Contribution of shape and surface reflectance information to kinship detection in 3D face images

Vanessa Fasolt University of Glasgow, Glasgow, Scotland, UK ✉

Iris J. Holzleitner University of Glasgow, Glasgow, Scotland, UK

Anthony J. Lee University of Stirling, Stirling, Scotland, UK

Kieran J. O'Shea University of Glasgow, Glasgow, Scotland, UK

Lisa M. DeBruine University of Glasgow, Glasgow, Scotland, UK

Previous research has established that humans are able to detect kinship among strangers from facial images alone. The current study investigated what facial information is used for making those kinship judgments, specifically the contribution of face shape and surface reflectance information (e.g., skin texture, tone, eye and eyebrow color). Using 3D facial images, 195 participants were asked to judge the relatedness of 100 child pairs, half of which were related and half of which were unrelated. Participants were randomly assigned to judge one of three stimulus versions: face images with both surface reflectance and shape information present (*reflectance and shape version*), face images with shape information removed but surface reflectance present (*reflectance version*), or face images with surface reflectance information removed but shape present (*shape version*). Using binomial logistic mixed models, we found that participants were able to detect relatedness at levels above chance for all three stimulus versions. Overall, both individual shape and surface reflectance information contribute to kinship detection, and both cues are optimally combined when presented together. Preprint, preregistration, code, and data are available on the Open Science Framework (osf.io/7ftxd).

Introduction

Numerous studies have found evidence for allocentric kin recognition, showing that individuals are able to detect relatedness when shown face images of people unknown to them (Alvergne, Perreau, Mazur, Mueller, & Raymond, 2014; Bressan & Dal Martello, 2002; Bressan & Grassi, 2004; Dal Martello, DeBruine, & Maloney, 2015; DeBruine et al., 2009; Maloney & Dal Martello, 2006; Nesse, Silverman, & Bortz, 1990). Generally, previous research has examined this ability by asking raters to judge whether a pair of 2D facial images are related or not, or by asking raters to match up a related pair out of a number of options. The standard of the stimuli used in these studies varies considerably, with some image sets being sent in by families (e.g., using photographs from family holidays), while other image sets were collected by researchers under more controlled conditions.

Some of this research has found that different facial areas are important when making kinship judgments (Alvergne et al., 2014; Dal Martello & Maloney, 2006). For instance, Dal Martello and Maloney (2006) found that the upper half of the face contains more informative cues of kinship than the lower half of the face, but that these cues are optimally combined when assessing a full face, and that featural information (e.g., the shape of the nose) is more informative than configurational information (the relationship between features) when making kinship judgments. Alvergne et al. (2014) found that raters were not able to detect kin when only the lower half of the face was shown, but again, featural information was more important than configurational information. Dal Martello et al.'s (2015) finding that facial inversion or rotation does not affect kinship judgments further supports this notion that featural, rather than configurational, information is important for kin judgments. This converging evidence suggests that face shape cues play an important role in kinship detection. Yet, this has never been directly examined. Face shape is highly heritable (Djordjevic, Zhurov, & Richmond, 2016; Kim et al.,

2013; Tsagkrasoulis, Hysi, Spector, & Montana, 2017; Weinberg, Parsons, Marazita, & Maher, 2013). Genetic factors explain over 70% of the variance in facial traits such as face size; nose height, width, and prominence; interocular distance; and lip prominence. As kin have a more similar genetic make-up than non-kin, they also have a more similar facial shape, and hence are more likely to look more similar than non-kin. While environmental factors contribute to the variance in facial morphology as well, families typically live in a shared environment which might further contribute to facial similarity. Thus, facial shape is likely to be an informative cue of kinship.

Facial skin tone is another highly heritable facial trait that has not yet been explicitly examined in the allocentric kin recognition literature. Heritability has been estimated to account for around 56% to 83% of the variance in skin tone, mainly due to ethnicity (Clark, Stark, Walsh, Jardine, & Martin, 1981; Frisancho, Wainwright, & Way, 1981; Williams-Blangero & Blangero, 1991). Environmental factors also contribute to the variance in tan, as well as red and yellow skin tones. Skin yellowness as measured by spectrophotometry has been positively linked to the intake of the antioxidant carotenoid through fruit and vegetables (Alaluf, Heinrich, Stahl, Tronnier, & Wiseman, 2002; Pezdirc et al., 2015; Stephen, Coetzee, & Perrett, 2011; Tan, Graf, Mitra, & Stephen, 2015; Whitehead, Re, Xiao, Ozakinci, & Perrett, 2012), redness has been positively linked to skin vascularization and blood oxygenation through cardiovascular, hormonal, and circulatory health and physical exercise (Charkoudian, Stephens, Pirkle, Kosiba, & Johnson, 1999; Johnson, 1998; Piérard, 1998; Thornton, 2002), and tan/melanin has been linked to sun exposure, with tanning potential being genetically determined (Kalla, 1972; Williams-Blangero & Blangero, 1991). As most families tend to live in a shared or similar environment (e.g., are likely to have a similar diet, exercise routine, or sun exposure), facial tone, too, might be an informative cue of kinship. Moreover, eye color can be an informative cue of kinship, as eye color is highly heritable (Larsson, Pedersen, & Stattin, 2003; Zhu et al., 2004). Dal Martello and Maloney (2006) tested the contribution of the eye region (rather than eye color specifically) to allocentric kin recognition, finding that kinship judgment accuracy decreased by 20% when the eye region was obscured. Yet, this decrease in accuracy levels was not significant and the study did not specifically speak to the importance of eye color alone in allocentric kin recognition, as both eye color and shape were obscured. Still, observing a decrease in accuracy suggests that the eye region is to some extent an informative cue to kinship that needs to be tested further.

In light of the fact that both shape and texture/tone cues have been implicated but not explicitly investigated in the allocentric kin recognition literature, the current study investigated the direct contribution of facial shape and surface reflectance information to kinship detection in a sample of 3D images. We use the term surface reflectance information to refer to facial cues as captured by the texture map of our 3D images, such as skin tone, texture, and eye color. We created three different versions of 3D face stimuli: one version combined both individual surface reflectance and shape information (*reflectance and shape version*), one version that retained individual surface reflectance information but was standardized in shape (*reflectance version*), and one that showed individual shape but no surface reflectance information (*shape version*). This allowed us to directly investigate how surface reflectance and shape information independently influence kin judgments.

We hypothesized that:

1. Regardless of reflectance and shape information, people would be able to detect relatedness at levels above chance, judging related pairs to be related more often than unrelated pairs. This would be demonstrated in the analysis by a positive main effect of relatedness.
2. Both reflectance and shape information would contribute significantly to accuracy of relatedness judgments, with judgment accuracy being higher for stimuli with reflectance information than without, and for stimuli with shape information than without. This would be demonstrated by a positive two-way interaction between relatedness and reflectance, and a positive two-way interaction between relatedness and shape.

Methods

The methods and analyses for this study were preregistered on the Open Science Framework (osf.io/7ftxd/). Planned analysis script and data are available at this site, as well as details about the hypotheses, stimuli, and procedure. All procedures and analyses below follow this preregistration. Additional non-preregistered analyses are clearly marked and improved visualizations of findings have been added.

Stimuli

Face images were collected from children visiting a local science center who volunteered to take part in a study of facial cues of family relatedness. Parental consent and child assent were obtained from each child to use their face photograph in studies of family

resemblance detection. Children were photographed sitting or standing at a distance of 90 cm to the camera rig, looking straight at the camera with hair pulled back and any glasses, scarves, and hats removed, once with a smiling and once with a neutral facial expression.

Images were collected using a DI3D system (http://www.di4d.com/). This is a passive stereo photogrammetry-based solution for the creation of accurate, ultra-high resolution, full-color 3D surface images using six standard digital cameras (Canon EOS100D; lenses: Canon EF 50 mm f/1.8 STM; Canon, Tokyo, Japan). Two remote-controlled flash units (Elinchrom D-Lite RX 2; Elinchrom, Renens, Switzerland) were used for lighting. The software DI3Dcapture (version 6.8.4) was used to capture participants' faces from six different angles. The 3D images were generated using DI3Dview (version 6.8.9), which creates both a texture map in the BMP file format (at a resolution of 1MP minimum) as well as a three-dimensional mesh from the raw data that was exported in the Wavefront OBJ file format.

Extraneous parts of each face scan were removed using MeshLab (Visual Computing Lab ISTI-CNR; http://www.meshlab.net/) and Blender (Blender Foundation; https://www.blender.org/) and faces were delineated in MorphAnalyser 2.4 (Tiddeman, Duffy, & Rabey, 2000). More details on image collection and processing are available at osf.io/bvtnj.

The standard of photographs from previous studies varied; for instance, one common method of building a stimulus set of related individuals has been asking family members to send photos from family albums. This method is problematic because photographs can be easily ascribed to one family unit due to properties of the picture extraneous to facial kinship cues (e.g., individuals from the same family can match in background, illumination, or image quality and therefore be judged to be related based solely on these similarities). The varying standard of photographs in general is a concern for the field and might be a factor in the plethora of diverging and contradicting findings in the literature. The current study used highly standardized photographs, from which all background information was removed.

The use of highly standardized 3D photographs is novel in the allocentric kin recognition literature. It allows participants to view the faces from different angles, enabling participants to perceive the actual depth, curvature, and protrusion of facial features, rather than making inferences based on shadows in a 2D image. Moreover, as environmental factors explain some variance in face shape and texture/tone, we used face images of children under the age of 17, as younger siblings are more likely to share an environment. We were not able to collect data on whether siblings shared an environment due to time constrictions; however, families came into the science center together, indicating that they spend at least some time together. Lastly, we have previously shown that a smiling facial expression decreases kin recognition accuracy (Fasolt, Holzleitner, Lee, O'Shea, & DeBruine, 2018); hence, we only used stimuli with a neutral facial expression in the current study.

From a set of approximately 2,000 images of individuals of varying age, sex, and relatedness, we algorithmically chose the maximum number of sibling pairs fitting a number of criteria. Both siblings were required to be fully genetically related (same biological father and mother) and were required to be non-twin full siblings under the age of 18. We also required that a pair of age-matched (within 1 year), ethnicity-matched, and sex-matched foil images were available from family units that were not represented elsewhere in the image set. Specifically, the two individuals in each sibling pair are related to each other, but not to any other individual in the set, while all individuals in unrelated pairs are related to no individuals in the set.

This matching procedure is crucial as it ensures that there are no interdependencies of stimuli within the set, as this could result in judgment biases. For example, most studies in the field use individuals from one family as both experimental and control stimuli; hence, the same faces are seen in multiple trials. This means that a rater might already have matched a child to a parent, and when this same child comes up again in other trials, the rater might infer unrelatedness based on the previous cognitive "relatedness" decision, rather than evaluating facial kinship cues again.

This procedure produced 50 sibling pairs and 50 matched unrelated pairs. In each group, 13 pairs were both male, 15 pairs were both female, and 22 pairs were male and female. The individuals ranged from 3 to 17 years of age (mean age = 9.44, $SD = 2.92$) and the age difference between individuals in a pair ranged from 0 to 7 (mean = 2.96, $SD = 1.64$) years. The age difference between individuals in related and unrelated pairs was approximately equal due to the matching of foil pairs to related pairs. All children were white.

Three versions of these 100 pairs of stimuli were created, a reflectance and shape version, a reflectance version, and a shape version. The reflectance and shape versions were the original 3D photographs, showing both individual shape and surface reflectance information. A shape version was created by showing only the 3D shape but no surface reflectance information. A reflectance version was created by mapping children's individual surface reflectance information onto an average face shape, which was computed by averaging the face shape of all 200 children.

Stimulus pairs showed each face from three different perspectives (i.e., −40°, frontal view, and +40°; see Figure 1).
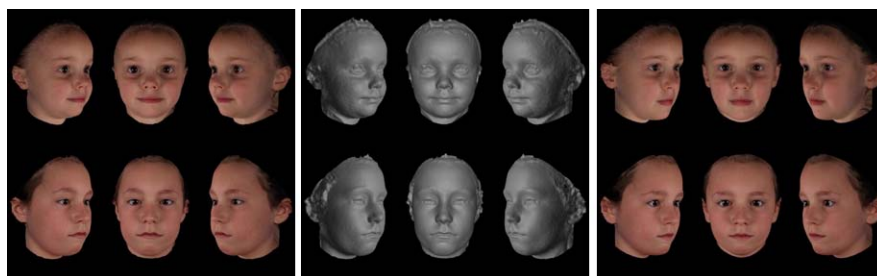
Figure 1. Presentation of the three versions of the stimuli (between subjects), (1) reflectance and shape version (original photograph), (2) shape version (individual shape information retained but surface reflectance information removed) and (3) reflectance version (individual surface reflectance information retained but shape standardized).

Procedure

Raters were recruited online through social media (e.g., Facebook, Twitter) and social bookmarking sites. The study itself was completed online at faceresearch.org on raters' own computers and lasted around 10 minutes.

Raters were randomly assigned to one of three versions of the study, either the reflectance and shape version, the shape version, or the reflectance version. Each rater was presented with only one version. Within each version, stimulus pairs were presented in a random order. Before the study began, raters received the following instructions: "In this experiment you will be shown 100 pairs of faces. Some are siblings, some are an unrelated pair. You will be asked to determine whether each pair is 'unrelated' or 'related'." Raters were shown one pair of child faces at a time and chose their answer by clicking on buttons labeled unrelated or related without any time restrictions.

Raters

The study was started by a total of 270 people across versions. We excluded 68 raters who did not rate all 100 stimuli and were therefore left with 202 raters. As specified in the preregistration, based on a power calculation we only included the first 65 raters to complete each version of the study, resulting in 195 raters included in the following analysis. The full data set including all 270 raters is available at osf.io/7ftxd/. The inclusion of all raters did not change the main findings of the analysis reported below but did show an additional significant main effect of surface reflectance information, whereby stimuli with no reflectance information were judged to be related less often, independent of actual relatedness.

Overall, the responses from 45 men (mean age = 29.63; $SD = 11.6$) and 144 women (mean age = 28.67; $SD = 11.1$) were analyzed. Six raters (mean age = 30.46; $SD = 5.18$) did not indicate their gender. Most raters identified as white (155 out of 195 raters).

Analysis

We used a logistic mixed model to predict relatedness judgments from actual relatedness (effect-coded as related = +0.5 and unrelated = −0.5), surface reflectance information (effect-coded as reflectance on = +0.5 and reflectance off = −0.5), shape information (effect-coded as shape on = +0.5 and shape off = −0.5), the interactions between surface reflectance information and relatedness, and shape information and relatedness. We included the rater ID and stimulus ID as random effects and specified our slopes maximally (Barr, Levy, Scheepers, & Tily, 2013). Analyses were conducted in the programming software R version 3.5.0 (R Core Team, 2017) in conjunction with lme4 version 1.1.17 (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest version 3.0.1 (Kuznetsova, Brockhoff, & Christensen, 2016).

We use a mixed model as this allowed us to account for variation among both raters and stimuli. This prevents the inflated false positive rates that can come from aggregating responses: analyses aggregating over raters do not generalize beyond the specific set of stimuli used, while analyses aggregating over stimuli do not generalize beyond the specific raters. These limitations are overcome in a mixed model analysis where responses are not aggregated.

Results

Supporting Hypothesis 1, we found a main effect of relatedness ($\beta = 0.96$, $SE = 0.17$, $z = 5.73$, $p < 0.001$),
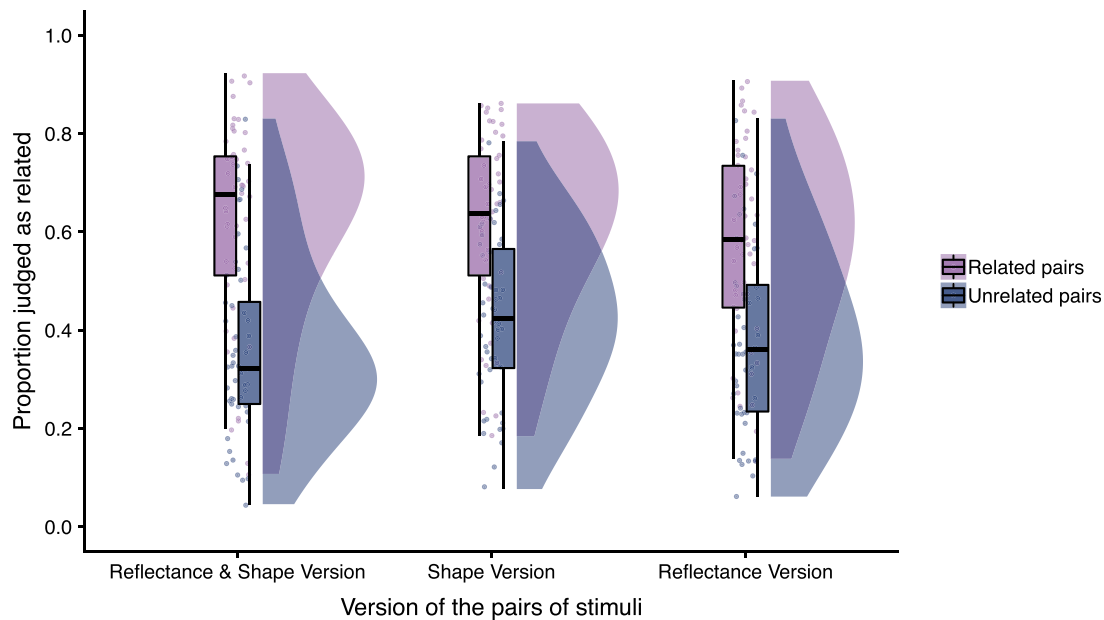
Figure 2. The effects of stimulus version and actual relatedness on average kinship judgments (0 = unrelated judgment, 1 = related judgment). The box plots, points, and distributions represent the average relatedness score for each individual stimulus pair. The box plots are showing the median, first and third quartile, and the minimum and maximum relatedness score for related (pink) and unrelated (blue) pairs. The distribution "clouds" also give more information about patterns in the data; for example, more or less overlap in average relatedness score for actually related (pink) or unrelated (blue) pairs in the different stimulus versions.

whereby actually related pairs were 2.61 times more likely to be judged as related than unrelated pairs (see Figure 2).

Hypothesis 2 was partially supported by our results (see Figure 2). As predicted, there was a significant positive interaction between relatedness and shape information ($\beta = 0.32$, $SE = 0.14$, $z = 2.2$, $p = 0.028$, odds ratio $= 1.38$). The interaction between relatedness and surface reflectance information was also positive but not significant ($\beta = 0.28$, $SE = 0.17$, $z = 1.68$, $p = 0.093$, odds ratio $= 1.32$). Both shape and reflectance information contributed to the accuracy of relatedness judgments, though the latter not significantly so. Yet, the difference in effect size between these two interactions was small. Higher powered studies are needed to conclusively determine whether shape

contributes more to kinship judgments than surface reflectance (see Table 1).

Further analyses

Next, to further clarify the individual importance of shape and reflectance cues in kinship judgments, we conducted additional analyses not included in the preregistration. First, we ran three logistic mixed effects models, one for each stimulus version. Again, actual relatedness was entered as a fixed effect. These analyses revealed that raters accurately identified related and unrelated pairs in all three versions of the study (see Table 2).

Following Dal Martello and Maloney (2006), we conducted a signal detection analysis obtaining esti-

Effect	Estimate (β)	SE	z	p	Odds ratio
Intercept	0.00	0.10	−0.03	0.973	1.00
Relatedness	0.96	0.17	5.73	<0.001	2.61
Surface reflectance	−0.15	0.13	−1.12	0.263	0.86
Shape	0.04	0.13	0.35	0.725	1.04
Relatedness × Surface reflectance	0.28	0.17	1.68	0.093	1.32
Relatedness × Shape	0.32	0.14	2.20	0.028	1.38

Table 1. Results from main analysis.

Version	Related pairs	Unrelated pairs	Estimate	*SE*	*z*	*p*	Odds ratio
Reflectance and shape version	61.7%	36.2%	1.25	0.21	6.08	<0.001	3.49
Reflectance version	57.2%	38.6%	0.95	0.20	4.75	<0.001	2.59
Shape version	61.7%	42.4%	0.98	0.18	5.35	<0.001	2.66

Table 2. The table shows the rate of identifying related pairs as related (hit rate), and the rate of identifying unrelated pairs incorrectly as related (false alarm rate) as well as the results from the mixed effects models for each stimulus version.

mates of sensitivity d' and likelihood criteria β, which allowed us to further examine performance rates in the three different versions of the stimuli (Green & Swets, 1966). Performance accuracy in all three versions was above chance, which was indicated by a d' value being significantly bigger than 0 (see Table 3). The z statistic which determined whether the d' value was in fact bigger than 0 was computed by dividing the estimate d' by the Bootstrap estimate of its SD. Performance rates were significantly worse in the shape version ($z = -3.558$, $p < 0.001$) and skin reflectance version ($z = -4.022$, $p < 0.001$) compared to the reflectance and shape version. Performance rates in the shape version and the reflectance version did not differ from each other ($z = -0.464$, $p = 0.643$).

Lastly, and also following Dal Martello and Maloney (2006), we calculated the predicted d'_{rs} value for the reflectance and shape version from the two independent d' values of the shape version (d'_s) and the reflectance version (d'_r) with the following formula (Green & Swets, 1966):

$$d'_{rs} = \sqrt{(d'_s)^2 + (d'_r)^2}$$

The predicted $d'_{rs} = 0.68$ value and the actual $d'_{rs} = 0.65$ value from the reflectance and shape version were not significantly different from each other ($z = -0.619$, $p = 0.536$), which suggests that the reflectance and shape version did not provide any additional, independent information, but that reflectance and shape are optimally combined to make kinship judgments from the original images. All the information affecting performance in the reflectance and shape version is already present in the shape version and reflectance version independently. Thus, it is clear that reflectance information is optimally combined with shape information to detect kinship.

Discussion

We found that third-party raters were able to reliably identify related and unrelated child sibling pairs, a robust finding across the literature (Alvergne, Perreau, Mazur, Mueller, & Raymond, 2014; Bressan & Dal Martello, 2002; Bressan & Grassi, 2004; Dal Martello et al., 2015; DeBruine et al., 2009; Maloney & Dal Martello, 2006). Raters were able to detect kinship accurately in all stimulus versions (i.e., even when only shape or surface reflectance information was available). We also found that individual shape and reflectance information are optimally combined to make kinship judgments in the reflectance and shape version, and that the presentation of the combined cues does not add any further, independent information that is not already present in shape only or reflectance only versions.

These findings highlight the importance of shape and surface reflectance information in allocentric kin recognition and complement research showing that facial morphology and skin texture/tone cues are heritable (Clark et al., 1981; Djordjevic et al., 2016; Frisancho et al., 1981; Kim et al., 2013; Tsagkrasoulis et al., 2017; Weinberg, Parsons, Marazita, & Maher, 2013; Williams-Blangero & Blangero, 1991). However, the current study was unable to distinguish whether kinship judgments were based on face similarities due to genetic or shared environmental sources. While the use of stimuli showing child sibling-pairs (between 3 to 17 years of age) may minimize the effect of unique environmental and lifestyle factors on facial shape and reflectance (at least compared to adult sibling-pairs), we did not collect data on whether related stimuli pairs actually shared an environment or not. Hence, we cannot exclude the possibility that reflectance information varied within related pairs due to living in

Version	*d*	*d_SD*	β	β_SD	*z*	*p*
Shape version	0.491	0.032	0.974	0.008	15.557	<0.001
Reflectance version	0.470	0.032	1.025	0.008	14.649	<0.001
Reflectance and shape version	0.652	0.032	1.019	0.011	20.278	<0.001

Table 3. The d' estimate and the likelihood criterion β for the signal detection analysis are shown for each version. Standard deviations were estimated by a bootstrap procedure (Efron & Tibshirani, 1993) based on 1,000 replications.

different environments which could have led to reflectance being less informative of kinship than shape. This limitation could be addressed by assessing kinship judgments between individuals of varying genetic relatedness, or modeling for unique/shared environment in child siblings and adult siblings.

The current study expands on past research looking at which specific regions of the face influence kin recognition (Alvergne et al., 2014; Dal Martello & Maloney, 2006). While these previous studies implicitly assumed that shape or reflectance information of different regions are informative kinship cues, here we were able to explicitly confirm that shape and reflectance information are both cues of kinship and are used as such. Studies investigating facial regions did not test what specific information was extracted from these regions in order to make kinship judgments (i.e., whether it was shape or reflectance information, or an optimal combination of both). This would be an important next step, as facial regions may vary in the information they provide. For example, the eye region has been found to hold kinship cues (Dal Martello & Maloney, 2006), but it is unclear what exact information from the eye region is used to make kinship judgments. It is possible that eye color or eye shape is used as kinship cue, as both are heritable (Larsson et al., 2003; Tsagkrasoulis et al., 2017; Zhu et al., 2004), or that both are optimally combined.

Furthermore, a difficulty when looking at reflectance independently of shape information is that the used texture maps still contained some shape and depth information through shadows from protruding and deep features, and through reflectance information specific to face regions (e.g., redness of cheeks, lips). This intrinsic shape information in the reflectance version might have been redundant when judging reflectance and shape version stimuli. However, our predicted $d'_{rs} = 0.68$ is nearly identical to the actual performance $d'_{rs} = 0.65$, which suggests that there is no redundant information in the two separate versions when combining them in the reflectance and shape version. Alternatively, this could be the result of having both redundant and interacting information cancelling each other out when combining shape and reflectance information. Our results cannot distinguish between these two possibilities.

To conclude, raters can detect relatedness among strangers based on facial cues alone. Facial shape and surface reflectance cues can be independently used to make correct kinship decisions but are optimally combined when they are both available as in the reflectance and shape version of our 3D stimuli.

Keywords: kinship, face perception, allocentric kin recognition, facial resemblance, 3D face shape, surface reflectance information

References

Alaluf, S., Heinrich, U., Stahl, W., Tronnier, H., & Wiseman, S. (2002). Dietary carotenoids contribute to normal human skin color and UV photosensitivity. *The Journal of Nutrition*, *132*(3), 399–403, https://doi.org/10.1093/jn/132.3.399.

Alvergne, A., Perreau, F., Mazur, A., Mueller, U., & Raymond, M. (2014). Identification of visual paternity cues in humans. *Biology Letters*, *10*(4): 20140063, https://doi.org/10.1098/rsbl.2014.0063.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278, https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48, https://doi.org/10.18637/jss.v067.i01.

Bressan, P., & Dal Martello, M. F. (2002). Talis pater, talis filius: Perceived resemblance and the belief in genetic relatedness. *Psychological Science*, *13*(3), 213–218, https://doi.org/10.1111/1467-9280.00440.

Bressan, P., & Grassi, M. (2004). Parental resemblance in 1-year-olds and the Gaussian curve. *Evolution and Human Behavior*, *25*(3), 133–141, https://doi.org/10.1016/j.evolhumbehav.2004.03.001.

Charkoudian, N., Stephens, D. P., Pirkle, K. C., Kosiba, W. A., & Johnson, J. M. (1999). Influence of female reproductive hormones on local thermal control of skin blood flow. *Journal of Applied Physiology*, *87*(5), 1719–1723, https://doi.org/10.1152/jappl.1999.87.5.1719.

Clark, P., Stark, A., Walsh, R., Jardine, R., & Martin, N. (1981). A twin study of skin reflectance. *Annals of Human Biology*, *8*(6), 529–541, https://doi.org/10.1080/03014468100005371.

Dal Martello, M. F., DeBruine, L. M., & Maloney, L. T. (2015). Allocentric kin recognition is not affected by facial inversion. *Journal of Vision*, *15*(13):5, 1–

11, https://doi.org/10.1167/15.13.5. [PubMed] [Article]

Dal Martello, M. F., & Maloney, L. T. (2006). Where are kin recognition signals in the human face? *Journal of Vision*, 6(12):2, 1356–1366, https://doi.org/10.1167/6.12.2. [PubMed] [Article]

DeBruine, L. M., Smith, F. G., Jones, B. C., Roberts, S. C., Petrie, M., & Spector, T. D. (2009). Kin recognition signals in adult faces. *Vision Research*, 49(1), 38–43, https://doi.org/10.1016/j.visres.2008.09.025.

Djordjevic, J., Zhurov, A. I., & Richmond, S. (2016). Genetic and environmental contributions to facial morphological variation: A 3D population-based twin study. *PLoS One*, 11(9):e0162250, https://doi.org/10.1371/journal.pone.0162250.

Fasolt, V., Holzleitner, I. J., Lee, A. J., O'Shea, K. J., & DeBruine, L. M. (2018). Facial expressions influence kin recognition accuracy. *Human Ethology Bulletin*, 33(4), 19–27, https://doi.org/10.22330/heb/334/019-027.

Frisancho, A. R., Wainwright, R., & Way, A. (1981). Heritability and components of phenotypic expression in skin reflectance of Mestizos from the Peruvian lowlands. *American Journal of Physical Anthropology*, 55(2), 203–208, https://doi.org/10.1002/ajpa.1330550207.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). Oxford, UK: Wiley.

Johnson, J. M. (1998). Physical training and the control of skin blood flow. *Medicine & Science in Sports & Exercise*, 30(3), 382–386, https://doi.org/10.1097/00005768-199803000-00007.

Kalla, A. K. (1972). Parent-child relationship and sex differences in skin tanning potential in man. *Human Genetics*, 15(1), 39–43, https://doi.org/10.1007/bf00273430.

Kim, H.-J., Im, S.-W., Jargal, G., Lee, S., Yi, J.-H., Park, J.-Y., . . . Seo, J-S. (2013). Heritabilities of facial measurements and their latent factors in Korean families. *Genomics & Informatics*, 11(2), 83–92, https://doi.org/10.5808/gi.2013.11.2.83.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). *LmerTest: Tests in linear mixed effects models*. Retrieved from https://CRAN.R-project.org/package=lmerTest.

Larsson, M., Pedersen, N. L., & Stattin, H. (2003). Importance of genetic effects for characteristics of the human iris. *Twin Research*, 6(3), 192–200, https://doi.org/10.1375/136905203765693843.

Maloney, L. T., & Dal Martello, M. F. (2006). Kin recognition and the perceived facial similarity of children. *Journal of Vision*, 6(10):4, 1047–1056, https://doi.org/10.1167/6.10.4. [PubMed] [Article]

Nesse, R. M., Silverman, A., & Bortz, A. (1990). Sex differences in ability to recognize family resemblance. *Ethology and Sociobiology*, 11(1), 11–21, https://doi.org/10.1016/0162-3095(90)90003-O.

Pezdirc, K., Hutchesson, M., Whitehead, R., Ozakinci, G., Perrett, D., & Collins, C. (2015). Fruit, vegetable and dietary carotenoid intakes explain variation in skin-color in young Caucasian women: A cross-sectional study. *Nutrients*, 7(7), 5800–5815, https://doi.org/10.3390/nu7075251.

Piérard, G. (1998). EEMCO guidance for the assessment of skin colour. *Journal of the European Academy of Dermatology and Venereology*, 10(1), 1–11, https://doi.org/10.1016/S0926-9959(97)00183-9.

R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Stephen, I. D., Coetzee, V., & Perrett, D. I. (2011). Carotenoid and melanin pigment coloration affect perceived human health. *Evolution and Human Behavior*, 32(3), 216–227, https://doi.org/10.1016/j.evolhumbehav.2010.09.003.

Tan, K. W., Graf, B. A., Mitra, S. R., & Stephen, I. D. (2015). Daily consumption of a fruit and vegetable smoothie alters facial skin color. *PLoS One*, 10(7): e0133445, https://doi.org/10.1371/journal.pone.0133445.

Thornton, M. (2002). The biological actions of estrogens on skin. *Experimental Dermatology*, 11(6), 487–502, https://doi.org/10.1034/j.1600-0625.2002.110601.x.

Tiddeman, B. P., Duffy, N., & Rabey, G. (2000). Construction and visualisation of three-dimensional facial statistics. *Computer Methods and Programs in Biomedicine*, 63, 9–20, https://doi.org/10.1016/S0169-2607(00)00072-9.

Tsagkrasoulis, D., Hysi, P., Spector, T., & Montana, G. (2017). Heritability maps of human face morphology through large-scale automated three-dimensional phenotyping. *Scientific Reports*, 7(1): 45885, https://doi.org/10.1038/srep45885.

Weinberg, S. M., Parsons, T. E., Marazita, M. L., & Maher, B. S. (2013). Heritability of face shape in twins: A preliminary study using 3D stereophotogrammetry and geometric morphometrics. *Dentistry 3000*, 1(1), 7–11, https://doi.org/10.5195/d3000.2013.14.

Whitehead, R. D., Re, D., Xiao, D., Ozakinci, G., &

Perrett, D. I. (2012). You are what you eat: Within-subject increases in fruit and vegetable consumption confer beneficial skin-color changes. *PLoS One*, *7*(3):e32988, https://doi.org/10.1371/journal.pone.0032988.

Williams-Blangero, S., & Blangero, J. (1991). Skin color variation in eastern Nepal. *American Journal of Physical Anthropology*, *85*(3), 281–291, https://doi.org/10.1002/ajpa.1330850306.

Zhu, G., Evans, D. M., Duffy, D. L., Montgomery, G. W., Medland, S. E., Gillespie, N. A., ... Martin, N. G. (2004). A genome scan for eye color in 502 twin families: Most variation is due to a qtl on chromosome 15q. *Twin Research*, *7*(2), 197–210, https://doi.org/10.1375/136905204323016186.