CrossMark

ORIGINAL ARTICLE

Discrimination of Attractiveness and Health in Men's Faces: the Impact of Color Cues and Variation in Relation to Sex and Age of Rater

S. Craig Roberts[1] · Anthony C. Little[2] ·
Lisa M. DeBruine[3] · Marion Petrie[4]

Abstract Cues available in facial skin are used to assess mate quality in humans and non-human primates. In men, perception of skin healthiness and facial attractiveness are associated with heterozygosity at genes in the major histocompatibility complex, with potential implications for securing direct benefits through mate choice. There is, however, some debate as to the precise nature of the information available in skin that is used in health and attractiveness judgments. Here we investigate whether color cues are important in discrimination of skin healthiness and facial attractiveness in men's faces. We presented images of men judged to have attractive or unattractive faces, and healthy or less healthy skin, to independent raters in either full-color or gray-scale monochrome. Differences in ratings across these conditions indicate that hue carries additive information over that available in other skin cues (e.g. texture and tone) and that this aids discrimination of individual men's quality, especially in judgments of skin condition. We also found significant sex and age effects on discrimination. Our results are consistent with findings from other species that color cues can signal underlying quality and that sexual selection may have contributed to the evolution of color vision in primates.

Keywords Color preference · Red · Mate choice · Good-genes · MHC

✉ S. Craig Roberts
craig.roberts@stir.ac.uk

1   Division of Psychology, University of Stirling, Stirling, UK

2   Department of Psychology, University of Bath, Bath, UK

3   Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

4   Institute of Health and Society, Newcastle University, Newcastle, UK

 Springer

Introduction

Several authors, dating as far back as Darwin (1871), have suggested that within-population variation in skin color and condition might influence judgments of attractiveness, particularly of women's attractiveness. It has been argued, for example, that men prefer women with relatively pale skin as this is associated with youth (van den Berghe and Frost 1986, but see Fink et al. 2001) and that redness of cheeks could be a signal of health (Zahavi and Zahavi 1997). Indeed, skin redness has since been shown to be associated with both perceived health and attractiveness (Stephen et al. 2009, 2012) in facial rating studies. Furthermore, relating color attributes to attractiveness ratings, Fink et al. (2001) found that attractiveness was negatively correlated with variability in blue color space in women's faces, but was also predicted by homogeneity of contrast, suggesting that both color and textural cues influence judgments. In subsequent studies, Fink and colleagues showed that both homogeneity of skin color distribution and skin surface topography predicts perception of attractiveness, age and health in women's (Fink et al. 2006; Fink and Matts 2008) and men's faces (Fink et al. 2012).

The theoretical basis for such effects is that facial skin coloration and condition may provide cues of underlying mate quality, along with a suite of other facial traits including symmetry, averageness and sexual dimorphism (reviews in e.g. Fink and Penton-Voak 2002; Rhodes 2006; Roberts and Little 2008). While many of these traits may be inter-correlated if they are underpinned by a common currency of genetic quality (cf. Thornhill and Grammer 1999; see also Feinberg et al. 2005; Roberts and Little 2008), each may carry independent and additive contributions to such judgments (e.g. Saxton et al. 2009).

Consistent with this idea, skin condition accurately predicts overall facial attractiveness independently of information about facial shape (Jones et al. 2004a, b; Fink et al. 2006; Fink and Matts 2008). Furthermore, perceived health of facial skin patches cropped from the cheek area of digital images positively correlate with ratings of the attractiveness of men's faces (Jones et al. 2004a), and men with relatively symmetric faces were perceived as having healthier facial skin than those with asymmetric faces (Jones et al. 2004b). Color cues may also contain cues to underlying genetic indicators of health, such as heterozygosity at genes in the major histocompatibility complex, MHC, as healthiness ratings of skin patch images correlate with both MHC-heterozygosity and ratings of whole-face attractiveness (Roberts et al. 2005).

Such work has been supported by studies using a variety of novel methodological approaches to investigate the links between facial attractiveness, health and color cues. In studies in which participants manipulate color-calibrated facial images to optimize the appearance of health, raters appear to be sensitive to differences arising from variation in relative levels of oxygenated to deoxygenated blood color, suggesting this variation is important to perceptions of health in facial skin (Stephen et al. 2009; Re et al. 2011). Furthermore, changes in skin coloration also occur within as little as 1 h following experimental infection with a bacterial endotoxin, with changes varying in facial skin (becoming lighter and less red) compared to elsewhere on the body (Henderson et al. 2017).

Thus, color information in facial skin appears to be linked to the discrimination of both attractiveness and health. This is further evidenced by the fact that some studies investigating the link between attractiveness and actual measures of healthiness return

mixed results: those that do not find a link have often used monochrome images (Kalick et al. 1998; Rhodes et al. 2003), while those that find support for an attractiveness-healthiness relationship have used images presented in full color (e.g. Shackelford and Larsen 1999).

In this paper, we directly tested the effect of color on discrimination of facial attractiveness and skin healthiness in men's faces. We presented participants with sets of whole-face and skin patch images either in gray-scale monochrome or color. Image sets included several faces or skin patches perceived, respectively, to be attractive or healthy (from what we term "high quality men") and several faces or skin patches perceived to be unattractive or unhealthy (from "low quality men"). Our aim was specifically to compare the extent to which individual raters preferred the high quality individuals over those of low quality when images were presented in either color or monochrome. Our hypothesis was that, if color cues are important to discrimination of mate quality in faces, the difference in ratings awarded to high and low quality individuals should be more pronounced when images are presented in color. We carried out this experiment during a public science exhibition, which enabled us to obtain a large sample with a wide participant age range, and thus to also explore differences in ratings of high and low-quality individuals across different ages and between sexes.

Methods

Participants

The study was run during a public science exhibition, where visitors to the exhibition were invited to take part. A total of 409 individuals (284 women, 125 men) completed the task. Participants were aged between 8 and 82. Because the task involved ratings of adult attractiveness and tested the importance of color, we excluded from the analyses participants who were aged below 16 and 4 adults who reported that they were color-blind. This left 392 participants (278 women, 114 men), with a median age of 29.

Stimuli

We used images of whole faces and patches of skin cropped from the cheek area of whole face images. Men from whom images were taken were 92 students or staff at the University of Newcastle who provided informed consent.

Digital color photographs of the men's faces were taken under standard lighting conditions using a Nikon Coolpix 775 digital camera. Men were instructed to look directly at the camera and adopt a neutral expression. Each image was normalised on the inter-pupillary distance (Fink et al. 2001) and digitally masked so that only the face was visible, obscuring potentially confounding information about hairstyle and clothing with plain black shading (e.g. Roberts et al. 2004). Face images had a resolution of 1600 × 1200 pixels and were presented to raters with an on-screen face size of approximately 12 × 18 cm. Skin patch images were squares of skin from the right cheek of each man, the equivalent of 2.5 cm square, the lower edge being aligned with the bottom of the nose and the right-hand edge immediately next to the right nostril.

Following Jones et al. (2004a), skin images were magnified by 300% to facilitate ratings. These images have been used elsewhere (e.g. Roberts et al. 2005).

These images were presented to a reference panel of 50 women (age 18–49, mean = 23) on a liquid crystal display computer screen. Women rated facial attractiveness and skin healthiness using a 7-point rating scale (high scores = attractive, healthy). The two types of images were presented separately, with block order alternated between participants and within-block image order being fully randomised so that each participant saw images in a different order. Scores were standardised within individual raters to control for inter-rater variability in the use of rating scales, and mean standardised scores were then calculated for each image (there was high inter-rater agreement for both faces and skin patches: Cronbach α = .95 and .94, respectively; see Roberts et al. 2005 for further details).

For this experiment, we then selected six men from each extreme of the distributions of facial attractiveness scores and skin healthiness scores. Mean ($\pm$ s.e.) standardized scores for the six attractive and unattractive faces were $.827 \pm .31$ and $-.570 \pm .11$, respectively. The mean scores for the high and low skin healthiness images were $.716 \pm .11$ and $-.738 \pm .07$. Two men were represented in both face and skin images, the remaining ten were different.

Procedure

The two sets of images (12 faces, 12 skin patches) were presented to each rater. Each set was printed onto a large poster, with images set out in a 3 × 4 grid and numbered 1–12. Within each set, odd-numbered images were low-quality men and even-numbered images were high-quality men. Each set was printed in both full color and monochrome.

During the exhibition, we invited visitors to our stand to take part in our study. If they agreed, the task was fully explained and they were given a checksheet to complete on their own, on which was marked two 3 × 4 grids (marked 'Faces' and 'Skin'), each grid cell being numbered in the same scheme as on the poster, to minimize the possibility of scoring errors. The checksheets also had details of the rating scales to be used and checkboxes to indicate the rater's age and sex, and a box to indicate whether they were color blind (no other personal information was requested). Raters judged the faces first, followed by the skin patches, again scoring them for attractiveness or healthiness on 7-point scales. For each rater, we summed scores for the 6 images in each set (i.e. those perceived to be attractive/less unattractive, or healthy/less healthy) for analysis. Only the color or monochrome conditions were visible to visitors at any one time, and these were alternated across morning, afternoon and evening sessions over the three days of the exhibition. Each image presented was seen at approximately 10 × 15 cm on the poster. Of those included in the analyses (see (a) above) 194 individuals (140 women, 54 men) did the rating tasks in the color condition and 198 (138 women, 60 men) in the monochrome condition.

Analysis

Data and analyses are available at https://osf.io/gdekz/. We used a repeated measures ANOVA, with Image type (face, skin) and Quality (high, low) as the within-subjects variables, Sex and Color condition (color, monochrome) as between-subjects factors, and Age as a covariate.

For post hoc investigation of correlations with age, we also calculated, for each image type, the ratio of the scores awarded to high quality and low quality groups (sum of scores for high quality men divided by sum of scores for low quality men), which provides an indication of the degree to which each individual discriminated between individuals classed by the reference panel as high or low quality (hereafter termed 'discrimination quotient').

Results

Repeated measures ANOVA revealed a significant three-way interaction between Image type, Quality and Color condition (Table 1), such that the ratings given to the skin (but not the face) images of low quality individuals were particularly low when viewed in color compared to monochrome (post hoc t test: $t(390) = 3.37$, $p = 0.001$) and there was little effect of color condition on high quality individuals ($t(390) = 1.12$, $p = 0.262$; see Fig. 1a). This result is consistent with the hypothesis that color cues are important in discriminating facial differences associated with markers of low quality in men's faces.

The analyses also demonstrated a significant main effect of Image type, in which higher scores were recorded for skin than face images (Fig. 1a). The significant Image type x Color condition interaction showed that the difference in scores associated with Image type was also modulated by whether the images were in color or not, such that raters awarded higher scores overall to faces in color than monochrome, and higher scores overall to skin images in monochrome than color (Fig. 1b).

Table 1 Results of repeated measures ANOVA, showing effects on discrimination of face and skin quality recorded from 392 participants

Effect	F	p
Within-subjects		
Image type	98.60	<.001
Image type x Color condition	7.50	.006
Image type x Age	22.85	<.001
Quality x Age	11.35	.001
Quality x Sex	4.13	.043
Image type x Quality x Color condition	3.88	.049
Between-subjects		
Color condition	.015	.902
Age	3.51	.062
Sex	2.92	.088
Color condition x Sex	1.84	.176

Degrees of freedom = 1, 387 in each case. A significant main effect of Quality ($p << 0.001$) is not included since it simply retrieves the experimental design. Other non-significant within-subjects effects and interaction terms have been omitted
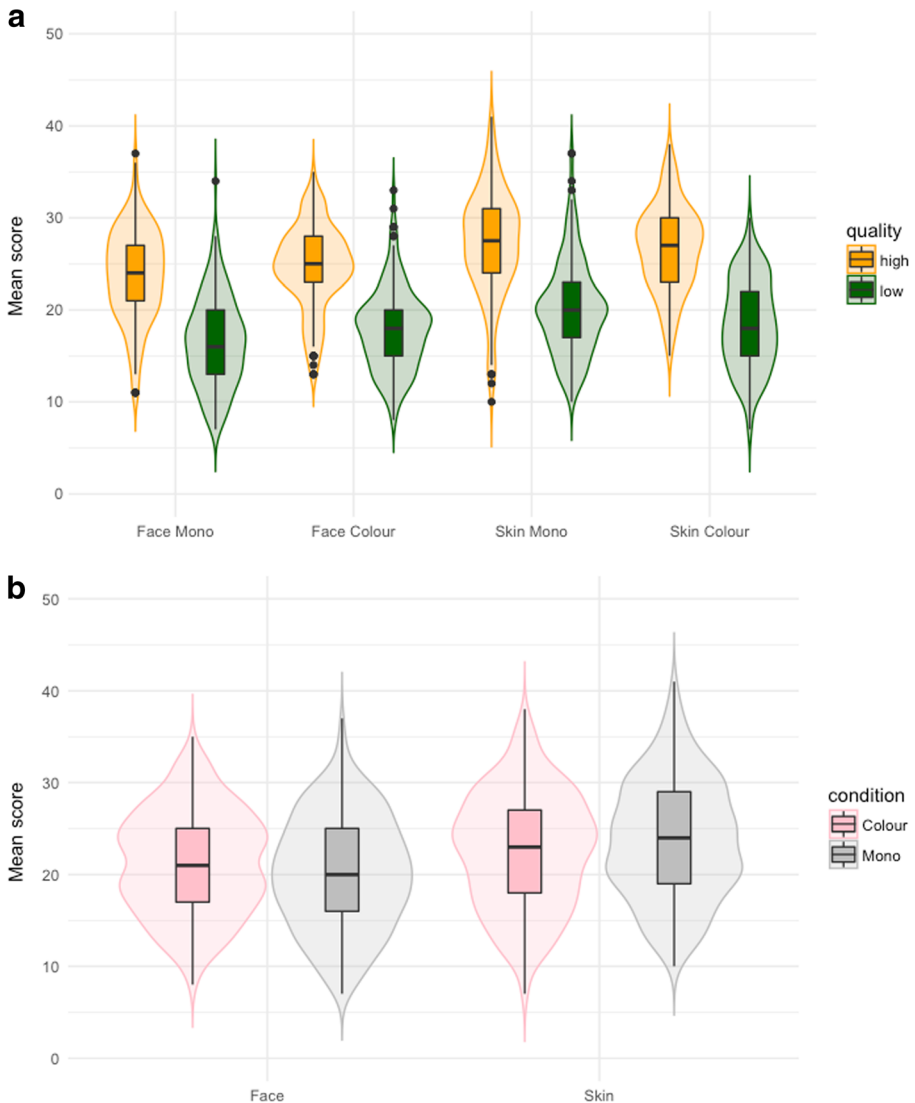
Fig. 1 Facial attractiveness and skin healthiness scores of 392 participants. **a** Differences in scores awarded to high-quality and low-quality faces and skin patches when presented in color or monochrome. **b** Scores awarded by all raters in color and monochrome. Plots show standard boxplots superimposed on violin plots, which indicate the full distribution of the data: the box spans the first to third quartile, with a line at the median, and the whiskers extend to the largest values no further than 1.5 times the interquartile range above the third quartile or below the first quartile. Data points outside this range are plotted individually

The significant Quality x Sex interaction term indicated that, while both sexes awarded higher scores to high quality images, women awarded lower scores than men to low quality individuals (Fig. 2a).

The analysis also revealed significant effects of rater Age, through interactions with both Image type and Quality (all statistically significant effects listed in Table 1 remained such when the model was repeated without this covariate). To further
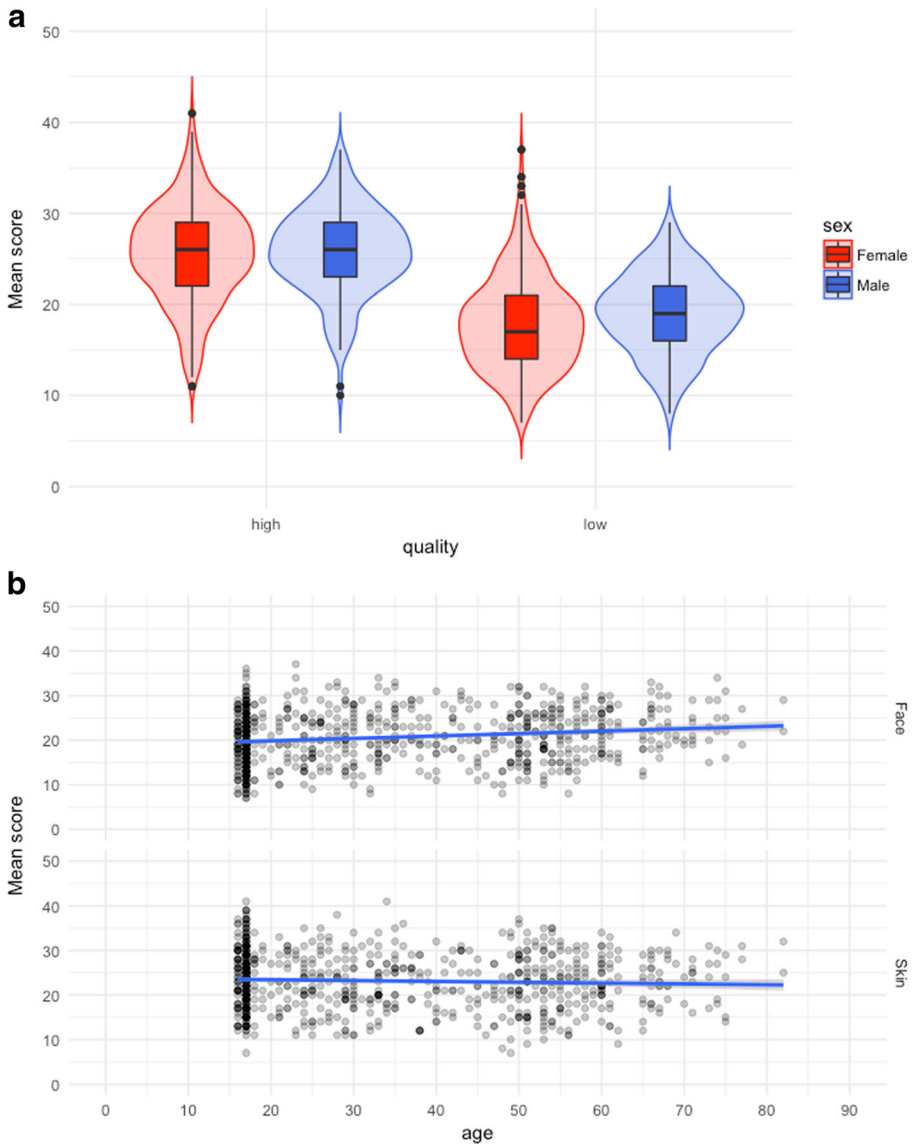
Fig. 2 Differences in scores awarded by male and female raters to images (data show summed scores for faces and skin) of high and low quality individuals. **b** Relationships between rater age and scores awarded to faces and skin patches

investigate the interaction with Image type, we correlated age with the sum of face or skin scores given to the twelve images. Age was positively correlated with face ($r_s = .248$, $n = 392$, $p < .001$) but not skin scores ($r_s = -.092$, n = 392, $p = .070$), that is, younger participants gave lower scores to the face, but not the skin, images (Fig. 2b). To investigate the interaction with Quality, we correlated discrimination quotients against age for both face and skin images. In both sexes, discrimination quotients were negatively correlated with age for face judgments (Fig. 3; men: $r_s = -.298$, $n = 114$,
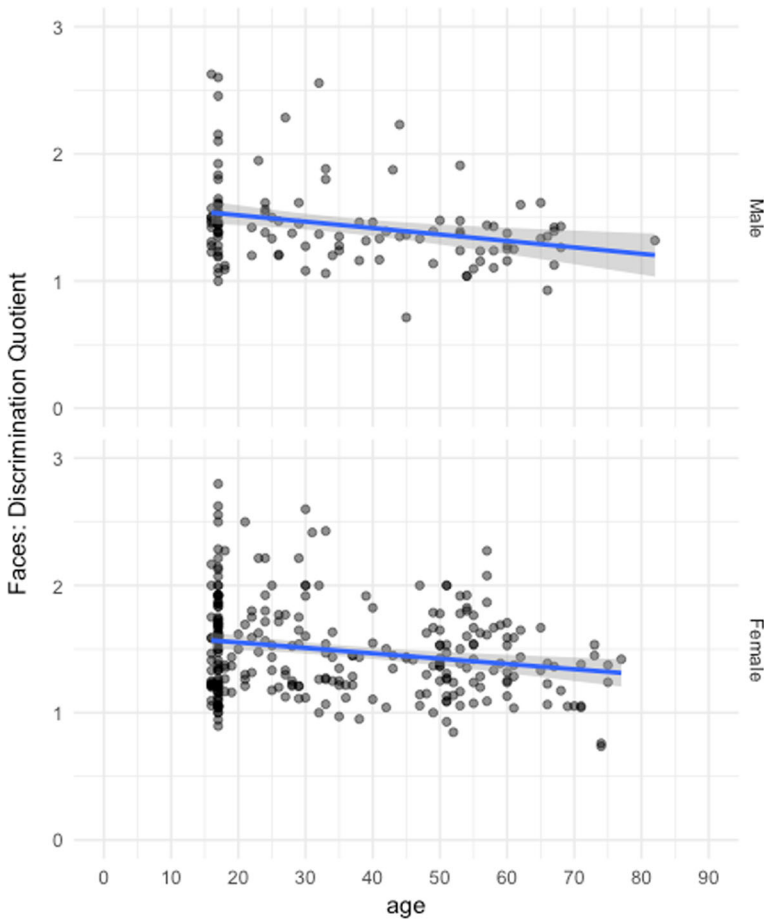
Fig. 3 Relationship between age and discrimination quotient for attractive and unattractive faces in male (upper panel) and female raters (lower panel). Higher quotients indicate clearer discrimination of attractiveness (relative to the ratings of the reference panel)

$p = .001$; women: $r_s = -.137$, $n = 278$, $p = 0.023$) but not for skin ratings (men: $r_s = .012$, n = 114, $p = .90$; women: $r_s = 0$, n = 278, $p = 1$).

None of the between-subjects effects were significant, although there were tendencies for women to award lower scores than men ($p = .088$) and for younger participants to award lower scores than older participants ($p = .062$).

Finally, we investigated how individual preferences for quality were related across the two image presentations. Discrimination quotients obtained in the face test significantly correlated with quotients from the skin test in women ($r_s = .156$, n = 278, $p = .009$) but not men ($r_s = .049$, n = 114, $p = .61$).

Discussion

Our results show that discrimination of high and low quality images for health varies with color information for the skin patches but that color had limited impact on

discrimination of attractiveness for the whole faces. The significant Image type x Quality x Color condition interaction suggests that higher scores are awarded to high quality individuals compared with low quality individuals when presented in the color condition than in monochrome, but only for the skin images. Our results are consistent with recent findings that skin color and the distribution of skin coloration are important in judgements of mate quality (Fink et al. 2001, 2006; Jones et al. 2004a, b; Stephen et al. 2009). Color information is not the only cue to quality, however, since high quality individuals received higher scores than low quality individuals, even when judged in monochrome, in both face and skin image presentations. Even within images of small patches of skin, monochrome images retain cues of skin condition, including information about tone, contrast and homogeneity (see Fink et al. 2001). Nonetheless, color cues appear to provide additive information that contributes to perception of skin condition, and our results suggest that these may be important in distinguishing underlying quality in faces.

One issue with our design is that the initial ratings, on which we based our classification of stimuli into high and low quality, were made for color images. It is then possible that the ratings of images when seen in monochrome might have been influenced by simply being different from those initially rated. However, if this was a major influence on the results obtained, we would expect to see a significant Quality x Color Condition interaction. This interaction was not significant (F (1387) = 2.03, $p = 0.16$), suggesting that there was no overall effect of simple stimulus consistency (or otherwise) in the initial and subsequent ratings. Furthermore, the variation according to rater age and sex, and the significant interactions with image type, suggest that there are real differences in the perception of the images when presented in color and in monochrome.

Indeed, in this large sample of participants of variable age, our results also revealed several significant effects on the relative ratings of high and low quality men. On average, women and younger participants tended to give lower scores than men and older participants, respectively, indicating that there is a tendency for the two former groups to use the lower ends of the rating scale, although these main effects were not significant. However, women did award lower scores to low quality individuals and this was not simply due to differential scale use because there was no sex difference in scores awarded to high quality men. Women may thus be better at detecting low-quality men than men are. Cross-task correlations indicated that women who scored highly on the face task also scored highly on the skin task, indicating robustness for within-sex discrimination of mate quality. No cross-task correlations were significant for male participants, however. Furthermore, the difference between scores awarded to high and low quality individuals was inversely related with age. The pattern of these results suggests that younger individuals and women appeared to be more able to discriminate quality in their rating than older participants and men.

We cannot be certain that these differences with respect to sex and age are a result of differences in ability to discriminate quality, or whether alternatively they reflect differences in preference (e.g. Ling and Hurlbert 2011). For example, because the task involved discriminating between high and low attractive/healthy images as judged by young women, it is possible that differences in preferences between younger and older women could lead to lower discrimination for older women. Alternatively, older women might have been less motivated to do the task because the age gap between

themselves and the stimuli was relatively large. Similarly, men may have approached the task differently or with less motivation than women. However motivational or preference differences cannot account for the effects due to color or monochrome stimulus presentation. We therefore think it more likely that these effects are due to differences in perceptual acuity associated with gender and age; research suggests that younger people and women have better perceptual acuity than older people and men respectively (Fiorentini et al. 1996; Bimler et al. 2004), and these differences in perceptual acuity may play an important role in the between-groups differences in quality discrimination observed in the current study.

Our results indicate that color information in facial skin plays an important role in the discrimination of health in humans. Women in particular appeared to be more successful in discriminating high and low quality stimuli. Our results are consistent with, and provide further evidence for, the idea that sexual selection through skin color discrimination may have contributed to the evolution of primate color vision in females (Waitt et al. 2003; Changizi et al. 2006). An enhanced ability to detect and respond to color cues in facial skin is likely to have fitness effects through more accurate discrimination of mate quality and health, potentially providing increased indirect and direct benefits to women.

Compliance with Ethical Standards

Ethical Statement   The research described here was carried out with approval from the Ethics Committee of the Newcastle and North Tyneside NHS Trust and has therefore been performed in accordance with the ethical standards laid down in the Declaration of Helsinki.

References

Bimler, D. L., Kirkland, J., & Jameson, K. A. (2004). Quantifying variations in personal color spaces: Are there sex differences in color vision? *Color Research and Application, 29*, 128–134.

Changizi, M. A., Zhang, Q., & Shimojo, S. (2006). Bare skin, blood and the evolution of primate colour vision. *Biology Letters, 2*, 217–221.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. London: John Murray.

Feinberg, D. R., Jones, B. C., DeBruine, L. M., Moore, F. R., Law Smith, M. J., Cornwell, R. E., Tiddeman, B. P., Boothroyd, L. G., & Perrett, D. I. (2005). The voice and face of woman: One ornament that signals quality? *Evolution and Human Behavior, 26*, 398–408.

Fink, B., & Matts, P. J. (2008). The effects of skin colour distribution and topography cues on the perception of female facial age and health. *Journal of the European Academy of Dermatology and Venereology, 22*, 493–498.

Fink, B., & Penton-Voak, I. S. (2002). Evolutionary psychology of facial attractiveness. *Current Directions in Psychological Science, 11*, 154–158.

Fink, B., Grammar, K., & Thornhill, R. (2001). Human (*Homo sapiens*) facial attractiveness in relation to skin texture and color. *Journal of Comparative Psychology, 115*, 92–99.

Fink, B., Grammer, K., & Matts, P. J. (2006). Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evolution and Human Behavior, 27*, 433–442.

Fink, B., Bunse, L., Matts, P. J., & D'Emiliano, D. (2012). Visible skin colouration predicts perception of male facial age, health and attractiveness. *International Journal of Cosmetic Science, 34*, 307–310.

Fiorentini, A., Porciatti, V., Morrone, M. C., & Burr, D. C. (1996). Visual ageing: Unspecific decline of the responses to luminance and colour. *Vision Research, 36*, 3557–3566.

Henderson, A. J., Lasselin, J., Lekander, M., Olsson, M. J., Powis, S. J., Axelsson, J., & Perrett, D. J. (2017). Skin colour changes during experimentally-induced sickness. *Brain, Behavior, and Immunity, 60*, 312–318.

Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2004a). When facial attractiveness is only skin deep. *Perception, 33*, 569–576.

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

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.

Ling, Z., & Hurlbert, A. (2011). Age-dependence of colour preference in the U.K. population. In C. P. Biggam, C. Hough, C. Kay, & D. R. Simmons (Eds.), *New directions in colour studies* (pp. 347–360). Amsterdam: John Benjamins Publishing Company.

Re, D. E., Whitehead, R. D., Xiao, D., & Perrett, D. I. (2011). Oxygenated-blood colour change thresholds for perceived facial redness, health, and attractiveness. *PLoS One, 6*, e17859.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57*, 199–226.

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

Roberts, S. C., & Little, A. C. (2008). Good genes, complementary genes and human mate choice. *Genetica, 132*, 309–321.

Roberts, S. C., Havlicek, J., Flegr, J., Hruskova, M., Little, A. C., Jones, B. C., Perrett, D. I., & Petrie, M. (2004). Female facial attractiveness increases during the fertile phase of the menstrual cycle. *Proceedings of the Royal Society London B (Suppl.), 271*, S270–S272.

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

Saxton, T. K., Burriss, R. P., Murray, A. K., Rowland, H. M., & Roberts, S. C. (2009). Face, body and speech cues independently predict judgments of attractiveness. *Journal of Evolutionary Psychology, 7*, 23–35.

Shackelford, T. K., & Larsen, R. J. (1999). Facial attractiveness and physical health. *Evolution and Human Behavior, 20*, 71–76.

Stephen, I. D., Coetzee, V., Law Smith, M., & Perrett, D. I. (2009). Skin blood perfusion and oxygenation colour affect perceived human health. *PLoS One, 4*, e5083.

Stephen, I. D., Oldham, F. H., Perrett, D. I., & Barton, R. A. (2012). Redness enhances perceived aggression, dominance and attractiveness in men's faces. *Evolutionary Psychology, 10*, 562–572.

Thornhill, R., & Grammer, K. (1999). The body and face of woman: One ornament that signals quality? *Evolution and Human Behavior, 20*, 105–120.

Van den Berghe, P. L., & Frost, P. (1986). Skin color preference, sexual dimorphism and sexual selection: A case of gene-culture co-evolution? *Ethnic and Racial Studies, 9*, 87–118.

Waitt, C., Little, A. C., Wolfensohn, S., Honess, P., Brown, A. P., Buchanan-Smith, H. M., & Perrett, D. I. (2003). Evidence from rhesus macaques suggests that male coloration plays a role in female primate mate choice. *Proceedings of the Royal Society London B (Suppl.), 270*, S144–S146.

Zahavi, A., & Zahavi, A. (1997). *The handicap principle: a missing piece of Darwin's puzzle*. Oxford: Oxford University Press.