Personality and Individual Differences xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Personality and Individual Differences

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

Accuracy in assessment of self-reported stress and a measure of health from static facial information

Anthony C. Little [a,*], Jean McPherson [a], Lillias Dennington [a], Benedict C. Jones [b]

[a] School of Natural Sciences, University of Stirling, Stirling FK9 4LA, UK
[b] School of Psychology, University of Aberdeen, Aberdeen AB24 2UB, UK

ARTICLE INFO

Article history:
Received 23 March 2011
Received in revised form 1 June 2011
Accepted 6 June 2011
Available online xxxx

Keywords:
Stress
Health
Appearance
Accuracy
Social perception

ABSTRACT

The human face is important for social communication and in attractiveness judgements. Previous studies indicate that several facial traits may be related to mental and physical health and there is some evidence that individuals are able to judge past health on the basis of facial appearance. The current study builds on this prior work, examining the relationship between static facial appearance and self-reported stress and health. Specifically, we examined (1) within and between individual stress (Study 1) by photographing the same participants at two times, once in a relatively stress free and once in a stressful time, and (2) between individual health (Studies 2A and 2B) by examining self-reported past number of colds as a measure of immune function. All studies demonstrated that individuals could judge the stress and physical health of others from static facial appearance alone at rates greater than chance. Such accuracy may reflect selection pressures to identify stress free and healthy social partners.

1. Introduction

The human face is probably the most important visual trait in social communication. People believe that faces provide clues to future behaviour and personality (Hassin & Trope, 2000; Liggett, 1974) and studies have demonstrated some accuracy in guessing aspects of personality from facial photographs of individuals (Borkenau & Liebler, 1993; Little & Perrett, 2007). In this study, we examined the perception and accuracy of detecting stress and health.

There is a large and obvious selective advantage in detecting stress free and healthy partners both for social exchange and mate choice. There have been several studies linking facial appearance and healthiness. Asymmetry has received much attention as it is thought to represent an indicator of developmental stress. Symmetric human faces are perceived as healthier than asymmetric faces (Jones et al., 2001, 2004; Rhodes et al., 2001). While some studies have reported no relationships (Rhodes et al., 2001), other studies using more controlled stimuli report a positive relationship between facial symmetry and health (Thornhill & Gangestad, 2006) and between sex-typical traits and symmetry (Little et al., 2008). Facial symmetry is also positively related to measures of psychological well being (Shackelford & Larsen, 1997) and measures of intelligence (Furlow, ArmijoPrewitt, Gangestad, & Thornhill, 1997; Penke et al., 2009), indicating that symmetry may be associated with mental health as well as physical health. Symmetry in faces appears important in non-human primate species. For example, asymmetry in canine teeth has been found to positively correlate with environmental deterioration in western lowland gorillas (Manning & Chamberlain, 1994). Additionally, measures of health as rated by keepers and zoo staff are positively associated with symmetry in zoo populations of chimpanzees (Sefcek & King, 2007).

Alongside symmetry, studies have shown positive relationships between sex-typical face traits and health in men (Rhodes, Chan, Zebrowitz, & Simmons, 2003; Thornhill & Gangestad, 2006) and women (Thornhill & Gangestad, 2006). Feminine women and masculine men report fewer health problems. One study addressing rated distinctiveness, the converse of averageness, has shown negative associations with health problems in men and women (Rhodes et al., 2001). Both traits may reflect health for different reasons. Sex-typicality of face shape has been linked to hormone production (Law-Smith et al., 2006; Pound, Penton-Voak, & Surridge, 2009) and averageness has been linked to genetic diversity (Lie, Rhodes, & Simmons, 2008).

Studies have also examined perceived traits. Two studies have addressed whether rated facial attractiveness is related to long-term health measures, although results are mixed (Kalick, Zebrowitz, Langlois, & Johnson, 1998; Shackelford & Larsen, 1999). These two studies draw on a historical set of photographs of individuals associated with long-term health records. Shackelford and Larsen did find that attractiveness is positively related to health measures while Kalick et al. did not. Importantly, Kalick et al. did find ratings

* Corresponding author. Tel.: +44 01786 467651; fax: +44 01786 467641.
E-mail address: anthony.little@stir.ac.uk (A.C. Little).

0191-8869/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.paid.2011.06.010

of health were positively related to measures of health, however. Other studies have shown that facial attractiveness is positively related to longevity in humans while ratings of health are not (Henderson & Anglin, 2003). Facial attractiveness is also associated with genes coding for the immune system. Faces of individuals who are heterozygous at the major histocompatibility complex, genes associated with immune defence, are rated as more attractive/healthier than those who are homozygous (Roberts et al., 2005). Together these studies suggest that ratings based on facial appearance are linked to health, though so far the evidence for accurate perceptions of health based on facial appearance are mixed.

The current study examined relationships between ratings of stress and health from static facial appearance and self-reported stress and health. We examine accuracy in cues to current mental well being within and between individuals using rated and self-reported stress (Study 1). We also use a measure of health (self-reported colds) to examine whether judgements of health of different individuals are accurate (Studies 2A and 2B).

2. Study 1: judging self-reported stress from individual faces across time

Some individuals are more likely to experience stress than others and stress is also changeable over time depending on situation. In Study 1 we address how stress within and between individuals might relate to differences in their facial appearance.

2.1. Participants

Participants were 12 females (M = 23.6, SD = 9.1) and 12 males (M = 21.5, SD = 3.1) who were photographed twice. All photographed participants were of white European appearance. Forty-five different participants (31 females, mean age = 19.0, SD = 1.4, 14 males, mean age = 19.7, SD = 2.1) rated these face images for stress. These latter participants received course credit to take part in laboratory testing.

2.2. Photography

Full frontal colour facial photographs were taken of all participants under standardised diffuse lighting conditions and against a constant background. Participants were asked to pose with a neutral facial expression and were asked to pull their hair back from their face. Women were asked to remove their make-up prior to coming into the laboratory and were asked to remove make-up if they appeared to be wearing it on arrival. Participants were asked to remove spectacles and males were clean shaven. Photographs were taken at two points to provide a stressful versus a less stressful time, the middle of an autumn semester after a mid-semester break (12–18th October 2006) and the end of the same semester during the university exam period (23–30th November 2006). It was expected students would experience more stress at exam time and that the one month intervening time span would limit other types of change over time.

2.3. Measuring self-reported stress

After photographs were taken, participants filled out a short questionnaire which included demographic information, age, sex, race, nationality and the Hospital Anxiety and Depression scale (Zigmond & Snaith, 1983). Questions addressed recent feelings: participants rated "how you have been feeling in the past week". Scores were computed to create measures of both current anxiety and current depression. An example of a question assessing anxiety was: "I feel tense or 'wound up'". An example of a question assessing depression was: "I feel cheerful". Answer options differed by question but all were on 4-point scales. For example, most of the time, a lot of the time, time to time, occasionally, not at all.

2.4. Procedure for stress judgements

Judgements were carried out under laboratory conditions. Participants were asked to rate the 48 faces for stress with the following question: "How STRESSED is the person in this picture?" Ratings were on a 7-point scale (1 = low and 7 = high). Faces were presented to participants on computer screen individually and in a random order. Rating the face from 1 to 7 and pressing enter brought up the next face. There was no time limit for the ratings.

2.5. Results

2.5.1. Reliability of ratings

Reliability analysis using Coefficient Alpha revealed very high between-judge agreement for ratings of stress for the 48 faces (alpha = .932). Averaging male and female ratings separately, there was a strong correlation between these ratings (r = .87, p < .001).

2.5.2. Effects of time: manipulation checks on self-perception

We conducted mixed model ANOVAs to examine the effects of time of testing within-participants with self-rated anxiety scores as the dependent variable, time (stressed/unstressed) as a within-participant factor, and sex of participant as a between participant factor. This revealed a significant effect of time ($F_{1,22}$ = 5.89, p = .024, η^2 = .211), no significant effect of sex of participant ($F_{1,22}$ = .02, p = .895, η^2 = .001), and no interaction between these factors ($F_{1,22}$ = .85, p = .367, η^2 = .037). A correlation revealed a significant relationship between anxiety scores at the low stress and high stress time (r = .564, p = .004). Participants rated themselves as more stressed in the stressful time and between-individual differences in self-rated stress were consistent across test sessions. Mean scores can be seen in Fig. 1.

The same analysis for self-rated depression scores revealed no significant effect of time ($F_{1,22}$ = .22, p = .640, η^2 = .010), sex of participant ($F_{1,22}$ = .04, p = .848, η^2 = .002), and no interaction between these factors ($F_{1,22}$ < .01, p > .999, η^2 < .001). A correlation revealed a positive but not significant relationship between scores at the low stress and high stress time (r = .257, p = .225). Participants did not differ in their depression between the two test sessions and self-rated depression was not significantly consistent across time between participants.

2.5.3. Effects of time: other rated stress

The same analysis was conducted with other-rated stress as the dependent measure. This revealed a significant effect of time ($F_{1,22}$ = 4.53, p = .045, η^2 = .171), no significant effect of sex of participant ($F_{1,22}$ = .03, p = .863, η^2 = .001), and no interaction between these factors ($F_{1,22}$ = .16, p = .690, η^2 = .007). We followed up this analysis by computing a difference score of other rated stress by subtracting ratings at the low stress time from the higher stress time. Positive scores indicate that other rated stress was higher at the stressful than the less stressful time. A one sample t-test against chance (no change, 0) revealed a significant effect of time (M = .25, SD = .57, t = 2.17, p = .041). A correlation revealed a significant relationship between scores at lower stress and higher stress (r = .685, p < .001). Participant faces were rated as more stressed in the stressful time and rated stress was consistent across time between participants. Mean scores can be seen in Fig. 1.

2.5.4. Between participant correlations: self and other rated stress

Finally, we used Pearson correlations revealing positive relationships between self and other rated anxiety/stress at the less

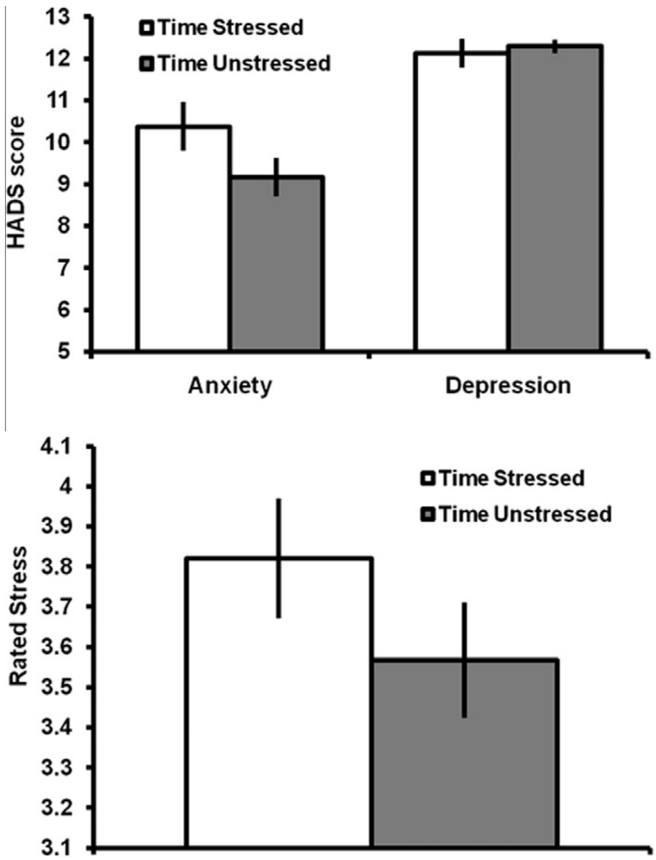

Fig. 1. HADS scores (top) and rated health of photographs (bottom) of participants according to time of measurement/photograph (±1 SEM, between participant variance).

stressful time ($r = .425$, $p = .038$) and the more stressful time ($r = .393$, $p = .057$), though the latter was only approaching significance.

3. Study 2: judging self-reported health from composite and individual faces

Studies 2A and 2B addressed accuracy for medical/physical health using number of reported colds as a measure of resistance to infection. Study 2A used composite images in which characteristics common to the individual faces combined are maintained, while idiosyncratic variations that are not common to the set are 'averaged out'. Study 2B examined relationships between self- and other-rated health using unmanipulated individual faces.

3.1. Study 2A: composite images

3.1.1. Participants

Participants for the photographs were 60 individuals (30 women, mean age = 20.9, SD = 1.7, 30 men, mean age = 22.0, SD = 3.1) who were paid £5 to come into the laboratory to take part. A different 148 participants judged the faces for health (85 women, mean age = 22.5, SD = 5.8, 63 men, mean age = 21.8, SD = 3.7). These participants were recruited over the internet via a research website and completed the test remotely.

3.1.2. Photography

Photographs were taken under the same conditions as Study 1. Participants filled out age and gender information as well as health related questions outlined below.

3.1.3. Measuring health

We asked photographed participants to state whether they currently had a cold (yes/no) and how many times a year they would suffer from a cold. The following options were listed and corresponding scores were: 0 = never, 1 = 1–2 times a year, 2 = 3–4 times a year, 3 = 5–6 times a year, 4 = 7–8 times a year, 5 = 9–10 times a year, 6 = 11+ times a year.

3.1.4. Making the composites and test

The 10 highest and 10 lowest scorers on the self-reported colds question for men and women were selected to make up composites. Due to some overlap in scores where multiple individuals for a category of score were tied, faces were selected at random to make the 10. For the low female face, this meant 9 faces were chosen from 16 tied scoring "1" and for the high face 5 faces were chosen from 8 scoring "2". For males, this meant 10 faces were chosen from 19 tied scoring "1" and for the high face 8 faces were chosen from 9 scoring "2". Some participants reported having current colds but it was possible to control for this variable by matching numbers in each of the four sets (one individual in each). Ten faces were deemed sufficient to capture the average configuration of high and low scoring individuals as the perception of individuality or distinctiveness in composite images changes little after six images (Little & Hancock, 2002). The mean difference in colds on the above scale between the highest 10 and lowest 10 scorers was 1.5 for men and 2.1 for women. The numbers for the high and low cold groups were significantly different from each other for both women ($t_{18} = 4.85$, $p < .001$, $d = 2.29$) and men ($t_{18} = 3.74$, $p = .002$, $d = 1.76$).

The composite faces were created using specially designed software (Tiddeman, Burt, & Perrett, 2001). Key locations (174 points) were manually marked around the main features (e.g., points outline, eyes, nose, and mouth) and the outline of each face (e.g., jaw line and hair line). The average location of each point in the 10 faces for each composite was then calculated. The features of the individual faces were then morphed to the relevant average shape before superimposing the images to produce a photographic quality result. For more information on this technique see Tiddeman et al. (2001). Images can be seen in Fig. 2.

3.1.5. Procedure for health judgements

Judgements were carried out remotely over the internet. Participants were asked to judge the two pairs of faces (high versus low composite for female and male faces) for health with the following question: "Please indicate which face you think looks most HEALTHY by clicking on the face below". Faces were presented to participants on computer screen in pairs and in a random order. Selection of a face from the pair brought up the next face pair. There was no time limit for the judgements.

3.2. Results

A mixed model ANOVA with accuracy scores as the dependent variable, sex of face (male/female) as a within-participant factor, and sex of participant as a between participant factor revealed no significant effect of sex of face ($F_{1,146} = .42$, $p = .680$, $\eta^2 = .002$), no significant effect of sex of participant ($F_{1,146} = .02$, $p = .867$, $\eta^2 < .001$), and no interaction between these factors ($F_{1,146} = 1.12$, $p = .292$, $\eta^2 = .008$). Combining scores for male and female faces into an average accuracy score, a one sample t-test against chance (50%) revealed that average accuracy was significantly higher than chance (M = 59.8%, SD = 32.3, $t_{147} = 3.69$, $p < .001$, $d = .61$). Chi square tests revealed that accuracy was significantly greater than chance for both female ($\chi^2 = 6.92$, DF = 1, $p = .009$) and male ($\chi^2 = 4.57$, DF = 1, $p = .033$) faces separately. Mean proportion scores can be seen in Fig. 3.

Please cite this article in press as: Little, A. C., et al. Accuracy in assessment of self-reported stress and a measure of health from static facial information. *Personality and Individual Differences* (2011), doi:10.1016/j.paid.2011.06.010

Fig. 2. Composite images representing the shape and colour of individuals with high (left) and low (right) self-assessments of number of colds.

3.3. Study 2B: individual images

3.3.1. Participants

Participants for the photographs were 51 individuals (27 women, mean age = 19.8, SD = 2.6, 24 men, mean age = 19.0, SD = 1.3) who received course credit to come into the laboratory to take part. A different 11 participants rated the faces for health (eight women, mean age = 32.1, SD = 14.7, three men, mean age = 27.7, SD = 7.5) who also received course credit to come into the laboratory to take part.

3.3.2. Photography

Photographs were taken under the same conditions as Study 1. Participants filled out age and gender information as well as health related questions.

3.3.3. Measuring health

The same questions were used as in Study 2A.

3.3.4. Procedure for health ratings

Judgements were carried out under laboratory conditions. Participants were asked to rate the 51 faces for healthiness with

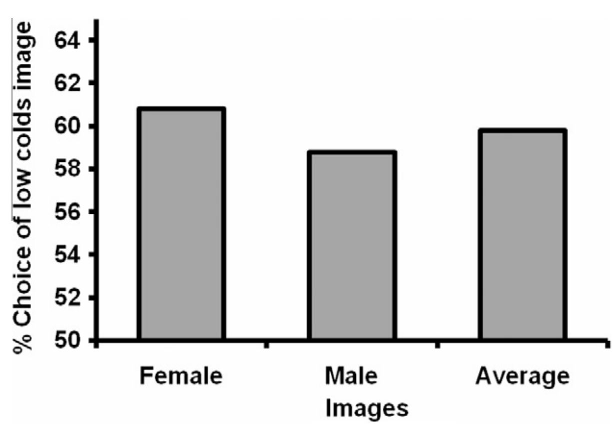

Fig. 3. Percent choice of the low colds faces when asked to choose the healthiest appearing.

the following question: "How HEALTHY is the person in this picture?" Ratings were on a 7-point scale (1 = low and 7 = high). Faces were presented to participants on computer screen individually and in a random order. Rating the face from 1 to 7 and pressing enter brought up the next face. There was no time limit for the ratings.

3.4. Results

3.4.1. Reliability of ratings

Reliability analysis using Coefficient Alpha revealed very high between-judge agreement for ratings of health (alpha = .885). Averaging male and female ratings separately, there was a strong correlation between these ratings ($r = .81$, $p < .001$).

3.4.2. Self-reported health

Sex differences were analysed using independent samples *t*-tests. These revealed that women were perceived as more healthy than men (women mean = 4.34, SD = .78, men mean = 3.88, SD = .75, $t_{49} = 2.15$, $p = .037$, $d = .61$) and that women perceived themselves as having less colds than men (women mean = 1.37, SD = .84, men mean = 1.96, SD = 1.30, $t_{49} = 1.94$, $p = .058$, $d = .55$), though the latter was only approaching significance.

A regression analysis entering age, gender, current cold, and number of colds to predict other-rated health revealed a significant overall model ($F_{3,189} = 4.23$, $p = .005$, $R^2 = .27$). There was a significant negative relationship for number of colds ($\beta = -.476$, $p = .002$), close to significant negative relationships for age ($\beta = -.254$, $p = .062$) and gender ($\beta = -.236$, $p = .087$), and no significant relationship for current cold ($\beta = .157$, $p = .275$).

Pearson correlations revealed a positive relationship between self-rated health and other-rated health for both women ($r = -.352$, $p = .072$) and men ($r = -.364$, $p = .081$), though both were only approaching significance. The difference between these correlations was not significant ($Z = .08$, $p = .936$), justifying not splitting by sex. The correlation was significant when these data were not split by sex ($r = -.401$, $p = .004$) and converting male and female self- and other-rated health scores separately to *Z*-scores and rerunning this correlation also did not affect the relationship ($r = -.357$, $p = .010$).

4. Discussion

The current studies demonstrate that individuals can, to some extent, accurately judge the mental stress and immunological competence of an unknown individual based only on static cues from facial appearance for the measures used here. Study 1 demonstrated that individual variation in self-reported stress is observable based only on facial photographs. There were also positive correlations between self and other rated anxiety/stress, although we note that this was significant for the low stress images and only close to significance for the higher stress images. Observers could then detect cues to an individual's internal state of stress by examining their face. Study 2A demonstrated that in composite face images of those experiencing most and least number of colds in a 12 month period the two images differed in perceived health. For both men and women, the composites of individuals indicating a higher number of colds were seen as less healthy. Study 2B demonstrated that accuracy was also evident when individual unmanipulated faces were rated for health. Again, for both men and women, those indicating they were more likely to have colds were rated as less healthy than those indicating they had lower numbers of colds.

Our data demonstrates that cues to stress are visible in faces and that this variation can be seen within and between individuals, addressing current state stress and perhaps stress as a trait. Such accuracy is consistent with links between certain traits such as facial symmetry being positively related to measures of psychological well-being (Shackelford & Larsen, 1997). Accuracy in perceptions of health are also consistent with previous studies that have demonstrated links between particular traits and physical health such as symmetry (Thornhill & Gangestad, 2006), sex-typicality (Rhodes et al., 2003; Thornhill & Gangestad, 2006), and averageness (Rhodes et al., 2001). Potentially it is perception of such traits in the faces here that lead to accurate judgements given, for example, symmetric human faces are perceived as healthier than asymmetric faces (Jones et al., 2001; Rhodes et al., 2001).

Previous studies have generated mixed evidence for accurate perceptions of health. Attractiveness has been found to be related to real physical health and longevity (Henderson & Anglin, 2003; Shackelford & Larsen, 1999) and perceived health has been found to be positively related to measures of health in some studies (Kalick et al., 1998) but not longevity (Henderson & Anglin, 2003). Overall, this work does suggest accuracy in health perception and our data adds to this literature. In Study 2A people identified healthier from less healthy individuals about 10% greater than chance and in Study 2B ratings of health explained about 16% (R^2 of correlation for both men and women together = .161) of the variance in self-rated health. Together this suggests that people can detect another's health from their face but that such detection is far from perfect (in Study 2A 40% guessed incorrectly). Such limited accuracy might be predicted given the face is unlikely to be a perfect signal of current mental or physical health. Given our relatively crude measure of past physical health (self-reported number of colds) it appears unlikely that such a measure would capture health very accurately. Such a measure likely reflects immunological capabilities and perhaps is less closely tied to physical fitness and diet, for example. We also note that while individuals seem likely to remember the number of colds they have suffered, there is likely some bias in interpretation here as participants may under or over report. Future studies can investigate other measures of health, and we would perhaps expect stronger effects in studies measuring more serious health issues.

The mechanism for accuracy remains unclear. Previous authors have suggested that accuracy may come about via self-fulfilling prophecies (Snyder, Tanke, & Berscheid, 1977) whereby facial appearance affects social perception, leading individuals to behave in a manner consistent with how they are perceived. This seems unlikely for the current results because it is unclear how an individual believing another to be unhealthy or stressed could act in such a way as to cause the perceived person to actually be unhealthier or more stressed.

Alternatively, as noted above, several face traits, such as symmetry, are associated with both real health and perceptive health and examining such traits could then lead to accurate perceptions of health. Stress and health may also be seen via emotional leaks visible in faces. The personality dispositions of elderly people are reflected in their faces, with, for example, those of a hostile disposition tending to look angry (Malatesta, Fiore, & Messina, 1987). People with an irritable temperament may tense certain facial muscles in a way that yields different jaw development from that shown in people who are more relaxed (Kreiborg, Jensen, Moller, & Bjork, 1978). Potentially, within-person changes in stress may also cause the tensing of the same muscles leading to increased perceived stress. Subtle expression differences between individuals could also relate to accuracy in perceiving physical health. Poor health or stress may also be related to poor diet and/or poor grooming practices impacting on health through mediation of appearance. There is evidence that good diet (Stephen, Coetzee, & Perrett, 2011) and grooming (Cash, 1990) improve appearance. Stress and poor health could also have direct impact on perceptions of health via factors such as skin appearance. The mechanism for accuracy in perceiving health from facial appearance is an important avenue for future research.

In conclusion, the current studies demonstrate that individuals can judge the stress and health of another from facial appearance

Please cite this article in press as: Little, A. C., et al. Accuracy in assessment of self-reported stress and a measure of health from static facial information. *Personality and Individual Differences* (2011), doi:10.1016/j.paid.2011.06.010

at rates greater than chance. Potentially, such accuracy may reflect pressures to find stress free and healthy social partners. Our data are consistent with ideas that the face may be an honest cue of within and between individual stress and between individual differences in health.

Acknowledgment

Anthony Little is supported by a Royal Society University Research Fellowship.

References

Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology, 65*, 546–553.

Cash, T. F. (1990). The psychology of physical appearance: Aesthetics, attributes, and images. In T. F. Cash & T. Pruzinsky (Eds.), *Body images: Development, deviance, and change*. New York: Guilford.

Furlow, F. B., ArmijoPrewitt, T., Gangestad, S. W., & Thornhill, R. (1997). Fluctuating asymmetry and psychometric intelligence. *Proceedings of the Royal Society of London Series B – Biological Sciences, 264*, 823–829.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology, 78*, 837–852.

Henderson, J. J. A., & Anglin, J. M. (2003). Facial attractiveness predicts longevity. *Evolution and Human Behavior, 24*, 351–356.

Jones, B. C., Little, A. C., Feinberg, D. R., Penton-Voak, I. S., Tiddeman, B. P., & Perrett, D. I. (2004). The relationship between shape symmetry and perceived skin condition in male facial attractiveness. *Evolution and Human Behavior, 25*, 24–30.

Jones, B. C., Little, A. C., Penton-Voak, I. S., Tiddeman, B. P., Burt, D. M., & Perrett, D. I. (2001). Facial symmetry and judgements of apparent health – Support for a "good genes" explanation of the attractiveness–symmetry relationship. *Evolution and Human Behavior, 22*, 417–429.

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*, 8–13.

Kreiborg, S., Jensen, B. L., Moller, E., & Bjork, A. (1978). Craniofacial growth in a case of congenital muscular-dystrophy – Roentgencephalometric and electromyographic investigation. *American Journal of Orthodontics and Dentofacial Orthopedics, 74*, 207–215.

Law-Smith, M. J., Perrett, D. I., Jones, B. C., Cornwell, R. E., Moore, F. R., Feinberg, D. R., et al. (2006). Facial appearance is a cue to oestrogen levels in women. *Proceedings of the Royal Society B – Biological Sciences, 273*, 135–140.

Lie, H. C., Rhodes, G., & Simmons, L. W. (2008). Genetic diversity revealed in human faces. *Evolution, 62*, 2473–2486.

Liggett, J. (1974). *The human face*. London: Constable.

Little, A. C., & Hancock, P. J. B. (2002). The role of masculinity and distinctiveness in judgments of human male facial attractiveness. *British Journal of Psychology, 93*, 451–464.

Little, A. C., Jones, B. C., Waitt, C., Tiddeman, B. P., Feinberg, D. R., Perrett, D. I., et al. (2008). Symmetry is related to sexual dimorphism in faces: Data across culture and species. *PLos One, 3*, e2106.

Little, A. C., & Perrett, D. I. (2007). Using composite images to assess accuracy in personality attribution to faces. *British Journal of Psychology, 98*, 111–126.

Malatesta, C. Z., Fiore, M. J., & Messina, J. J. (1987). Affect, personality, and facial expressive characteristics of older-people. *Psychology and Aging, 2*, 64–69.

Manning, J. T., & Chamberlain, A. T. (1994). Fluctuating asymmetry in gorilla canines – A sensitive indicator of environmental-stress. *Proceedings of the Royal Society of London Series B – Biological Sciences, 255*, 189–193.

Penke, L., Bates, T. C., Gow, A. J., Pattie, A., Starr, J. M., Jones, B. C., et al. (2009). Symmetric faces are a sign of successful cognitive aging. *Evolution and Human Behavior, 30*, 429–437.

Pound, N., Penton-Voak, I. S., & Surridge, A. K. (2009). Testosterone responses to competition in men are related to facial masculinity. *Proceedings of the Royal Society B – Biological Sciences, 276*, 153–159.

Rhodes, G., Chan, J., Zebrowitz, L. A., & Simmons, L. W. (2003). Does sexual dimorphism in human faces signal health? *Proceedings of the Royal Society of London B, 270*, S93–S95.

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 Behaviour, 22*, 31–46.

Roberts, S. C., Little, A. C., Gosling, L. M., Perrett, D. I., Carter, V., Jones, B. C., et al. (2005). MHC-heterozygosity and human facial attractiveness. *Evolution and Human Behavior, 26*, 213–226.

Sefcek, J. A., & King, J. E. (2007). Chimpanzee facial symmetry: A biometric measure of chimpanzee health. *American Journal of Primatology, 69*, 1257–1263.

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*, 456–466.

Shackelford, T. K., & Larsen, R. J. (1999). Facial attractiveness and physical health. *Evolution and Human Behaviour, 21*, 71–76.

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behaviour: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35*, 656–666.

Stephen, I. D., Coetzee, V., & Perrett, D. I. (2011). Carotenoid and melanin pigment coloration affect perceived human health. *Evolution and Human Behavior, 32*, 216–227.

Thornhill, R., & Gangestad, S. W. (2006). Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. *Evolution and Human Behavior, 27*, 131–144.

Tiddeman, B. P., Burt, D. M., & Perrett, D. I. (2001). Prototyping and transforming facial texture for perception research. *IEEE Computer Graphics and Applications, 21*, 42–50.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*, 361–370.