



OPEN

## Colour information biases facial age estimation and reduces inter-observer variability

Jean Y. J. Hsieh, W. Paul Boyce, Erin Goddard & Colin W. G. Clifford

Age estimation is a perceptual task that people perform automatically and effortlessly on a daily basis. Colour has been identified as one of the facial cues that contributes to age perception. To investigate further the role of colour in age perception, we manipulated the chromatic content of facial images holistically. In Experiment 1, images were shown in colour or grey scale; in Experiment 2, images were shown with red–green contrast increased or decreased; in Experiment 3, images were shown with modified yellow–blue contrast. We examined whether the presence of chromatic information biases the perception of age and/or affects inter-observer variability in age judgements, and whether specific chromatic information affects the perception of age. We found that the same face tended to be judged as younger with increased red–green contrast compared to decreased red–green contrast, suggesting that red–green contrast directly affects age perception. Inter-observer variability in age ratings was significantly lower when participants were asked to rate colour compared with grey scale versions of images. This finding indicates that colour carries information useful cues for age estimation.

Age is an important social attribute, making age estimation an ecologically relevant component of face perception. It determines rights, responsibilities and even social status<sup>1</sup>. However, perceived age does not always reflect an individual's biological age, since facial appearance is also influenced by environmental and lifestyle factors<sup>1</sup>. A longitudinal study of Danish twins reported that perceived age is a better biomarker of health than chronological age due to the influence of lifestyle and social factors<sup>2</sup>. This social relevance is also reflected in most people wanting to be perceived as younger, since looking youthful is associated with health and attractiveness<sup>3–5</sup>. Additionally, researchers have also found that attractive people are judged and treated more positively than unattractive people<sup>6</sup>.

The eyes and lips have been reported to be particularly relevant to estimating age compared to other facial features. The surrounding area, shape, and size of the eyes and sclera colour have been shown to help people make a judgement of age<sup>7</sup>. Lips are also potentially a reliable cue to age since their volume<sup>8</sup> and redness<sup>9</sup> decrease significantly with age. Gunn et al.<sup>8</sup> conducted a study to investigate the factors that contribute to a youthful appearance. They collected ratings from participants on perceived age, attractiveness, and health based on images of females ranging in age from 16 to 96 years old. The study found that lip height, skin wrinkling, and hair greying were all significant factors that were independently associated with how old the women appeared for their age.

Besides facial features, there are other surface-based cues associated with perceived age<sup>10–12</sup>. These cues can be divided into two categories. The first category, known as “pigmentation,” includes variations in skin tone and texture<sup>10</sup>. Faces with uneven skin tone are more likely to be judged as older compared to those with even skin tone<sup>8</sup>. Sun exposure is also associated with a higher perceived age<sup>13,14</sup> as it causes uneven skin tone<sup>15</sup> and wrinkling<sup>8</sup>. The second category is the pattern of shading and shadow created by the three-dimensional (3D) shape of the face<sup>10,11,16</sup>. For example, nose shape changes significantly with age<sup>17</sup>, but it is not yet clear if people use the nose as a cue to estimate age. Empirically, digitally reducing skin wrinkles, skin sagging, or skin unevenness has been shown to give a more youthful appearance<sup>15,18</sup>.

Colour information of the face has not only been reported to be important in the perception of age<sup>15,19</sup>, but also to contribute significantly to perceptions of sex/gender and ethnicity<sup>16,20</sup>, health<sup>7,15,21</sup>, attractiveness<sup>21,22</sup>, and identity<sup>23</sup>. When investigating age perception, Fink et al.<sup>18</sup> applied skin colour distributions from different ages to identical 3D shape-standardised female faces. Participants rated age, health, and attractiveness of the 3D rendered faces. Faces with younger surface-based information were perceived as younger and received significantly higher ratings for attractiveness and health. Analysis of the images revealed that the skin colour distribution of younger faces was more even compared to older faces. Fink et al.<sup>18</sup> suggested that skin colour distributions provide information independent of facial form and skin surface topography that contributes to

School of Psychology, UNSW Sydney, Sydney, NSW 2052, Australia. email: colin.clifford@unsw.edu.au

the perception of female facial age and judgements of attractiveness and health. However, further research is needed to determine if these findings can be applied to actual face images. Therefore, we intend to explore this topic further in our proposed study.

Facial contrast, defined as the contrast between facial features such as the eyes and lips and the surrounding skin, is a potential cue for age estimation<sup>24,25</sup>. It should be noted that this is different from ‘skin contrast’ as defined by other researchers<sup>15</sup>. Studies have found that female faces have greater facial contrast than male faces, and facial contrast plays an important role in sex/gender classification and attractiveness<sup>22,25</sup>. Porcheron et al.<sup>24</sup> proposed that facial contrast decreases with age and serves as a cue for identifying age. They measured the facial contrast of Caucasian female faces within the CIE L\*a\*b\* colour space (Lab space) and asked participants to estimate age from the original and contrast-manipulated images. Female face images were judged to be significantly younger when shown with greater a\* (red–green) contrast around the mouth, greater luminance contrast around the eyes, or greater luminance contrast around the eyebrows. It is worth noting that these estimates of facial contrast did not distinguish between skin pigmentation and luminance cues related to topographic factors, and that measuring facial contrast required individual labelling of facial regions of interest, which could make standardisation difficult. Overall, these findings suggested that colour contrast both within features/skin and between features is a potential cue to age.

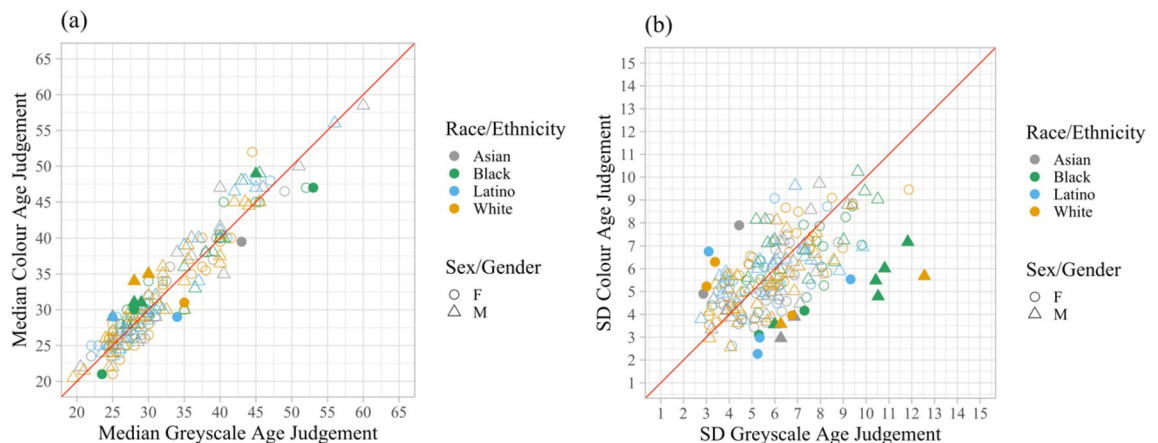
A key challenge for facial age estimation research is the fact that faces are complex stimuli, and the space of possible variables driving age perception is large. A common approach is to manipulate specific facial features and observe changes in age ratings to identify relevant features<sup>18,24</sup>. While this approach is valid, it relies on the researcher first identifying candidate facial features, leaving open the possibility that more complex or subtle features are missed. To circumvent this limitation, we proposed to manipulate the colour contrast of entire face images using the three dimensions of Lab space, as opposed to manipulating specific facial features. We hypothesised that faces would be perceived as younger while increasing both the level of redness<sup>18,24</sup> and yellowness. Yellowness is associated with skin health<sup>26,27</sup> and health is a highly associated with age<sup>2</sup>. Whether colour rather than grey scale presentation influences the perception of age is unknown.

## Results

**Experiment 1.** We tested whether colour rather than grey scale presentation systematically biased the perception of age, or affected inter-observer variability in age judgements. We conducted data cleaning before running any analyses. There were 7800 ratings collected from 39 included participants. From these, we excluded ratings under 13 years since images from the Chicago face dataset (CFD) are all images of adults. We also excluded absolute deviations from the CFD average rated age larger than 25. By doing so, 32 of 7800 trial ratings were excluded.

To establish if presentation in colour versus grey scale systematically biases ratings of age, we compared the median rating of each image in colour ( $M = 32.54$ ,  $SD = 9.47$ ) and grey scale ( $M = 32.37$ ,  $SD = 9.64$ ). A paired-sample  $t$  test indicated there was no significant difference between image conditions ( $t_{199} = 1.37$ ,  $p = 0.19$ ; Fig. 1a).

The standard deviation ( $SD$ ) in ratings gives a measure of inter-observer variability, which reflects the degree of agreement among independent observers, for each image in each of the two conditions. Specifically, the lower the  $SD$ , the lower the variability and the higher the agreement. To investigate whether image condition affected inter-observer variability, we compared  $SD$ s of the rating of each image in the colour ( $M = 5.74$ ,  $SE = 1.61$ ) and



**Figure 1.** Rating comparison in Experiment 1. *Note:* Each of the 200 data points represents a particular stimulus identity. The red line is a reference line of equality. The legends describe the characteristics of the images. (a) Median age ratings. Points above the red line are images that have older ratings in the colour condition compared to the grey scale condition. Filled points are images that have significantly different mean age ratings between the two conditions (independent samples  $t$  tests, uncorrected for multiple comparisons). No cases exceed the false discovery rate (FDR) threshold for  $q < 0.05$  when we corrected these  $p$  values for multiple comparisons across images<sup>28</sup>. (b) Between participants  $SD$  of age ratings. Points above the red line are images that have higher  $SD$  in the colour condition compared to the grey scale condition. Filled points are images that have significantly different  $SD$ s between the two conditions ( $F$ -tests, uncorrected for multiple comparisons). 4 cases exceed the false discovery rate (FDR) threshold for  $q < 0.05$ .

grey scale ( $M=6.06$ ,  $SE=1.96$ ) conditions. A paired-sample  $t$  test revealed that inter-observer variability was significantly higher in the grey scale condition than in the colour condition ( $t_{199} = -2.65$ ,  $p=0.009$ , Fig. 1b). Inspection of the data indicates an ethnicity effect, such that there is a disproportionately high number of black faces among those that have a significantly higher SD in grey scale conditions (Fig. 1b). A post-hoc chi-squared test confirmed that the ethnicities of these images did indeed differ from the expected distribution,  $X^2(3, N=200) = 8.63$ ,  $p=0.05$ .

**Experiment 2.** We further investigated if facial age estimation is affected by specific chromatic information, specifically red–green contrast. There was a total of 8200 ratings collected from 41 included participants. Using the same data cleaning criteria as in Experiment 1, we excluded 20 of 8200 ratings. To clarify if presentation in increased versus decreased red–green contrast systematically biases ratings of age, we compared the median rating of each image in the increased ( $M=31.99$ ,  $SD=9.31$ ) and decreased red–green contrast conditions ( $M=32.57$ ,  $SD=9.41$ ) with a paired-sample  $t$  test ( $t_{199}=4.46$ ,  $p<0.001$ ). There was a significant difference between image conditions such that faces tend to be rated younger when presented in the higher red–green contrast condition (Fig. 2a).

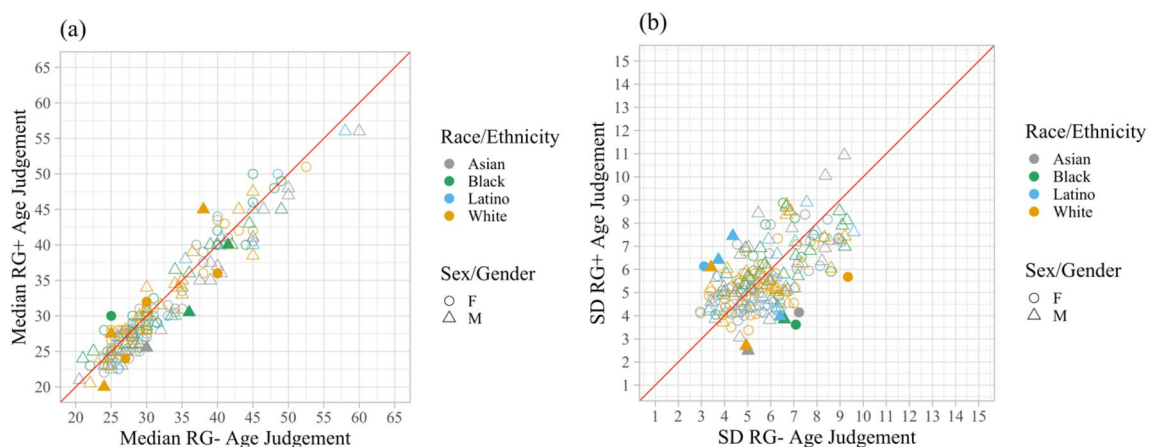
We also investigated whether image conditions affected inter-observer variability. A comparison of the  $SD$  of the rating of each image in both lower red–green contrast ( $M=5.78$ ,  $SE=1.54$ ) and higher red–green contrast ( $M=5.64$ ,  $SE=1.42$ ) condition with a paired-sample  $t$  test ( $t_{199}=1.36$ ,  $p=0.17$ ) revealed no significant difference for inter-observer variability (Fig. 2b).

**Experiment 3.** We further investigated if facial age estimation is affected by yellow–blue contrast. There was a total of 8600 ratings collected from 43 included participants. Using the same data cleaning criteria as in Experiment 1, we excluded 57 of 8600 ratings. Comparing the increased ( $M=33.18$ ,  $SD=9.91$ ) and decreased yellow–blue contrast conditions ( $M=33.28$ ,  $SD=10.13$ ), the paired-sample  $t$  test on the median rated age ( $t_{199} = -0.50$ ,  $p=0.62$ ) was not significant (Fig. 3a).

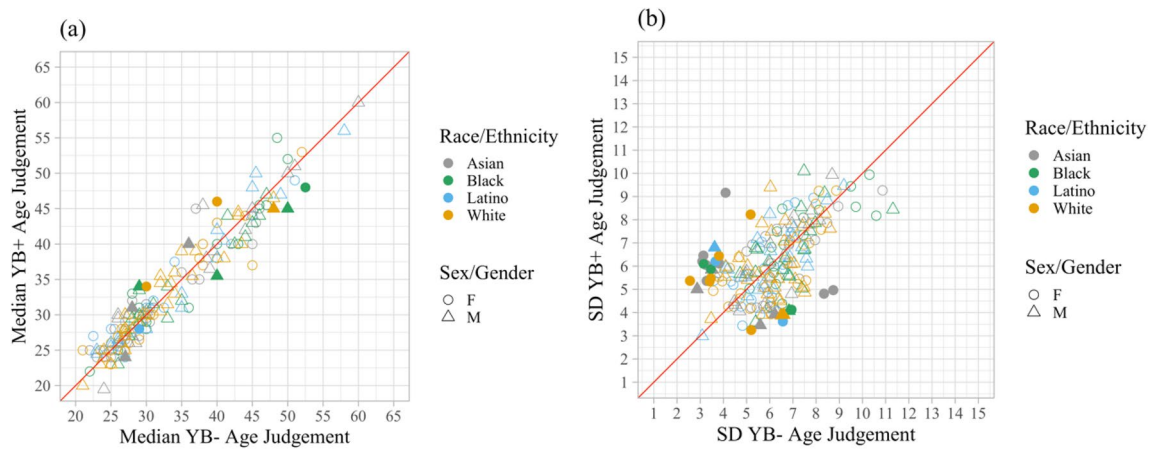
We also investigated whether image conditions affected inter-observer variability. A comparison of the  $SD$  of the rating of each image in the lower yellow–blue contrast ( $M=6.15$ ,  $SE=1.62$ ) and higher yellow–blue contrast ( $M=6.18$ ,  $SE=1.58$ ) conditions with a paired-sample  $t$  test ( $t_{199} = -0.27$ ,  $p=0.79$ ) revealed no significant difference for inter-observer variability (Fig. 3b).

## Discussion

The aim of the current study was to establish whether colour and chromatic contrast contribute to facial age judgements. We examined whether colour information biases the perception of age, or affects inter-observer variability in age judgements. Additionally, we also examined which specific chromatic channel affects the perception of age. In Experiment 1, participants were asked to rate images in either colour or grey scale version. The inter-observer  $SD$  of ratings for individual images was found to be significantly lower for colour presentation. This implies that the participants' responses were more consistent when the images were presented in colour. Comparing median ratings within Experiment 2, we found that face images with higher red–green contrast



**Figure 2.** Rating comparison in Experiment 2. *Note:* Each of the 200 data points represents a particular stimulus identity. The red line is a reference line of equality. The legends describe the characteristics of the images. (a) Median age ratings. Points above the red line are images that have older ratings in the increased red–green contrast (RG+) condition compared to the decreased red–green contrast (RG–) condition. Filled points are images that have significantly different mean age ratings between the two conditions (independent samples  $t$  tests, uncorrected for multiple comparisons). No cases exceed the false discovery rate (FDR) threshold for  $q<0.05$  when we corrected these  $p$  values for multiple comparisons across images. (b) Between participants  $SD$  of age ratings. Points above the red line are images that have higher  $SD$  in the RG+ condition compared to the RG– condition. Filled points are images that have significantly different  $SD$ s between the two conditions ( $F$ -tests, uncorrected for multiple comparisons). No cases exceed the false discovery rate (FDR) threshold for  $q<0.05$  when we corrected these  $p$  values for multiple comparisons across images.

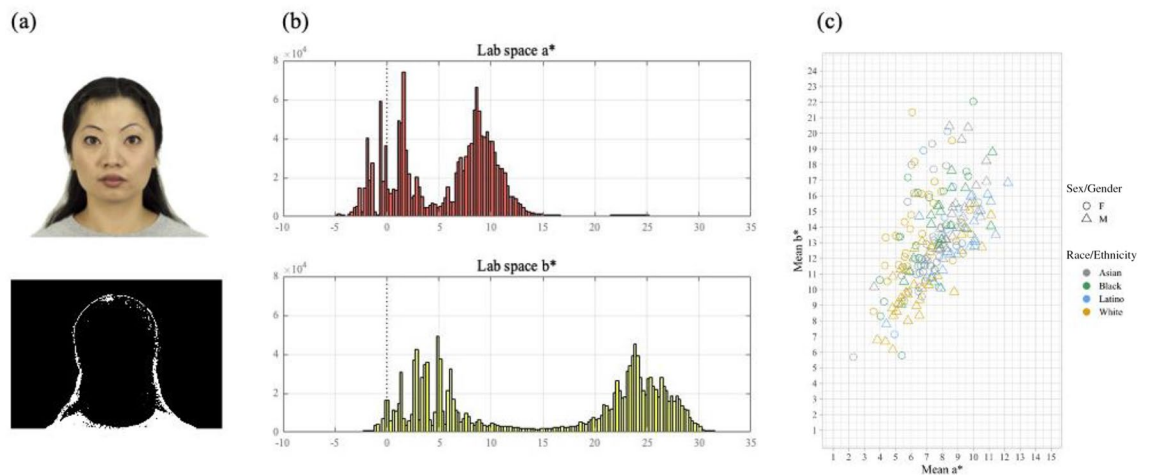


**Figure 3.** Rating comparison in Experiment 3. *Note:* Each of the 200 data points represents a particular stimulus identity. The red line is a reference line of equality. The legends describe the characteristics of the images. (a) Median age ratings. Points above the red line are images that have older ratings in the higher yellow–blue contrast (YB+) condition compared to the lower yellow–blue contrast (YB–) condition. Filled points are images that have significantly different mean age ratings between the two conditions (independent samples *t* tests, uncorrected for multiple comparisons). No cases exceed the false discovery rate (FDR) threshold for  $q < 0.05$  when we corrected these *p* values for multiple comparisons across images. (b) Between participants *SD* of age ratings. Points above the red line are images that have higher *SD* in the YB+ condition compared to the YB– condition. Filled points are images that have significantly different *SD*s between the two conditions (*F*-tests, uncorrected for multiple comparisons). No cases exceed the false discovery rate (FDR) threshold for  $q < 0.05$  when we corrected these *p* values for multiple comparisons across images.

were rated significantly younger. This finding demonstrates that red–green contrast can directly influence the perception of age such that the same face tends to be rated as younger with increased compared to decreased red–green contrast. No significant effect was found while manipulating yellow–blue contrast in Experiment 3.

Recent studies have suggested that redness is associated with the perception of various facial attributes, such as age, attractiveness, and health. However, it is unclear whether hues<sup>22</sup> or colour contrasts<sup>24</sup> are driving the results. Representing face images in Lab space, it is evident that there is a predominance of red ( $> 0$ ) and yellow ( $> 0$ ); there is very little green and even less blue (Fig. 4). Across 200 images, 81.64% of pixels had  $a^* > 0$  and 91.52% of pixels had  $b^* > 0$ . Thus, by manipulating different Lab space contrasts, we were mainly manipulating “redness” and “yellowness” in the facial images. Specifically, increasing red–green contrast increased redness while increasing yellow–blue contrast increased yellowness.

In Experiment 1, we compared the median and *SD* of ratings between colour and grey scale images. We found that the inter-observer *SD* of ratings for individual images was significantly lower for colour presentation. That is, when images were in colour, the participants’ responses were less varied, compared to when the images were presented in grey scale. While we found no evidence that the simple presence or absence of colour systematically



**Figure 4.** Face colour in Lab space. *Note:* The example image was drawn from the Chicago Face Database<sup>29</sup>. (a) Example stimulus image and map of pixels where  $a^* < 0$ . (b) Pixel counts of Lab space  $a^*$  and  $b^*$  values for this image. Most pixels are red ( $a^* > 0$ ) and yellow ( $b^* > 0$ ). (c) Mean chromaticity of face images in Lab space. Each of the 200 data points represents a particular stimulus image. All have mean  $a^*$  and mean  $b^*$  greater than zero.

affects our perception of age, this finding indicates that chromatic information provides extra information about age, beyond that carried by the achromatic content of the image, and that this is used when making age judgements, in a way that is broadly consistent across observers.

A chi-squared test revealed an imbalance in the ethnicity of faces that have significantly different *SDs* between the two conditions (Fig. 1b), primarily due to a higher number of black faces. All black faces where the difference was significant had a higher *SD* in the grey scale condition, suggesting that colour may influence the perceived age of black faces more than other ethnicities. Additionally, these findings align somewhat with the previously reported ‘other-race effect’, where age estimation is more accurate for individuals of the same ethnicity<sup>30</sup> and higher familiarity<sup>30,31</sup>. A study in Belgium revealed that Caucasian participants performed better in evaluating ages for Caucasian faces compared to African faces, while African participants performed equally well with both face types, possibly due to the African participants’ time of residence in Belgium<sup>30</sup>. Although our experiment did not control participant ethnicity, the observed differences may be attributable to the unfamiliarity of our primarily domestic student participants with Afro-American black faces, which are underrepresented in Sydney’s otherwise diverse racial profile. However, it is notable that in our study we did not observe a significant other-race effect per se, but rather a race-specific decrement in performance when colour was removed from the images, suggesting that processing the age of other-race faces may be disproportionately affected when image information is impoverished. Further investigation should explore the impact of observers’ ethnicity and time of residence in Australia, as well as other means of image degradation, for a more comprehensive understanding.

In Experiment 2, we found that increasing red–green contrast (effectively increasing redness) made faces look younger, which complements previous research. Previous studies have shown that increasing the redness of particular facial features (e.g., lips, skin) makes faces look younger<sup>14,15,18,24</sup>. While youthfulness is often associated with health and attractiveness, increasing redness can also make faces look healthier<sup>32,33</sup> or more attractive<sup>15,18,22</sup>. Facial redness is an indicator of the level of blood oxygenation, which may be interpreted by observers as a cue to underlying physiological health or quality<sup>21,32</sup>.

In Experiment 3, we did not find any evidence that yellow–blue contrast (effectively increasing yellowness) is associated with facial age perception. Furthermore, the variance of ratings between the two conditions did not exhibit any significant differences. Despite evidence that increased yellowness is related to health, which is itself strongly associated with age, we found no evidence that yellowness has any impact on the perception of age in faces. Previous studies have suggested that the *b\** component of skin colour is greatly influenced by two pigments—melanin and primarily carotenoids<sup>27</sup>. Although carotenoids can serve as a cue for perceiving health, they may not be a reliable indicator of age perception, which could explain the non-significant results obtained in this experiment.

The current study has several limitations that should be considered. Firstly, the CFD images did not come with actual ages, and all selected images used were limited to between 20 and 40 years old based on the subjective age ratings provided, meaning that the findings may not generalise to other age groups<sup>34</sup>. Additionally, the chromatic information used in the images did not isolate surface-based cues, as shading caused by facial topography may have influenced chromatic content. Even though the facial bones almost stop growing after the age of 20, increasing age leads to a dissipation of the soft tissue in the face resulting in the bony appearance of the face in older adults<sup>35</sup>. Nose shape<sup>17</sup> and the thickness of the soft-tissue<sup>36</sup> change significantly with age. Such changes in facial topography may cause differential shading that influences the chromatic content of facial images. Therefore, while the current study sheds light on the role of colour and chromatic contrast in facial age perception, it is important to recognise the limitations associated with the age range of face stimuli and the complexity of chromatic information in facial images.

Furthermore, we have an imbalance in the gender distribution of our participants, with a disproportionately low number of male participants. This limits our ability to compare ratings between genders. A previous study has found that when providing age ratings to female faces, female raters tend to be more accurate than male raters<sup>37</sup>. Some researchers have further suggested that this gender effect might be limited to older female faces<sup>38</sup>. However, when both male and female participants were asked to rate faces of both genders<sup>39</sup>, found no significant difference between genders. Given this mixed evidence, exploring gender differences in facial age estimation could be a potential direction for future research.

In summary, our experiments demonstrate that the mere presence of chromatic information reduces inter-observer variability in age judgements and that manipulating red–green contrast biases the perception of age such that redness tends to be associated with youthfulness. Possible directions for future work are to test whether the red–green contrast effect is driven by a hue shift rather than by red–green contrast, and whether the red–green contrast effect generalises to stimuli of a larger age range.

## Methods

**Participants.** A different set of participants was recruited for each of the 3 experiments (Table 1 for demographics). All participants had normal or corrected to normal vision. A summary of participant demographics and the stimulus characteristics across all 3 experiments as shown in Table 1. All participants provided informed consent and ethical approval for this study was granted by the UNSW Human Research ethics committee. All methods were performed in accordance with the Declaration of Helsinki. Participants were all UNSW psychology undergraduate students and were recruited through the UNSW SONA system (Research management system) and received course credit.

**Chicago face database (CFD).** Stimuli were based on 200 full face images from the main CFD set of models recruited in the United States<sup>29</sup>. All selected images showed models photographed with neutral facial expressions, in standardised lighting against a white background. There is no ground-truth age information in the main

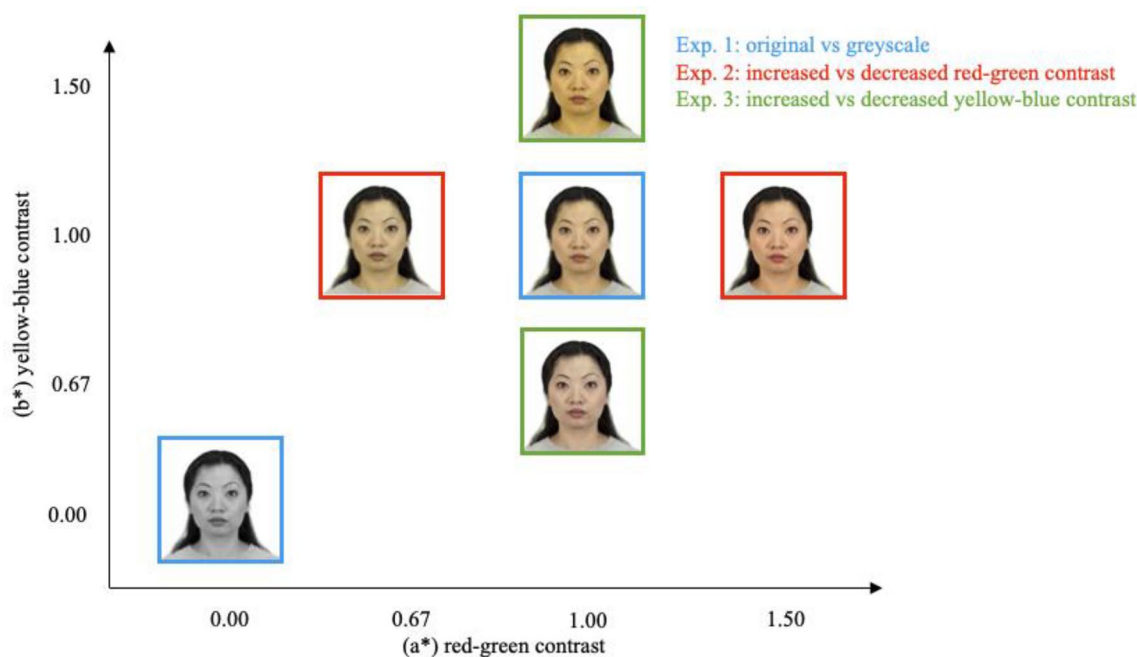
Experiment	n	Excluded	Included (M/F)	Image conditions
1	40	1	9/30, with a mean age of 19.05 ( $SD=1.52$ )	Colour (original)/grey scale
2	45	4	12/29, with a mean age of 20.59 ( $SD=5.88$ )	Higher/lower red–green contrast
3	44	1	3/40, with a mean age of 19.05 ( $SD=1.52$ )	Higher/lower yellow–blue contrast

**Table 1.** Summary of participant demographics and stimulus characteristics.

CFD set, but there are subjective age rating norms for each image (based on  $n=1087$  ratings). All selected images used were limited between 20 and 40 years old based on subjective age rating norms. We used these average age ratings to identify outliers in our data. Subjective age rating norms for the images were not used for the analysis proper. By self-reported ethnicity, there are images of 60 Asian, 60 White, 40 Black, and 40 Latino faces. The number of self-identified female and male images was balanced for each ethnicity.

**Stimuli.** We utilised the CIE Lab\* colour space to adjust the colour contrast of the 200 images sourced from CFD. This particular colour space is recognised for its uniform perceptual discriminability, where distances between coordinates of stimuli are predictive of perceived colour difference between the stimuli<sup>40</sup>. The 3 orthogonal dimensions of this colour space are light–dark ( $L^*$ ), red–green ( $a^*$ ), and yellow–blue ( $b^*$ ). Here we manipulated contrast along the chromatic dimensions while keeping the achromatic content ( $L^*$ ) constant. Demonstration images for each experiment are shown in Fig. 5. In Experiment 1, images were shown in original colour or grey scale version; in Experiment 2, images were shown with red–green contrast increased or decreased by a factor of 1.5 while images were shown with similarly modified yellow–blue contrast in Experiment 3. For each image, we first transformed the image from sRGB to CIE Lab, performed the contrast manipulation, then transformed the image back from CIE Lab to sRGB space, using the white point of CIE standard illuminant D65. All image manipulations were conducted using Matlab. The magnitude of the manipulations in Experiments 2 and 3 was based on informal piloting. The difference is perceptible, but the faces still look natural.

**Analyses.** In each experiment, a minority of participants gave unreliable ratings and their data were excluded. We excluded participants where the correlation between their ratings and the mean rated ages from CFD was  $<0.5$ . Additionally, we excluded data from individual ratings  $<13$ , as all face images used were of adult faces. Finally, we excluded data where the difference between the referred age from CFD and the rating from participants was  $>25$ . To investigate whether the presentation of images in two different conditions affects the perception of age, we used paired-sample  $t$  tests to compare the median ratings of each image. We conducted a paired-sample  $t$  test to examine whether the image conditions had an impact on inter-observer variability by comparing the  $SD$  of ratings for each image. Furthermore, we performed independent samples  $t$  tests on each



**Figure 5.** Different manipulations across experiments. *Note:* A sample image from the CFD showing colour manipulations used in Experiments 1–3. The original image is presented in the centre of the graph with a blue frame.

image's median and *SD* to determine if there were any significant effects of presentation condition within the images.

**Procedure and design.** This was an online study that allowed participants to join from a personal computer running a Google Chrome browser. We used JATOS<sup>41</sup> and jsPsych<sup>42</sup> hosted by a server within the UNSW School of Psychology. Participants were required to run the experiment in full-screen mode. Image size was calibrated to 18 cm (w) × 25.6 cm (h) by adjusting an on-screen rectangle to the size of a credit card prior to the experiment. Participants were asked to rate the age of faces by entering a two-digit integer (between 10 and 99) on the number pad of a computer keyboard. Participants were not informed about the age range of the stimuli and no feedback was given as to the accuracy of their responses. On each trial, an image was presented for 2 s and then removed. A written prompt and a text box would pop up on the screen until the participant made their response. Each participant provided 200 ratings with 100 images presented in colour and 100 in grey scale. Each participant rated each face only once, in colour or grey scale. Stimulus presentation was counterbalanced such that, across participants, each stimulus identity was presented the same number of times in colour and grey scale. Therefore, in Experiment 1, a total of 8000 ratings were collected across all 40 participants. For each stimulus identity, there were 20 ratings in colour and 20 ratings in grey scale, as each participant provided a rating either in colour or in grey scale.

### Data availability

The datasets generated during the current study are available from the corresponding author on reasonable request.

Received: 13 September 2022; Accepted: 1 August 2023

Published online: 14 August 2023

### References

- Mayes, A. E. *et al.* Environmental and lifestyle factors associated with perceived facial age in Chinese women. *PLoS ONE* **5**(12), 15270. <https://doi.org/10.1371/journal.pone.0015270> (2010).
- Christensen, K. *et al.* Perceived age as clinically useful biomarker of ageing: Cohort study. *BMJ* <https://doi.org/10.1136/bmj.b5262> (2009).
- Barber, N. The evolutionary psychology of physical attractiveness: Sexual selection and human morphology. *Ethol. Sociobiol.* **16**(5), 395–424. [https://doi.org/10.1016/0162-3095\(95\)00068-2](https://doi.org/10.1016/0162-3095(95)00068-2) (1995).
- Hsieh, J. Y. *et al.* The “cheerleader effect” in facial and bodily attractiveness: A result of memory bias and not perceptual encoding. *Q. J. Exp. Psychol.* **74**(5), 972–980. <https://doi.org/10.1177/1747021820976087> (2021).
- Symons, D. *Beauty is in the adaptations of the beholder: The evolutionary psychology of human female sexual attractiveness.* *Sexual Nature, Sexual Culture* 80–118 (University of Chicago Press, 1995).
- Langlois, J. H. *et al.* Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychol. Bull.* **126**, 390–423. <https://doi.org/10.1037/0033-2909.126.3.390> (2000).
- Russell, R., Sweda, J. R., Porcheron, A. & Mauger, E. Sclera color changes with age and is a cue for perceiving age, health, and beauty. *Psychol. Aging* **29**(3), 626–635. <https://doi.org/10.1037/a0036142> (2014).
- Gunn, D. A. *et al.* Why some women look young for their age. *PLoS ONE* **4**(12), e8021. <https://doi.org/10.1371/journal.pone.0008021> (2009).
- Gomi, T. & Imamura, T. Age-related changes in the vasculature of the dermis of the upper lip vermilion. *Aging* **11**(11), 3551–3560. <https://doi.org/10.18632/aging.101996> (2019).
- Bruce, V. & Langton, S. The use of pigmentation and shading information in recognising the sex and identities of faces. *Perception* **23**(7), 803–822. <https://doi.org/10.1068/p230803> (1994).
- Kemp, R., Pike, G., White, P. & Musselman, A. Perception and recognition of normal and negative faces: The role of shape from shading and pigmentation cues. *Perception* **25**(1), 37–52. <https://doi.org/10.1068/p250037> (1996).
- Russell, R., Sinha, P., Biederman, I. & Nederhouser, M. Is pigmentation important for face recognition? Evidence from contrast negation. *Perception* **35**(6), 749–759. <https://doi.org/10.1068/p5490> (2006).
- Leyden, J. J. Clinical features of ageing skin. *Br. J. Dermatol.* **122**(s35), 1–3. <https://doi.org/10.1111/j.1365-2133.1990.tb16118.x> (1990).
- Ortonne, J.-P. Pigmentary changes of the ageing skin. *Br. J. Dermatol.* **122**(s35), 21–28. <https://doi.org/10.1111/j.1365-2133.1990.tb16121.x> (1990).
- Matts, P. J., Fink, B., Grammer, K. & Burquest, M. Color homogeneity and visual perception of age, health, and attractiveness of female facial skin. *J. Am. Acad. Dermatol.* **57**(6), 977–984. <https://doi.org/10.1016/j.jaad.2007.07.040> (2007).
- Hill, H., Bruce, V. & Akamatsu, S. Perceiving the sex and race of faces: The role of shape and colour. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **261**(1362), 367–373. <https://doi.org/10.1098/rspb.1995.0161> (1995).
- Sforza, C., Grandi, G., De Menezes, M., Tartaglia, G. M. & Ferrario, V. F. Age- and sex-related changes in the normal human external nose. *Forensic Sci. Int.* **204**(1), 205.e1–205.e9. <https://doi.org/10.1016/j.forsciint.2010.07.027> (2011).
- Fink, B., Grammer, K. & Matts, P. J. Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evol. Hum. Behav.* **27**(6), 433–442. <https://doi.org/10.1016/j.evolhumbehav.2006.08.007> (2006).
- Burt, D. M. & Perrett, D. I. Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour information. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **259**(1355), 137–143. <https://doi.org/10.1098/rspb.1995.0021> (1995).
- Tarr, M. J., Kersten, D., Cheng, Y. & Rossion, B. It's Pat! Sexing faces using only red and green. *J. Vis.* **1**(3), 337–337. <https://doi.org/10.1167/1.3.337> (2001).
- Re, D. E., Whitehead, R. D., Xiao, D. & Perrett, D. I. Oxygenated-blood colour change thresholds for perceived facial redness, health, and attractiveness. *PLoS ONE* **6**(3), e17859. <https://doi.org/10.1371/journal.pone.0017859> (2011).
- Stephen, I. D. & McKeegan, A. M. Lip colour affects perceived sex typicality and attractiveness of human faces. *Perception* **39**(8), 1104–1110. <https://doi.org/10.1068/p6730> (2010).
- Yip, A. W. & Sinha, P. Contribution of color to face recognition. *Perception* **31**(8), 995–1003. <https://doi.org/10.1068/p3376> (2002).
- Porcheron, A., Mauger, E. & Russell, R. 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> (2013).
- Russell, R. Sex, beauty, and the relative luminance of facial features. *Perception* **32**(9), 1093–1107. <https://doi.org/10.1068/p5101> (2003).

26. Fisher, C. I., Hahn, A. C., DeBruine, L. M. & Jones, B. C. 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> (2014).
27. Stephen, I. D., Law Smith, M. J., Stirrat, M. R. & Perrett, D. I. Facial skin coloration affects perceived health of human faces. *Int. J. Primatol.* **30**(6), 845–857. <https://doi.org/10.1007/s10764-009-9380-z> (2009).
28. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **57**(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x> (1995).
29. Ma, D. S., Correll, J. & Wittenbrink, B. The Chicago face database: A free stimulus set of faces and norming data. *Behav. Res. Methods* **47**(4), 1122–1135. <https://doi.org/10.3758/s13428-014-0532-5> (2015).
30. Dehon, H., & Brédart, S. An “other-race” effect in age estimation from faces. *Perception* **30**(9), 1107–1113. <https://doi.org/10.1068/p3122> (2001).
31. Walker, P. M., & Hewstone, M. A perceptual discrimination investigation of the own-race effect and intergroup experience. *Appl. Cogn. Psychol.* **20**(4), 461–475. <https://doi.org/10.1002/acp.1191> (2006).
32. Stephen, I. D., Coetzee, V., Law Smith, M. & Perrett, D. I. Skin blood perfusion and oxygenation colour affect perceived human health. *PLoS ONE* **4**(4), e5083. <https://doi.org/10.1371/journal.pone.0005083> (2009).
33. Perrett, D. I., Talamas, S. N., Cairns, P. & Henderson, A. J. Skin color cues to human health: Carotenoids, aerobic fitness, and body fat. *Front. Psychol.* <https://doi.org/10.3389/fpsyg.2020.00392> (2020).
34. Voegeli, R. *et al.* Differences between perceived age and chronological age in women: A multi-ethnic and multi-centre study. *Int. J. Cosmet. Sci.* **43**(5), 547–560. <https://doi.org/10.1111/ics.12727> (2021).
35. Enlow, D. H. & Moyers, R. E. *Handbook of facial growth* (WB Saunders Company, 1982).
36. Farkas, L. G. *et al.* Anthropometric measurements of the facial framework in adulthood: Age-related changes in eight age categories in 600 healthy white north Americans of European ancestry from 16 to 90 years of age. *J. Craniofac. Surg.* **15**(2), 288–298 (2004).
37. Nkengne, A. *et al.* Influence of facial skin attributes on the perceived age of Caucasian women. *J. Eur. Acad. Dermatol. Venereol.* **22**(8), 982–991. <https://doi.org/10.1111/j.1468-3083.2008.02698.x> (2008).
38. Vestlund, J., Langeborg, L., Sörqvist, P., & Eriksson, M. Experts on age estimation. *Scand. J. Psychol.* **50**(4), 301–307. <https://doi.org/10.1111/j.1467-9450.2009.00726.x> (2009).
39. Voelkle, M. C., Ebner, N. C., Lindenberger, U., & Riediger, M. Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age. *Psychol. Aging* **27**, 265–277. <https://doi.org/10.1037/a0025065> (2012).
40. Tkalcic, M. & Tasic, J. F. *Colour spaces: Perceptual, historical and applicational background. The IEEE Region 8 EUROCON 2003 Computer as a Tool* 304–308 (IEEE, 2003).
41. Lange, K., Kühn, S. & Filevich, E. “Just another tool for online studies” (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLoS ONE* **10**(6), e0130834. <https://doi.org/10.1371/journal.pone.0130834> (2015).
42. de Leeuw, J. R. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behav. Res. Methods* **47**(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y> (2015).

## Acknowledgements

This work was supported by Australian Research Council Discovery Project DP190100491 to CC.

## Author contributions

J.Y.J.H.: conceptualization, methodology, software, investigation, formal analysis, writing—original draft. W.P.B.: conceptualization, methodology, software, writing—review & editing. E.G.: conceptualization, methodology, formal analysis, writing—review and editing. C.W.G.C.: conceptualization, methodology, formal analysis, writing—review & editing, resources.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to C.W.G.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023