147470491501300113.

Jones, A. L., Porcheron, A., Sweda, J. R., Morizot, F., & Russell, R. (2016). Coloration in different areas of facial skin is a cue to health: The role of cheek redness and periorbital luminance in health perception. *Body Image*, *17*, 57–66. https://doi.org/10.1016/j.bodyim.2016.02.001.

Jones, A. L., Batres, C., Porcheron, A., Sweda, J. R., Morizot, F., & Russell, R. (2017). Positive facial affect looks healthy. *Visual Cognition*. https://doi.org/10.1080/13506285.2017.1369202 Advance online publication.

Kalick, S. M., Zebrowitz, L. A., Langlois, J. H., & Johnson, R. M. (1998). Does human facial attractiveness honestly advertise health? Longitudinal data on an evolutionary question. *Psychological Science*, *9*(1), 8–13.

Kaplan, H., Hill, K., Lancaster, J., & Hurtado, A. M. (2000). A theory of human life history evolution: Diet, intelligence, and longevity. *Evolutionary Anthropology*, *9*(4), 156–185.

Kleisner, K., Kočnar, T., Tureček, P., Stella, D., Akoko, R. M., Třebický, V., & Havlíček, J. (in press). African and European perception of African female attractiveness. Evolution and Human Behavior. doi:https://doi.org/10.1016/j.evolhumbehav.2017.07.002 (Advance online publication).

Kościński, K. (2012). Hand attractiveness—its determinants and associations with facial attractiveness. *Behavioral Ecology*, *23*(2), 334–342. https://doi.org/10.1093/beheco/arr190.

Kramer, R. S. S., & Ward, R. (2010). Internal facial features are signals of personality and health. *The Quarterly Journal of Experimental Psychology*, *63*(11), 2273–2287. https://doi.org/10.1080/17470211003770912.

Law Smith, M., Perrett, D., Jones, B., Cornwell, R., Moore, F., Feinberg, D., ... Hillier, S. (2006). Facial appearance is a cue to oestrogen levels in women. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1583), 135–140. https://doi.org/10.1098/rspb.2005.3296.

Lee, A. J., Mitchem, D. G., Wright, M. J., Martin, N. G., Keller, M. C., & Zietsch, B. P. (2016). Facial averageness and genetic quality: testing heritability, genetic correlation with attractiveness, and the paternal age effect. *Evolution and Human Behavior*, *37*(1), 61–66. https://doi.org/10.1016/j.evolhumbehav.2015.08.003.

Lefevre, C. E., & Perrett, D. I. (2015). Fruit over sunbed: carotenoid skin colouration is found more attractive than melanin colouration. *Quarterly Journal of Experimental Psychology (2006)*, *68*(2), 284–293. https://doi.org/10.1080/17470218.2014.944194.

Lefevre, C. E., Ewbank, M. P., Calder, A. J., Hagen, E. v. d., & Perrett, D. I. (2013). It is all in the face: carotenoid skin coloration loses attractiveness outside the face. *Biology Letters*, *9*(6), 20130633. https://doi.org/10.1098/rsbl.2013.0633.

Lie, H. C., Rhodes, G., & Simmons, L. W. (2008). Genetic diversity revealed in human faces. *Evolution; International Journal of Organic Evolution*, *62*(10), 2473–2486. https://doi.org/10.1111/j.1558-5646.2008.00478.x.

Little, A. C., Jones, B. C., Waitt, C., Tiddeman, B. P., Feinberg, D. R., Perrett, D. I., & Marlowe, F. W. (2008). Symmetry is related to sexual dimorphism in faces: Data across culture and species. *PLoS One*, *3*(5), e2106. https://doi.org/10.1371/journal.pone.0002106.

Little, A. C., DeBruine, L. M., & Jones, B. C. (2011). Exposure to visual cues of pathogen contagion changes preferences for masculinity and symmetry in opposite-sex faces. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*(1714), 2032–2039. https://doi.org/10.1098/rspb.2010.1925.

Matts, P. J., Fink, B., Grammer, K., & Burquest, M. (2007). Color homogeneity and visual perception of age, health, and attractiveness of female facial skin. *Journal of the American Academy of Dermatology*, *57*(6), 977–984. https://doi.org/10.1016/j.jaad.2007.07.040.

Mazumder, B., Almond, D., Park, K., Crimmins, E. M., & Finch, C. E. (2010). Lingering prenatal effects of the 1918 influenza pandemic on cardiovascular disease. *Journal of Developmental Origins of Health and Disease*, *1*(1), 26–34. https://doi.org/10.1017/S2040174409990031.

McCallum, J. (1995). The SF-36 in an Australian sample: validating a new, generic health status measure. *Australian Journal of Public Health*, *19*(2), 160–166.

Mogilski, J. K., & Welling, L. L. M. (2017). The relative importance of sexual dimorphism, fluctuating asymmetry, and color cues to health during evaluation of potential partners' facial photographs. *Human Nature*, *28*(1), 53–75. https://doi.org/10.1007/s12110-016-9277-4.

Møller, A. P., & Swaddle, J. P. (1997). *Asymmetry, developmental stability, and evolution*. Oxford, UK: Oxford University Press.

Monin, B., & Oppenheimer, D. M. (2005). Correlated averages vs. averaged correlations: Demonstrating the warm glow heuristic beyond aggregation. *Social Cognition*, *23*(3), 257–278.

Morrison, E. R., Morris, P. H., & Bard, K. A. (2013). The stability of facial attractiveness: Is it what you've got or what you do with it? *Journal of Nonverbal Behavior*, *37*(2), 59–67. https://doi.org/10.1007/s10919-013-0145-1.

Nash, R., Fieldman, G., Hussey, T., Lévêque, J. -L., & Pineau, P. (2006). Cosmetics: They influence more than Caucasian female facial attractiveness. *Journal of Applied Social Psychology*, *36*(2), 493–504. https://doi.org/10.1111/j.0021-9029.2006.00016.x.

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin*, *35*(12), 1661–1671. https://doi.org/10.1177/0146167209346309.

Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, *296*(5570), 1029–1031. https://doi.org/10.1126/science.1069675.

Oyetakin-White, P., Suggs, A., Koo, B., Matsui, M. S., Yarosh, D., Cooper, K. D., & Baron, E. D. (2015). Does poor sleep quality affect skin ageing? *Clinical and Experimental Dermatology*, *40*(1), 17–22. https://doi.org/10.1111/ced.12455.

Özener, B., & Fink, B. (2010). Facial symmetry in young girls and boys from a slum and a control area of Ankara, Turkey. *Evolution and Human Behavior*, *31*(6), 436–441. https://doi.org/10.1016/j.evolhumbehav.2010.06.003.

Penton-Voak, I. S., Jones, B. C., Little, A. C., Baker, S., Tiddeman, B., Burt, D. M., & Perrett, D. I. (2001). Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proceedings of the Royal Society B: Biological Sciences*, *268*(1476), 1617–1623. https://doi.org/10.1098/rspb.2001.1703.

Pezdirc, K., Rollo, M. E., Whitehead, R., Hutchesson, M. J., Ozakinci, G., Perrett, D., & Collins, C. E. (2017). Perceptions of carotenoid and melanin colouration in faces among young Australian adults. *Australian Journal of Psychology*. https://doi.org/10.1111/ajpy.12163.

Ponsonby, A. -L., Dwyer, T., & Couper, D. (1997). Sleeping position, infant apnea, and cyanosis: A population-based study. *Pediatrics*, *99*(1), e3. https://doi.org/10.1542/peds.99.1.e3.

Porcheron, A., Mauger, E., & Russell, R. (2013). Aspects of facial contrast decrease with age and are cues for age perception. *PLoS One*, *8*(3), e57985. https://doi.org/10.1371/journal.pone.0057985.

Pound, N., Lawson, D. W., Toma, A. M., Richmond, S., Zhurov, A. I., & Penton-Voak, I. S. (2014). Facial fluctuating asymmetry is not associated with childhood ill-health in

a large British cohort study. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*(1792), 20141639. https://doi.org/10.1098/rspb.2014.1639.
Rantala, M. J., Coetzee, V., Moore, F. R., Skrinda, I., Kecko, S., Krama, T., & Krams, I. (2013). Adiposity, compared with masculinity, serves as a more valid cue to immunocompetence in human mate choice. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1751). https://doi.org/10.1098/rspb.2012.2495.
Rao, A., & Rao, L. (2007). Carotenoids and human health. *Pharmacological Research*, *55*(3), 207–216. https://doi.org/10.1016/j.phrs.2007.01.012.
Re, D. E., & Rule, N. O. (2016). Heavy matters: The relationship between just noticeable differences in perceptions of facial adiposity and facial attractiveness. *Social Psychological and Personality Science*, *7*(1), 69–76. https://doi.org/10.1177/1948550615599829.
Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, *57*(1), 199–226. https://doi.org/10.1146/annurev.psych.57.102904.190208.
Rhodes, G., Zebrowitz, L. A., Clark, A., Kalick, S. M., Hightower, A., & McKay, R. (2001). Do facial averageness and symmetry signal health? *Evolution and Human Behavior*, *22*(1), 31–46.
Rhodes, G., Yoshikawa, S., Palermo, R., Simmons, L. W., Peters, M., Lee, K., ... Crawford, J. R. (2007). Perceived health contributes to the attractiveness of facial symmetry, averageness, and sexual dimorphism. *Perception*, *36*(8), 1244–1252. https://doi.org/10.1068/p5712.
Russell, R. (2003). Sex, beauty, and the relative luminance of facial features. *Perception*, *32*(9), 1093–1107. https://doi.org/10.1068/p5101.
Russell, R., Sweda, J. R., Porcheron, A., & Mauger, E. (2014). Sclera color changes with age and is a cue for perceiving age, health, and beauty. *Psychology and Aging*, *29*(3), 626–635. https://doi.org/10.1037/a0036142.
Russell, R., Porcheron, A., Sweda, J. R., Jones, A. L., Mauger, E., & Morizot, F. (2016). Facial contrast is a cue for perceiving health from the face. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1354–1362. https://doi.org/10.1037/xhp0000219.
Samson, N., Fink, B., & Matts, P. J. (2010). Visible skin condition and perception of human facial appearance. *International Journal of Cosmetic Science*, *32*(3), 167–184. https://doi.org/10.1111/j.1468-2494.2009.00535.x.
Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1431), 1913–1917.
Scott, I. M. L., Pound, N., Stephen, I. D., Clark, A. P., & Penton-Voak, I. S. (2010). Does masculinity matter? The contribution of masculine face shape to male attractiveness in humans. *PLoS One*, *5*(10), e13585. https://doi.org/10.1371/journal.pone.0013585.
Scott, I. M. L., Clark, A. P., Boothroyd, L. G., & Penton-Voak, I. S. (2013). Do men's faces really signal heritable immunocompetence? *Behavioral Ecology*, *24*(3), 579–589. https://doi.org/10.1093/beheco/ars092.
Shackelford, T. K., & Larsen, R. J. (1997). Facial asymmetry as an indicator of psychological, emotional, and physiological distress. *Journal of Personality and Social Psychology*, *72*(2), 456–466.
Smith, F. G., Jones, B. C., DeBruine, L. M., & Little, A. C. (2009). Interactions between masculinity–femininity and apparent health in face preferences. *Behavioral Ecology*, *20*(2), 441–445. https://doi.org/10.1093/beheco/arn141.
Stephen, I. D., & McKeegan, A. M. (2010). Lip colour affects perceived sex typicality and attractiveness of human faces. *Perception*, *39*(8), 1104–1110. https://doi.org/10.1068/p6730.
Stephen, I. D., Coetzee, V., Law Smith, M., & Perrett, D. I. (2009a). Skin blood perfusion and oxygenation colour affect perceived human health. *PLoS One*, *4*(4), e5083. https://doi.org/10.1371/journal.pone.0005083.
Stephen, I. D., Law Smith, M. J., Stirrat, M. R., & Perrett, D. I. (2009b). Facial skin coloration affects perceived health of human faces. *International Journal of Primatology*, *30*(6), 845–857. https://doi.org/10.1007/s10764-009-9380-z.
Stephen, I. D., Coetzee, V., & Perrett, D. I. (2011). Carotenoid and melanin pigment coloration affect perceived human health. *Evolution and Human Behavior*, *32*(3), 216–227. https://doi.org/10.1016/j.evolhumbehav.2010.09.003.
Sundelin, T., Lekander, M., Kecklund, G., Van Someren, E. J. W., Olsson, A., & Axelsson, J. (2013). Cues of fatigue: effects of sleep deprivation on facial appearance. *Sleep*, *36*(9), 1355–1360. https://doi.org/10.5665/sleep.2964.
Sundelin, T., Lekander, M., Sorjonen, K., & Axelsson, J. (2017). Negative effects of restricted sleep on facial appearance and social appeal. *Royal Society Open Science*, *4*(5), 160918. https://doi.org/10.1098/rsos.160918.
Tan, K. W., Graf, B. A., Mitra, S. R., & Stephen, I. D. (2017). Impact of fresh fruit smoothie consumption on apparent health of Asian faces. *Evolution and Human Behavior*, *38*(4), 522–529. https://doi.org/10.1016/j.evolhumbehav.2017.02.004.
Thornhill, R., & Gangestad, S. W. (2006). Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. *Evolution and Human Behavior*, *27*(2), 131–144. https://doi.org/10.1016/j.evolhumbehav.2005.06.001.
Tiddeman, B., Burt, D. M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications*, *21*, 42–50. https://doi.org/10.1109/38.946630.
Tybur, J. M., & Gangestad, S. W. (2011). Mate preferences and infectious disease: theoretical considerations and evidence in humans. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366*(1583), 3375–3388. https://doi.org/10.1098/rstb.2011.0136.
Vazire, S., Naumann, L. P., Rentfrow, P. J., & Gosling, S. D. (2008). Portrait of a narcissist: Manifestations of narcissism in physical appearance. *Journal of Research in Personality*, *42*(6), 1439–1447. https://doi.org/10.1016/j.jrp.2008.06.007.
Ware, J., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical Care*, *34*(3), 220–233.
Weatherall, I. L., & Coombs, B. D. (1992). Skin color measurements in terms of CIELAB color space values. *Journal of Investigative Dermatology*, *99*(4), 468–473. https://doi.org/10.1111/1523-1747.ep12616156.
Welling, L. L. M., Conway, C. A., Debruine, L. M., & Jones, B. C. (2007). Perceived vulnerability to disease is positively related to the strength of preferences for apparent health in faces. *Journal of Evolutionary Psychology*, *5*(1), 131–139. https://doi.org/10.1556/JEP.2007.1012.
Whitehead, R. D., Re, D., Xiao, D., Ozakinci, G., & Perrett, D. I. (2012). You are what you eat: Within-subject increases in fruit and vegetable consumption confer beneficial skin-color changes. *PLoS One*, *7*(3), e32988. https://doi.org/10.1371/journal.pone.0032988.
Younsi, M., & Chakroun, M. (2014). Measuring health-related quality of life: psychometric evaluation of the Tunisian version of the SF-12 health survey. *Quality of Life Research: an International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *23*(7), 2047–2054. https://doi.org/10.1007/s11136-014-0641-8.
Zebrowitz, L. A., & Rhodes, G. (2004). Sensitivity to "bad genes" and the anomalous face overgeneralization effect: Cue validity, cue utilization, and accuracy in judging intelligence and health. *Journal of Nonverbal Behavior*, *28*(3), 167–185. https://doi.org/10.1023/B:JONB.0000039648.30935.1b.
Zebrowitz, L. A., Franklin, R. G., Boshyan, J., Luevano, V., Agrigoroaei, S., Milosavljevic, B., & Lachman, M. E. (2014). Older and younger adults' accuracy in discerning health and competence in older and younger faces. *Psychology and Aging*, *29*(3), 454–468. https://doi.org/10.1037/a0036255.
Zhen, R., & Stevenson, R. L. (2015). Image demosaicing. In E. Celebi, M. Lecca, & B. Smolka (Eds.), *Color image and video enhancement* (pp. 13–54). Cham: Springer.