Third Party Kinship Recognition

Vanessa Fasolt, BSc, MSc

Submitted for the degree of Doctor of Philosophy (PhD)

Institute of Neuroscience and Psychology

College of Medical, Veterinary & Life Sciences

University of Glasgow

January 2021

Abstract

Research into third party kinship recognition has been sparse even though kinship provides crucial insight into the biological underpinnings of pro-social and sexual behaviour. Furthermore, the studies that have been conducted are of varying quality and consistency, resulting in a myriad of different findings and conclusions. My doctoral research addressed the common issues in the literature by conducting studies using high quality stimuli, a consistent methodology and appropriate analyses.

Study 1 investigated what facial information is used for making kinship judgments in 3D facial images, specifically the contribution of face shape and surface reflectance information (e.g., skin texture, tone, eye and eyebrow colour). 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.

Study 2 investigated whether a smiling facial expression increases the accuracy of judging relatedness compared to a neutral facial expression in human raters. Contrary to expectations, smiling decreased the accuracy of relatedness judgments compared to a neutral facial expression.

Study 3 aimed to replicate previous studies suggesting that birth order affects kinship detection ability. Our findings indicate that laterborns do not have an advantage in detecting child sibling pairs and that kinship judgment accuracy is therefore unaffected by rater birth order.

Study 4 compared the performance of participants across three commonly used methods (i.e., kinship judgment, similarity rating, matching paradigm), using the same highly-controlled stimulus set. We found that while responses on all three tasks were correlated, performance varied significantly across the tasks. Furthermore, when looking at the effect sex and age of the portrayed individuals had on performance, we found that different results are found dependent on which method is used.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to take this opportunity to thank a few people. Firstly, I would like to thank my supervisors Prof Lisa DeBruine and Prof Ben Jones. Thank you for your incredible support since my undergraduate degree, for believing in me and for providing me with incredible opportunities which have changed my professional and also my personal life. I will forever remember these years spent under your tutelage as an incredibly inspiring time, being challenged to think until my brain hurt and then finally having that lightbulb moment.

I would also like the ERC (European Research Council) for funding my doctoral work. I also wish to thank the whole kinship team who have given up so much time to explain things to me, to discuss professional projects and life projects, and who have always been there to pick me up after I got a nasty review or R just wouldn't do what I wanted it to do. Iris, Ant, Kieran and Victor, you guys deserve a medal for putting up with me, and I really miss sharing the office with you and having chats about everything and anything.

There are so many more people to thank, and I can't even begin to explain how much all of your friendships mean to me. I am incredibly lucky to have such a loving and supportive network of friends around me that are always up for a drink and a gossip. Just a few names that need to be mentioned here are Becki, Babsi, Jen, Martin, Amalia, Fraser, Emma, Fran, Nick, Marcia, Ross, Claire, Amanda H and Amanda R, Rachel, Rebecca, Tuchis, Chris and Walzl. You are the best people and I am so lucky to have you guys in my life.

My thanks also go to my parents who always supported me during my degrees and who were always just a phone call away even if we are physically far away from each other.

Finally, I wish to thank Dan. You believe in me a lot more than I do, and I would not be where I am now without you. You have seen the tears, the frustration and the crazy during this process, and I cannot thank you enough for still being here and still supporting me.

Declaration

I declare that this thesis is my own work and was completed under the normal terms of supervision.

Vanessa Fasolt

Publications

Chapter 2:

Fasolt, V., Holzleitner, I. J., Lee, A. J., O'Shea, K. J., & DeBruine, L. M. (2019). Contribution of shape and surface reflectance information to kinship detection in 3D face images. *Journal of Vision*, *19*(12), 9-9.

Chapter 3:

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.

Chapter 4:

Fasolt, V., Holzleitner, I. J., Lee, A. J., O'Shea, K. J., & DeBruine, L. M. (2019). Birth Order Does Not Affect Ability to Detect Kin. *Collabra: Psychology*, *5*(1).

Conferences

Data reported throughout this thesis has been presented at the following conferences:

Conference Talks:

Fasolt V, Holzleitner IJ, Lee A, O'Shea KJ & DeBruine LM (2017). The Effect of Methodology in Third-Party Kin Recognition Research. *Human Behavior and Evolution Society* in Boise, Idaho. May 2017.

Fasolt V, Holzleitner IJ, Lee A, O'Shea KJ & DeBruine LM (2019). Contribution of shape and surface reflectance information to kinship detection in 3D face images. *Human Behavior and Evolution Society* in Boston, Massachusetts. May 2019.

Conference Posters:

Fasolt V, Holzleitner IJ, Lee A, O'Shea KJ & DeBruine LM (2017). Facial expression & 3rd Party Kin Recognition. *European Human Behaviour and Evolution Association* in Paris, France. April 2017.

Fasolt V, Holzleitner IJ, Lee A, O'Shea KJ & DeBruine LM (2017). Third Party Kin Recognition. *6th International Society for Human Ethology Summer Institute* in Boise, Idaho, US. May 2017.

Fasolt V, Holzleitner IJ, Lee A, O'Shea KJ & DeBruine LM (2018). Children resemble women more than men, regardless of kinship status. *European Human Behaviour and Evolution Association* in Pecs, Hugary. April 2018. **(WON THE POSTER PRIZE)**

Fasolt V, Holzleitner IJ, Lee A, O'Shea KJ & DeBruine LM (2019) Birth Order does not affect ability to detect kin. *European Human Behaviour and Evolution Association* in Toulouse, France. April, 2019.

Chapter 1:
Kinship Detection

Kinship, or relatedness, is central to biological theories of social behaviour. Social species such as insects, mammals, primates and birds modulate their behavior according to relatedness (reviewed in Chapais & Berman, 2004). This modulation of behaviour also occurs in humans. Inclusive fitness theory (Hamilton, 1964), explains how natural selection can favour cooperation and prosocial behaviour contingent on relatedness. Optimal Outbreeding theory (Bateson, 1983), suggests that cues of close relatedness may decrease sexual interest to avoid inbreeding costs.

There are two main classes of cues that inform relatedness judgments, namely contextual and phenotypic cues. Contextual cues are for example maternal perinatal association and coresidence (i.e., shared experience in Liebermann, Tooby, & Cosmides, 2007). Contextual cues are not necessarily reliable indicators of genetic kinship, however, in cases such as adoption and children fathered outside the social family. Phenotypic cues are physical cues, so for example odour resemblance and facial resemblance. Here, individuals are matched on phenotype, a mechanism where an individual's different phenotypic cues are compared to one's own or someone else's phenotypic cues to identify relatedness. These cues could be used when contextual cues are uncertain.

It has been demonstrated in previous studies that those who share facial similarities with the observer, hence are phenotype matched to the observer, are favoured in prosocial contexts by the observer. For example, self-resembling faces trigger similar neural substrates as faces of actual kin (Platek & Kemp, 2009). Facial resemblance has been positively correlated with cooperation in economic games (DeBruine, 2002, Krupp et al., 2008), investment in potential offspring (Platek et al., 2003, DeBruine, 2004), and emotional closeness between siblings (Bressan et al., 2009, Lewis, 2011). Facial resemblance however decreases rated attractiveness of the opposite sex (DeBruine, 2005), which again shows how behaviour is biased depending on relatedness and supports the notion of distinct strategies for cooperation and mating.

In addition to the ability to detect those who are related to oneself, humans have the ability to detect kin among others. This ability has been illustrated repeatedly in previous literature, for both parent and children pairs (Alvergne, Oda, Faurie, Matsumono-Oda, Durand & Raymond, 2009; Alvergne, Perreau, Mazur, Mueller & Raymond, 2014; Bressan & Grassi, 2004; Kaminski, Gentaz & Mazens, 2012; Porter, Cernoch & Balogh, 1984) and sibling pairs (Dal Martello & Maloney, 2006; Dal Martello & Maloney, 2010; DeBruine, Smith, Jones, Roberts, Petrie & Spector, 2009). However, the more nuanced findings have been inconsistent and not reproducible, which we will discuss in sections 1.3. and 1.4.

This ability to detect third party kinship could have evolved as a by-product of the ability to detect one's own kin. Another possible explanation for third party kinship detection is that it enabled social species to predict alliances between others (Cheney & Seyfarth, 2004). This latter explanation is supported by evidence showing that macaques recruit allies that are not related with the opposing group in conflicts (Schino, Tiddi, & Di Sorrentino, 2006). In humans, political coalitions have mainly been formed by related males, hence recognizing own relatives but also kinship among others is crucial to inform coallitions (Rodseth & Wrangham, 2004).

This kinship detection ability is also apparent in other species, for example nonhuman primates (Bergman, Beehner, Cheney, & Seyfarth, 2003; Dasser, 1988) which we will investigate in the next section.

1.1. Kinship detection in the non-human primate literature

Recognizing kin is crucial to biological theories of social behaviour, leading to cooperative networks among relatives, inbreeding avoidance, and biased behaviour towards relatives in other groups (Parr & de Waal, 1999; Widdig et al. 2002).

A number of non-human primates have been found to recognise kin based on phenotypic cues such as facial resemblance, including chimpanzees (Parr & de Waal, 1999; Parr et al., 2010), long-tailed macaques (Dasser, 1987), Japanese macaques (Tomonaga, 1994), mandrills (Charpentier et al., 2017) and rhesus macaques (Parr et al, 1999; Pascalis & Bachevalier, 1998; Parr et al., 2010).

Parr et al. (2010) showed that captive chimpanzees and rhesus macaques can detect relatedness in a match to sample task, however, this task required extensive training and the sample size was small. A natural experiment was conducted by Pfefferle et al. (2014) with rhesus macaques using a differential looking time paradigm, as looking time should be prolonged if something is salient to the viewer. Hence, if the rhesus macaque is presented with two stimuli and looks for longer at one than the other it must be able to differentiate between the two stimuli along the dimension they differ on, here relatedness. This is a widely used technique in human infants (Langlois, et al. 1987) and other primate research (Schell et al., 2011; Waitt et al., 2003). They report that free ranging rhesus macaques can discriminate between facial images of their paternal half siblings and unrelated individuals, when both animals are unfamiliar to the tested individual. Specifically, for non-kin, they looked longer at unrelated macaques of the same sex (potential threat) compared to opposite sex (potential mate). Charpentier et al. (2017) conducted a similar study in a semi-free-ranging population of mandrills, and found that after controlling for familiarity, mandrills are able to discriminate between unfamiliar relatives using facial cues alone.

Bower et al. (2012) took facial measurements of related rhesus macaques, unrelated random and unrelated age matched rhesus macaques and compared the 5 principal components derived from an initial principal components analysis across the two groups. The difference in facial measurements was significantly smaller for related macaques than for any of the unrelated macaques, supporting the notion that kinship information is contained in the face.

Parr and de Waal (1999) found that chimpanzees can match mothers and sons when presented with digitized portraits of unfamiliar chimpanzees. The chimpanzees, however, could not match mothers and daughters significantly above chance. This difference in ability to detect kin could be explained from an evolutionary point of view as an adaptive response to the patrilineal structure of chimpanzee communities. In these communities, it is the males that form the stable core of related individuals, whereas adult females are unrelated, having immigrated in from other communities at sexual maturity. It is also the males that show high levels of social affiliation and cooperation, thereby potentially reaping the kin-selected fitness benefits associated with kin-biased social behavior (cf. Hamilton, 1964a, 1964b).

A possible explanation for this bias in females is inbreeding avoidance. A Migrating female might not want to settle in groups where males look like her mother, as the males might be related to her. Parr et al. (2010) tested the inbreeding avoidance theory against a male distinctiveness theory using chimpanzees and rhesus macaques. Both species accurately matched relatives at levels above chance. Furthermore, they argued that the effect of sex, namely the advantage of recognising mother-son pairs over mother-daughter pairs would be reversed in rhesus macaques as here males move between groups, hence they should be better at recognising mother-daughter pairs to avoid inbreeding with an unfamiliar female relative. However, this pattern was not found, instead, rhesus macaques were better at recognising pairs containing a male, namely father-offspring pairs and son-parent pairs, with the best recognition rate for father-son pairs. This suggests that rather than inbreeding avoidance, male distinctiveness is the driving force behind kin recognition in rhesus macaques. These results also indicate that effects of sex in kin recognition in nonhuman primates are not based on a primate-wide mechanism for inbreeding avoidance.

A specialized male distinctiveness face-recognition mechanism might arise in one of two ways. The first is the development of face-recognition mechanisms in the perceiver that are specialized for the detection of facial similarities between mothers and sons: There might be developmental differences in how sons resemble their mother compared to how daughters resemble their mothers. Given the male orientated social organization of this species, selection might then favor the evolution of face-recognition mechanisms specifically tuned to detect traits shared by mothers and sons. Daughters' faces may also resemble their mothers' faces in some features, but the absence of selection pressure for mother– daughter recognition means that no analogous mechanism for the detection of these similarities has evolved.

However, Vokey (2004) proposed an alternative explanation for the bias towards matching mothers and sons compared to matching mothers and daughters. They suggested that characteristics of the faces themselves, namely identifying behaviors (e.g., pose, expression), are used to bias the detection of the facial similarity of mothers and their male offspring. They argue that it is intuitively plausible that, for example, sons might tend to copy (perhaps unintentionally) the characteristic poses and facial expressions of their mothers precisely to encourage the detection of relatedness to her and, thereby, other male offspring. Because adult females disperse from kin at sexual maturity, there would be less requisite selective advantage for them to adopt the poses and expressions of their mothers or siblings.

In short, by one process, sons and daughters both resemble their mothers but in different ways, and selection has favored special recognition mechanisms in perceivers that preferentially detect only mother–son resemblances. Hence, the mechanism is perceiver specific. By the other process, recognition processes in perceivers are unspecialized, and selection has instead favored either the expression of maternal facial characteristics in sons and not daughters or variable behavioral dispositions in sons and daughters to emulate their mother in ways that influence facial appearance. Hence, the mechanism is sender specific. Of course, the two paths could develop concurrently, and the functional result in any case would be the same, namely, that the faces of male as compared with female offspring would be perceived as more like those of their mothers. However, if the process underlying kin recognition were principally the former, then the recognition systems of other species (that are not specifically tuned to the different ways male and female offspring resemble their mothers in chimpanzees) should not preferentially match sons and not daughters to their mothers. That is, they might well be able to detect kin similarity, but not preferentially for sons over daughters. Conversely, if the process underlying kin recognition were primarily the latter, then other recognition systems should respond much as the chimpanzee subjects did, seeing sons' faces as more similar to their mothers' faces than are daughters' faces. To investigate these possibilities, they conducted a series of experiments to test both possible explanations.
Vokey (2004) found that humans only showed a bias towards matching mothers and sons when using the original material, but when eliminating potential framing biases, either by cropping the photos tightly to the faces or by rebalancing the recognition foils, the bias towards matching mother and sons was removed, but not human participants' ability to recognize chimpanzee kin. This supports the notion that kin recognition mechanisms are not perceiver specific, but rather sender specific. It also highlights the importance of carefully controlling stimuli in studies of kin recognition, since confounds in the aspect ratio of the images seemed to be driving the mother-son effect here.

Human and non-human primate infants both show a preference for face-like stimuli over object-like stimuli which suggests that face processing is partly experience-independent in both species (Johnson et al., 1991, Sugita, 2008). However, some primates, namely rhesus macaques (Dufour et al., 2004; Pascalis & Bachevalier, 1998), tonkean macaques, and brown capuchins (Dufour et al., 2006) show a species-specific effect, which means that they are better at recognizing faces of their own species in comparison to faces of another species. This species-specific effect suggests that certain aspects of face processing could be highly dependent on exposure and highly plastic (Dufour et al., 2006; Pascalis & Bachevalier, 1998). This has been shown in chimpanzees, as those raised by humans from an early age showed a preference for human faces over chimpanzees' faces (Tanaka, 2003). A related effect in humans is the 'other-race effect', finding that face recognition is better for faces of the own ethnicity (Lindsay et al., 1991; O'Toole et al., 1994). This other race effect can be reversed by exposure to faces of a different ethnicity, which again suggests that face recognition processes are plastic and dependent on exposure (Sangrigoli et al., 2005). Kazem and Widdig (2013) found that experts were able to detect kinship in rhesus macaques based on facial similarity more accurately than naïve participants with no experience with primates, however, both groups were significantly better than chance at recognising related rhesus macaques.

The processes involved in facial identification and kinship identification using facial resemblance might differ considering the findings that humans can readily detect kinship in common chimpanzees, western lowland gorillas and mandrills, hence detect kinship across species (Alvergne et al., 2009; humans were however not able to detect kinship in baboons, which could be based on the lower facial variation in baboons compared to other species). Facial identification is mainly attributable to configural processing which uses information on the relationships between internal features within the face. Kinship identification has been suggested to use featural processing, which relies on featural information (e.g. the shape of the nose, eyes).

This distinction between facial processing and kinship processing is supported by Alvergne et al. (2009) who found no difference in the ability to judge relatedness for faces of the own ethnicity compared to faces of another ethnicity using exclusively humans. Moreover, Alvergne (2014) found that when participants are shown the wrong configuration of facial features, but with all features still present, humans are still able to detect kinship above chance.

1.1.1. Effect on Behaviour

In primates, evidence for biased behaviour contingent on kin recognition among not directly familiar kin is mixed (see for review: Widdig 2007). For example, in wild chimpanzees, members of the majority of highly affiliative and cooperative pairs are unrelated, and paternal brothers do not selectively affiliate and cooperate with each other (Langergraber et al. 2007). Similarly, paternal half-sisters in white-faced capuchins (Cebus capucinus) do not associate more often than distantly related pairs of females (Perry et al. 2008). In contrast, in free-ranging rhesus macaques and wild yellow baboons (Papio cynocephalus), adult females affiliate more with their paternal half-sisters than unrelated females (Silk et al. 2006; Smith et al. 2003; Widdig et al. 2001, 2002; and see for review in other primates: Widdig 2007). Additionally, female baboons avoid relatives of their aggressor for a longer period than any other unrelated individual (Wittig et al., 2007).

These contrasting findings, rather than questioning the validity or pervasive nature of social biases among unfamiliar (or not directly familiar) kin primates, such as paternal kin (as per: Chapais 2001; Rendall 2004), may reflect responses to different selective forces, including the risks posed by inbreeding, male reproductive skew, kin availability and patterns of sex-biased dispersal.

1.2. Third Party Kin Recognition Literature

Following this general introduction to kin recognition and a look at the non-human primate literature, I will now focus on the main topic of this research: third party kin recognition from facial photographs in humans and the literature that is available on this topic.

As mentioned previously, there is converging evidence that we are able to detect our own kin, but that we are also able to detect kinship pairs among strangers from just face photographs at levels above chance (e.g., 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). However, looking at the studies and the findings in more detail we see that there is little agreement outwith that general statement that we are indeed able to judge kinship accurately given only photos of individuals with no additional information. In this section, we will take a comprehensive and itemized look at the studies that have been conducted. To do this we have identified four main areas that are part of third party kinship studies, namely properties of stimuli (section 1.2.1), properties of participants (section 1.2.2.), methodologies (section 1.2.3.), and analyses (section 1.2.4.).

I have surveyed and am referring to 23 articles in these next sections. These are the only articles that exclusively focus on third party kinship. These studies are also detailed and analysed in the appendix, where the reader can find further details about the specific characteristics of these studies.

1.2.1 Properties of Stimuli

The stimuli used in these studies are all photographs of faces, only one study uses a video, and therefore displays a dynamic face rather than a static face. The stimuli are a major component of a study; hence it is crucial to carefully consider the properties of stimuli and how these might influence the results. These next sections will consider how for example the procurement of stimuli influences the quality of photographs that are used as stimuli, or how different static facial expressions distort facial features and therefore might decrease kinship detection accuracy.

1.2.1.1. Procurement of stimuli

The pictures used in the studies were procured in several ways, some were pictures that families had sent in upon request, depicting individuals in various natural environments, such as a family holiday. Some other pictures were taken by the experimenter in various natural environments, for example pictures taken of the parents and newborns at the hospital. Other pictures were taken by the experimenter under controlled conditions, such as a lab environment with a standardised procedure. The background of the stimuli varied accordingly, as pictures taken in a natural environment can contain information of location when using pictures of a beach holiday, but can also be neutral when using pictures taken in front of a wall for example. Some studies have therefore cropped the images to contain the faces but not any background information, and altered the background to a uniform colour such as black, grey or white. When cropping the pictures some studies cropped the faces to include the hair of the stimuli, others excluded the hair and cropped the picture tightly around the outline of the face.

Removing background information is useful as it could influence kinship judgments; raters might base their kinship judgments on information such as similar backgrounds (e.g., the same beach holiday) rather than facial resemblance. Moreover, including or excluding the hair of the stimuli removes some information that could help or hinder kinship judgments. Dyed hair could hinder and natural hair could help judgments if the natural hair colour is the same between the depicted kin. Clothing and hairstyle can even provide social/cultural cues used to match people likely to belong to the same family. To ensure that kinship judgments are made based on facial cues rather than other information, it is helpful to exclude any extraneous information.

1.2.1.2. Colour of photograph

20% of studies used black and white photographs while the remaining 80% studies used colour photographs. This could lead to a difference in results, as black and white photographs exclude colour information present in the face. While colour information from skin tone, eye colour, and hair/eyebrow colour can provide information about genetic relatedness, it also provides information about environmental similarity. Both are valid cues of kinship, since kin are more likely to share the same environment than non-kin, and kinship cues can still be used and useful even if not 100% accurate. Colour information, such as redness, yellowness and tan might be used as a cue to kinship among individuals who share a household as these colour cues are partially influenced by environmental factors. Namely, exposure to approximately the same diet, the same life events and partaking in approximately the same level of exercise could lead to similar levels of yellowness, redness and tan of faces. For individuals who do not share a household, colour information might be less indicative of kinship, although the skin tone of the face and colour of the eyebrows and eyes could still indicate kinship relations, with tanning potential also being genetically determined.

One study using black and white photographs and colour photographs found that kinship judgment accuracy was higher for black and white pictures than for colour pictures (Kaminski et al., 2010). This might be explained by the colour photographs containing colour information that confuses rather than helps the rater in interpreting kinship cues. This could be the case for kin who do not share the same lifestyle and therefore display different levels of facial yellowness, redness and tan. It could also be the case for kin who do not have the same eyebrow and eye colour as the kin displayed in the same trial, as genetic relatedness does not necessarily mean sharing the same hair and eye colour. Black and white pictures might draw more attention to the facial shape information shared among kin, which could hold more stable kinship information than colour information over time. Yet, children's facial shape, especially the lower half of the faces, changes considerably with age and throughout puberty, hence facial shape might be more informative as kinship cue in individuals of similar age or adults. Yet, black and white pictures do still give some indication of the lightness or darkness of features, and therefore some colour information could still be used to evaluate kinship in black and white pictures.

We investigated this contribution of shape and surface reflectance information (e.g., skin tone, texture, eye colour) in our own study (Fasolt et al. 2019, experimental chapter 7) to determine the importance of these two different cues on kinship judgments. We found that both shape and surface reflectance information contribute equally to kinship detection, with raters being able to judge relatedness accurately when only shape or surface reflectance information was present. Raters were, however, most accurate at detecting kinship when they were shown the same stimuli with both shape and surface reflectance information (the original picture).

1.2.1.3. Facial expression

Another difference between the stimuli used in different studies of kin recognition is the facial expression of the person depicted. Most studies use pictures displaying a neutral facial expression, however as some are holiday pictures sent in by families, facial expression may vary. This could lead to differences in results, as smiles might distort facial features which could be used for kinship detection. Anecdotally, similarities between smiles have been commented on as indicative of kinship, however, no study has so far looked at the effect of different facial expressions on kinship judgments. One study looking exclusively at smiling facial expressions used a short video of dynamic smiles of parent-child pairs and their computer algorithm was successful at detecting kinship at levels above chance. However, this study only used the computer algorithm for 3rd party kinship detection and did not compare accuracy levels to human raters. Moreover, the task only comprised smiling stimuli, therefore accuracy levels could not be compared to neutral stimuli. Our own study (Fasolt et al., 2018; experimental chapter 8) directly compared performance accuracy for neutral and smiling faces, finding that a smiling facial expression decreases kinship detection accuracy. This is a crucial finding, as some studies are unclear on what facial expression their stimuli displayed. This could mean that a study using stimuli with different facial expressions might find a difference in accuracy levels between pairs based on a difference in facial expression, but attributing this difference to other factors rather than facial expression. One infamous study (Christenfeld, Hill, 1995) that generated results that failed to replicate in further studies used stimuli with varying facial expressions, which might be why the findings have never been replicated with neutral faces. (However, it is important to note here that the study suffered many further shortcomings.)

1.2.1.4. Aspect Ratio

As mentioned in the section examining kinship relations and detection in non-human primates, aspect ratio is important to control as it can bias the accuracy of raters, independent of research question. Some studies control for it while some others don’t. It is unclear from a lot of publications whether this has been done or not.

1.2.1.5. Ethnicity

The stimuli in most studies are European, with a couple of studies using stimuli from the U.S. (Alvergne, Perreau, et al., 2014; McLain et al., 2000) and one study using stimuli from Senegal (Alvergne, Oda et al., 2009). The latter study used Senegalese and French raters and found no difference in performance when detecting parent-child pairs from the same or another ethnicity. This study suggests that ethnicity might not impact kin detection accuracy, nevertheless, the very limited range of ethnicities in studies to date concerning third party kin recognition does not allow us to conclude anything decisive about the role ethnicity plays in kin detection and hence needs to be expanded on in the future.

1.2.1.6. Age

The age of the collected stimuli varies widely, with some of the youngest stimuli being one to three days old (Alvergne et al., 2007; McLain et al., 2000). Most studies, however, do not provide the age of the parents, hence a maximum age is not known. DeBruine et al. (2009) reported their adult siblings to be up to 46 years old, and Dibeklioglu, Ali Salah, and Gevers (2013) reported their oldest stimuli to be 76 years old.

The ages of the related stimuli should be known and matched with the age of the unrelated stimuli to ensure that age cues are not driving the results. So, for example if a study is employing a 1-3 matching task whereby a baby is matched to three potential adults, the adults should all be around the same age.

Some studies took pictures of the same individual at different time points to investigate the effect of age of the stimuli (Brédart, French, 1999; Christenfeld, Hill, 1995). This longitudinal measure allows the researcher to directly investigate whether age has an effect on parental resemblance and judgment accuracy, while controlling for the possibility that an age effect is found due to individual differences between pairs' general resemblance.

The results of differences in ages of stimuli will be discussed in detail in section 1.3.

1.2.1.7. Sex

Most studies collected and used both male and female stimuli and both same-sex and different-sex pairs. The groups are not always equal, which on one hand could mean that the results are biased as same-sex pairs might be easier to judge than different-sex pairs, as the rater is not tasked with comparing two sexually dimorphic faces. In a matching task this could also lead to same-sex pairs being more easily matched than different-sex pairs. The myriad of findings based on the effect of sex in stimuli will be explored in section 1.3., and illustrate why it is so important to control for the possible effects sex of the stimuli can have by keeping the numbers equal.

On the other hand, this unequal number of female and male stimuli could mean that unrelated control pairs might not have the same sex constellation as the related pairs, hence resulting in findings that are not based on actual relatedness but rather on sex information. In this case the findings for related and unrelated pairs cannot soundly be compared with each other.

1.2.2. Properties of Raters

The raters are another important component of these studies, hence it is crucial to understand the properties of the raters that took part in the research and how these properties might have influenced the findings. These next sections will explore these properties, for example, how the number of raters in a study has important implications for the power of the study to actually detect an effect.

1.2.2.1. Number of Raters

Studies vary widely in how many raters they recruited, with numbers ranging from 50 raters (Porter, Cernoch, Balogh, 1984) to 362 raters (Alvergne et al., 2009). 60% of studies on third party kinship recognition studies have around 60 to 140 raters.

This number is especially important to determine whether the study had the power to detect an effect with the given number of participants. None of the studies so far explicitly reported a power calculation. This could be one of the factors that have led to the different findings in the field, as too low numbers of raters (especially in studies with low stimuli numbers) could mean that a true effect was not detectable due to low statistical power only. Or it could also mean that positive results were false positives or the effect sizes was overestimated.

Moreover, depending on whether the studies were within or between subjects' designs, the number of raters could be cut into smaller groups, which again might lower the statistical power to find a true effect.

1.2.2.2. Family network

Kaminski, Ravary, Graff and Gentaz (2010) found an effect of birth order. They proposed that the older siblings have a disadvantage in the ability to judge relatedness from facial features, as contextual cues such as perinatal association with the mother in addition to phenotypic cues informed their understanding of having a sibling. Younger siblings do not have any exposure to the perinatal cues, hence might rely more on phenotypic cues of kinship. However, there are still other contextual cues available to younger siblings apart from phenotypic resemblance such as cohabitation. Alvergne et al. (2014) and our own paper (Fasolt et al., 2019; experimental chapter 3) did not find birth order to affect kinship judgment accuracy, hence it is unclear how birth order influences ability to judge kinship exactly, if at all.

1.2.2.3. Age

The age of the raters varies between studies, however, most studies report a mean rater age in the mid-twenties. This stems from the recruitment of mainly undergraduates for studies. The age of rater and stimuli are not matched, hence, individuals of all ages judge facial resemblance of individuals who are younger and/or older than themselves in these studies. This could again lead to a bias, as exposure to faces of the same age as the rater might increase accuracy in judging facial cues. This possibility is further explored in section 1.4.

One study (Kaminski, Gentaz, & Mazens, 2012) divided raters into 6 rater age groups, with all the raters being between five and eleven years old. Each rater age group comprised individuals of the same age. They were in turn judging the facial resemblance between neonates (mean age of 110 hours) and their parent, so again a mismatch between the raters' age and the stimuli's age shown to the rater. It would have been interesting to assess the ability to judge child siblings of roughly the same age as the age groups to identify any advantage of exposure to own age faces.

All other studies did not divide raters into age groups, or did not indicate so.

1.2.2.4. Sex

Most studies recruited male and female raters, with a few studies not providing any information about the sex of rater. The numbers are seldom equal, which could lead to a bias in results. One study (Arrantes & Berg, 2012) only had male raters.

A few studies looked at the effect of rater sex on the ability to judge relatedness based on facial resemblance and found mixed results. This will be discussed in section 1.4. in detail.

1.2.2.5. Ethnicity

Most studies used raters and stimuli of the same ethnicity, focusing mainly on Caucasians. One study (Alvergne, Oda, Faurie, Matsumoto-Oda, Durand, & Raymond, 2009) used Senegalese and French raters and stimuli and found no difference in performance when assessing another ethnicity's facial resemblance. Oda, Matsumoto-Oda, and Kurashima (2002) used Japanese raters and stimuli.

As discussed before, the ability to recognize kinship cues does not show an "own race" bias and therefore suggests a different mechanism to be at play.

1.2.2.6. Other

Other factors that have been included in one study by Nesse, Silverman and Bortz (1990) were years of education and marital status, neither had a significant effect on the ability to judge kinship.

Marital status could theoretically influence the importance of assessing kinship in different-sex individuals, as married individuals are not looking for a mating partner and therefore recognising kinship cues to avoid inbreeding is irrelevant. On the other hand, by marriage, individuals gain more relatives, which might increase the need to judge kinship based on phenotypic matching, as contextual cues such as cohabitation are not available.

General cognitive abilities do not seem to significantly influence the ability to judge kinship, as small children are already proficient in making kinship judgments (however there is a possible increase in accuracy with age). And as shown by Nesse, Silverman and Bortz (1990), years of education also do not increase the accuracy in judging kinship. This suggests that kinship recognition is a mechanism that develops independently from other abilities.

1.2.3. Methodology

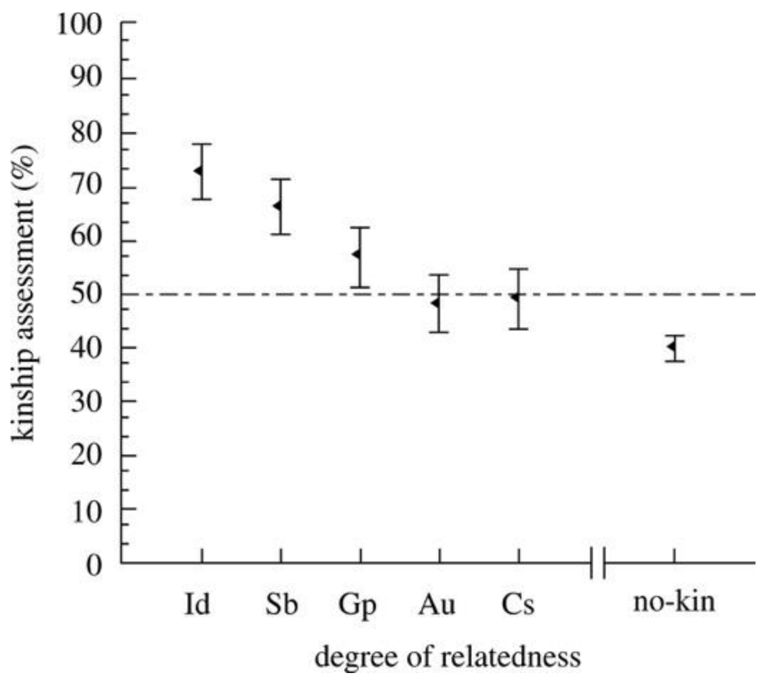
The methodology used in the studies is another crucial component that can have wide-ranging effects on the findings hence it is important to consider. These next sections look at different aspects of the methodology and their considerations, for example, how the study arrangement and task can influence the results or how the number of stimuli used can result in a memory task rather than a kin recognition task.

1.2.3.1. Degree of Kinship

Most studies looked at parent-child resemblance, with a couple exclusively looking at mother-daughter resemblance and one exclusively looking at father-son resemblance. Some studies also looked at sibling resemblance, with one study including non-identical twins (DeBruine et al., 2009). Only one study (Kaminski et al., 2009) has looked at facial resemblance in the wider family circle including siblings, aunts, uncles, grandparents and cousins. They showed that raters were able to accurately identify related and unrelated pairs at levels above chance no matter what degree of kinship. However, when analysing the effects of degree of kinship by multiple logistic regression with judge as random effect they found that sibling pairs and grandparent-grandchild pairs were judged to be related more often than by random chance, while cousin pairs and aunt/uncle-niece/nephew pairs were not judged to be related more often than by random chance (Figure 1). Yet, the latter were still judged differently than unrelated pairs, as those were significantly identified as unrelated pairs. This indicates that overall the closer the degree of kinship the more they are judged to be related, arguably based on a higher degree of facial resemblance between the closer kin.

Further investigations are needed to determine the effect of degrees of kinship on kin detection accuracy and what this means in terms of morphological differences and facial resemblance between kin of different degrees.

Figure 1: Kinship judgments based on varying degrees of kinship (Kaminski et al., 2009) Id = Same Individual at different ages, Sb = Siblings, Gp = Grandparent, Au = Aunt/Uncle, Cs = Cousins



1.2.3.2. Number of Stimuli used

The number of unique photographs of stimuli varied between studies, with studies with 30 unique stimuli (Bressan, Dal Martello, 2002) to up to 332 unique stimuli (Alvergne, Faurie, Raymond, 2007)

The stimuli were then employed differently in different studies, resulting in different numbers of trials. In a matching task where one trial requires a minimum of 4 unique stimulus photographs, the trial numbers were lower than in tasks showing pairs of faces, as one trial requires only two unique stimulus photographs. Hence, some matching studies had only10 trials (McLain, Setters, Moulton, Pratt, 2000)

Some studies used a unique stimulus only once in the study, therefore preventing any possible exposure effects, while some studies reused the same stimuli in a number of trials, hence using them as experimental and control stimuli. One way of controlling for a possible exposure effect when reusing the same stimuli is to employ a between-subjects design, so raters still only see a stimulus once even though the number of available stimuli to the researcher has been maximally used. Yet, some studies (Bressan, Dal Martello, 2002) show the same stimuli multiple times to the same rater, possibly confounding the results. For example, if a face was previously seen with a very strong resemblance to its paired face, that may affect responses when the same face is seen later paired with different faces and decrease the subsequent kinship judgment artificially. This again could be one of the reasons why the results across the field vary widely.

The number of stimuli used can also influence the statistical power of the studies, with a smaller number of stimuli leading to lower statistical power to detect a true effect. Additionally, most studies do not treat stimuli as sampled from a larger population, so the statistical conclusions are really only applicable to the exact stimuli used.

1.2.3.3. Arrangement & Task

The photographs and tasks were displayed in numerous ways. Most studies conducted their experiment on a computer (e.g. DeBruine et al., 2009), some studies printed out the photographs and pinned them up on boards, with the raters walking from one board to the next (e.g. McLain et al., 2000), and in some other studies raters were given photo albums with the stimuli photographs (e.g. Bressan & Grassi, 2004).

Various tasks have been employed in the literature. The most common task employed is matching one target stimulus to the real relative out of a number of stimuli. In most studies, this task comprises of one child or parent target and three possible parent or child options. Other variations are showing raters one neonate and six adults, with 3 being possible mothers and three being possible fathers (McLain et al., 2000), showing raters one neonate and 4 possible mothers or one adult female and 4 possible children (Porter et al., 1984), or showing raters one target female and two possible mothers (Arrantes & Berg, 2012). One of these possible relatives is always actually related to the target.

Other studies used a binary kinship task, whereby raters were shown two stimuli and then had to decide whether they thought the stimuli were related or not. These studies always had a mix of related and unrelated stimuli pairs.

Another task often employed is asking raters to judge the resemblance between stimuli. This is not explicitly mentioning or asking about kinship and the underlying idea is that facial similarity is a physical cue to kinship and similarity judgments can therefore be translated into kinship judgments. There are issues with this line of reasoning as DeBruine et al. (2009) showed that these two judgments are highly correlated but not necessarily synonymous.

One issue of using different methodologies is that studies are not directly comparable. Maloney and Martello (2006) and DeBruine et al. (2009) compared kinship judgments of pairs and similarity ratings of the same pairs, finding that they are highly correlated but as stated above not necessarily synonymous. We (Fasolt et al., 2019, experimental chapter 5) directly compared the three methods used in the literature and found that they significantly differ from each other in terms of kinship judgment accuracy levels and in terms of finding an effect of sex and age of stimuli. The use of different methods might be one explanation for how studies find such varying answers to the same questions.

Another issue that results from employing different tasks is that raters across methodologies are shown a different number of stimuli within the tasks. The matching task is showing one target stimulus and at least two possible relatives, which allows the rater to compare not only kinship cues between the target and the possible relative, but also between the possible relatives. So rather than just making a judgment about whether one person is related or not to the target stimulus, it is a judgment about which possible stimulus is more likely to be the actual relative taking into consideration all stimuli. Therefore, the matching task seems to ask a slightly different question from the kinship task and might result in different accuracy levels as raters know that there is an actual relative in the set and they have more context information to choose the actual relative.

1.2.3.4. Timing

In most studies raters were able to view the stimuli as long as they wanted with no time limit and no reaction time measurement taken. Only two studies limited their viewing time of the stimuli, with one study (Kaminski, Méary, et al. 2010) limiting the viewing time in a matching task to 25 seconds whereafter raters had to make their choice, with 5 second intervals between trials. In another study (Kaminski et al., 2009), they limited the viewing and decision time to 20 seconds. In this latter study, once the raters had indicated whether they thought a pair was related or unrelated the next trial would appear. Accuracy levels in both studies were still significantly above chance, hence a time restriction does not seem to have impacted raters' judgments negatively. However, it must be noted that 20 and 25 seconds of exposure are still a considerable amount of time for each trial, therefore it would be informative to implement different time limitations to conclude how fast accurate kinship judgments can be made. Another way of measuring this would be reaction times.

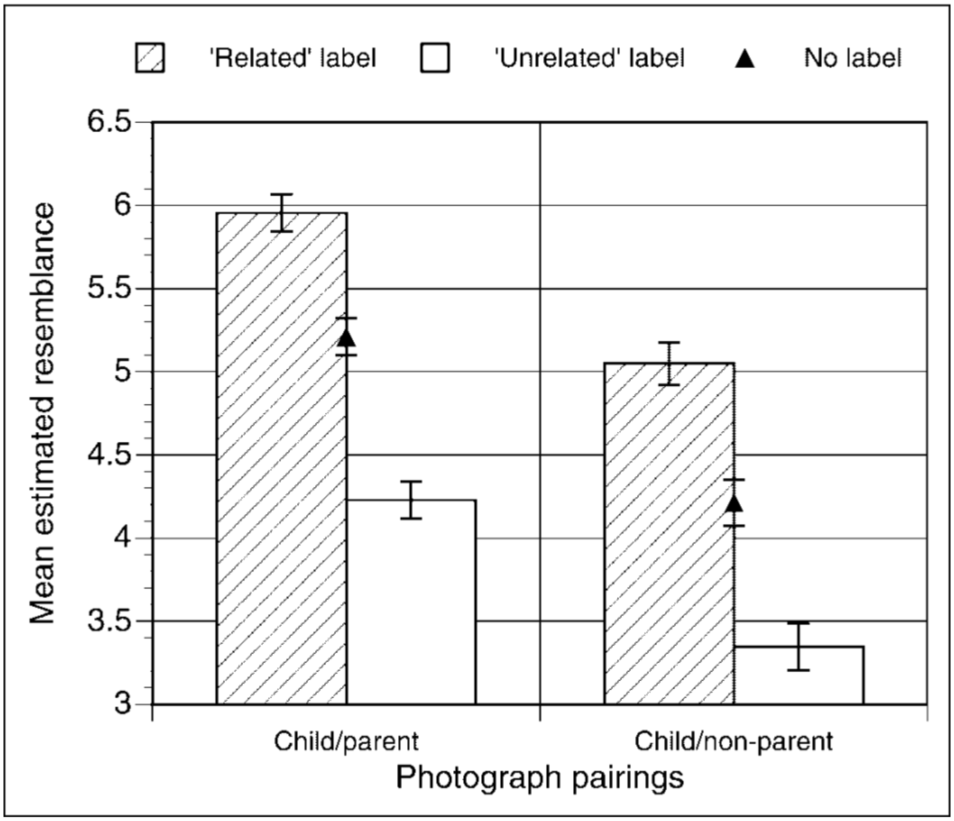
Kaminski et al. (2009) recorded the reaction times of the raters and found that raters took approximately 7 seconds to judge whether a pair was related or unrelated. When comparing reaction times between kin and non-kin pairs they did not find a significant difference, meaning that raters take roughly the same amount of time to judge related and unrelated pairs. However, when taking into consideration the degree of relatedness (from sibling to cousins) and accuracy there was a significant difference in reaction times, whereby raters were quicker to judge closer kin correctly than further removed kin or in cases their judgments were wrong. This suggests that reaction time could give us an interesting insight into the length of time it takes to process kinship information, and how factors such as degree of kinship or task can influence processing time.

1.2.3.5. Information about Kinship

Studies give different information about the kinship status of the stimuli shown to the raters. Most studies told the raters that they were about to complete an experiment concerning kinship detection, with an indication that some of the shown pairs were related but some might not be, however it isn't always explicitly stated what exactly the instructions were. Some studies were more specific and correctly told raters that half of the stimuli pairs were related and half of them were unrelated (Dal Martello, Maloney, 2006; 2010; DeBruine et al., 2009; Maloney, Dal Martello, 2006).

One study investigated the effect of labelling pairs as related or unrelated on similarity judgments (Bressan & Dal Martello, 2002). They found that being told that a pair was related was the main driver of similarity judgments (Figure 2 from Bressan and Dal Martello, 2002)

"***Figure 2. Mean estimated resemblance as a function of genetic relatedness (child-parent vs. child-nonparent) and belief in relatedness ("related" label vs. "unrelated" label) in Experiment 1. Filled triangles show the mean estimated resemblance when there was no information about relatedness (data from Experiment 2: no labels). Bars indicate the standard error of the mean.*** ", Bressan and Dal Martello, 2002.



1.2.4. Analyses

Analyses vary between studies. It is important to consider the unit of analysis when evaluating the findings of past studies. The unit of analysis found in the literature is either between (raters or stimuli) or within (rater and stimuli). Using only the raters or the stimuli as unit of analysis means that the characteristics of only one group is taken into account, ignoring the other group. This makes the results not generalizable to other groups of stimuli or other groups of raters, depending on the unit of analyses used. Including both the raters and the stimuli as units of analyses takes into account the characteristics of both the raters and the stimuli, making it a more robust and generalizable analyses.

1.3. The Effects of Stimulus Sex and Age

In this chapter, we will take a closer look at the effect of two properties of stimuli which have been intensely debated in the literature, namely the effect of sex of stimuli and the effect of age of stimuli.

1.3.1. The Effect of Sex of Stimuli

From an evolutionary perspective, paternity uncertainty could lead to an increased facial resemblance of children to their father to counteract negative effects of paternity uncertainty on care behaviour. Some research has shown that men are willing to invest more in a self-resembling child, while women are unaffected by self-resemblance to children (Platek et al., 2003, 2004) and that men's hypothetical adoption decisions are correlated more strongly with self-resemblance than women's adoption decisions (Volk & Quinsey, 2007). However, other research has found no difference in men's and women's preference for and investment in self-resembling child faces (DeBruine, 2004), or even found a preference for self-resembling children in women but not men (Bressan, Bertamini, Nalli, & Zanutto, 2009). Indeed, only one study has found that children resemble their fathers more than their mothers, and specifically, this was only found for one-year-old children (Christenfeld & Hill, 1995). This finding has never been replicated (Brédart & French, 1999; French et al., 2000). In contrast, the possibility of infidelity means that paternal resemblance could be disadvantageous and costly for children conceived outside the social pair (Daly & Wilson, 1996; French et al., 2000). Moreover, paternal resemblance can also be costly for males, considering that extramarital children could be identified and disadvantaged (Marlowe, 1999). Studies conducting interviews with relatives and observing family interactions with newborns found that the belief of resemblance is established and nurtured primarily by relatives commenting on a resemblance between fathers and their children, rather than by a strong phenotypic resemblance between children and their fathers (Alvergne et al., 2007; Daly & Wilson, 1982; McLain, Setters, Moulton, & Pratt, 2000; Regalski & Gaulin, 1993).

So, if children do not necessarily resemble their fathers more than their mothers, are there any systematic biases in who they resemble most? One study found that children resemble their mothers more than their fathers (McLain et al., 2000). Yet, this is the only study finding this specific result. In line with theories suggesting that sexually dimorphic facial characteristics influence face judgments, two studies found that boys resemble their fathers more and that girls resemble their mothers more (Alvergne et al., 2009; Kaminski et al., 2010). Another study found that children look most similar to females in general, with girls resembling females and mothers more than males or their fathers, and boys resembling females more, but both parents equally (Bressan & Dal Martello, 2002). Similarly, Bressan and Grassi (2004) found that children are rated to look more similar to females than males, but when taking into account only resemblance to the real parents, no effect of sex was found. This general resemblance to females could be based on the fact that young children's faces have attributes that are considered feminine facial traits, such as big eyes, round faces, and high eyebrows. However, it is unclear why this effect is not found when judging the resemblance of children to their parents. Maybe other kinship cues partially override any sexually dimorphic information used in similarity judgments of related pairs. Moreover, significant facial changes occur during puberty which might decrease the generic feminine facial traits in boys and decrease resemblance to females (Kohn, 1991). This suggests that the role of sexually dimorphic facial cues on kinship judgments and similarity judgments is not fully understood yet. Significantly less research has looked at siblings rather than parent-child pairs. DeBruine et al. (2009) found that unrelated same-sex pairs received higher similarity ratings than unrelated opposite-sex pairs, while sex composition had no effect on similarity ratings of sibling pairs, suggesting that when assessing facial similarity sexual dimorphism cues might play a role. To round it all off, some studies do not detect an effect of sex at all (Brédart & French, 1999; Kaminski, Dridi, Graff, & Gentaz, 2009; Maloney & Dal Martello, 2006).

Table 1: Summary of all findings in the literature concerning the effect of sex on third party kin recognition:

Author	**Finding**
Alvergne, A., Faurie, C., & Raymond, M. (2007)	Differential resemblance varies according to age, with boys resembling their mother more when newborn and then resembling their father more when between 1 and 5 years old. Girls resemble their mothers more than their fathers, at all ages considered
Alvergne, A., Oda, R., Faurie, C.,	For both French & Senegalese judges, interaction sex

Matsumoto-Oda, A., Durand, V., & Raymond, M. (2009)	child- sex parent= girls resembling more the mother and boys more their father
Brédart, S., & French, R. M. (1999)	No effect
Bressan, P., & Dal Martello, M. F. (2002)	Experiment 1: Children look more like women than men in general; girls resemble more females & mothers, boys resemble females more, but both actual parents equally
	Experiment 2: Children look more similar to female (M=5.13) than male (M=4.29) adults F(1,58)=81.84, p<.0001
	Experiment 3: Children look more similar to female (M=5.19) than male (M=4.75) adults F(1,78)=31.27, p<.0001
Bressan, P., & Grassi, M. (2004)	Children are rated to look more similar to females than males in general, but when taking into account only ratings for real mother and real father there is no difference in who children are rated to look more similar
Christenfeld, N. J., & Hill, E. A. (1995)	1 year olds look more like father than mother
DeBruine, L. M., Smith, F. G., Jones, B. C., Roberts, S. C., Petrie, M., & Spector, T. D. (2009).	In similarity judgment, unrelated same sex pairs rated to look more similar than opposite sex pairs

Kaminski, G., Dridi, S., Graff, C., & Gentaz, E. (2009)	Significant difference between MM pairs and FF pairs (Male-male 76.8% right match, female-female 67.8%); When other kinship degree better when a woman in pair than MM pairs; No significant difference in accuracy between same gender vs opposite gender pairs
Kaminski, G., Gentaz, E., & Mazens, K. (2012)	No neonate sex effect, but parent sex effect & interaction neonate and parent sex effect: male better chance being associated with neonate than female parent Contrast comparisons showed a neonate sex effect in male parents' items, with boys having a greater chance of being matched than girls (y=1.23 [1.04–1.45]; p=0.017), but no neonate sex effect in female parents' items
Kaminski, G., Méary, D., Mermillod, M., & Gentaz, E. (2010	No gender main effects, but an interaction: Girls more frequently paired with their Mothers (odd-ratio=1.29) and boys were more frequently paired with fathers (odd-ratio= 1.27)
Maloney, L. T., & Dal Martello, M. F. (2006).	No effect
McLain, D. K., Setters, D., Moulton, M. P., & Pratt, A. E. (2000)	Experiment 1: Neonates matched to mothers at significantly higher rate than fathers (p<.05)

	Experiment 2: Neonates matched to mothers at significantly higher rate than fathers ($p<.05$)
Nesse, R. M., Silverman, A., & Bortz, A. (1990)	Main effect sex parent= pairs including a mother were matched at higher rates than those with fathers; no effect sex child
Oda, R., Matsumoto-Oda, A., & Kurashima, O. (2002)	Only in condition 3 (reversed sex indication): significant sex difference in that boys are judged to resemble fathers more than girls resemble their fathers

1.3.2. Effect of Age of Stimuli

Age of the stimuli has also been suggested to influence kin recognition, yet again, contradictory findings do not allow us to conclude what this effect is. A few studies find that age does not affect kin recognition (Kaminski et al., 2009; Maloney & Dal Martello, 2006; Nesse et al., 1990). However, Alvergne, Faurie and Raymond (2007) found that newborn boys resemble their mothers more than their fathers, but between the ages of two and three years an inversion occurs, and they resemble their fathers more than their mothers. For girls, this inversion does not occur, as they resemble their mothers more at any age. Brédart and French (1999) found that raters were better at matching five-year-old boys to their parents than younger boys, while there was no such age effect for girls. Furthermore, Christenfeld and Hill (1995) found that one-year-old children resemble their fathers more than their mothers, with older children not being accurately matched to their parents at all. For siblings, DeBruine et al. (2009) found that age difference had an effect on similarity ratings but not kinship judgments, which could indicate an interaction between the effect of age and methodology used in studies. However, age and sex composition of the stimuli pairs were confounded, as the age difference in opposite sex pairs was larger than in same sex pairs (DeBruine et al., 2009).

An interesting question addressed by a couple of studies was what part of the face informs kinship judgments most. Dal Martello and Maloney (2006) found that the upper part of faces is crucial when judging the relatedness of children. This importance of the upper face could also facilitate judging the relatedness between children and adults. Children might express more kinship cues in the upper part of their face as the lower part is significantly developing and changing throughout childhood and adolescence. This importance of the upper part of the face might be reduced when judging the relatedness between adults.

Yet, in one study using pictures of adults between 21-26 years, judges were not able to match the father and son pairs when showing only the lower half of the face (Alvergne, Perreau, Mazur, Mueller, & Raymond, 2014). This suggests that the upper half of the face is crucial to make kinship judgments independent of the age of the stimuli.

Table 2: Summary of all findings in the literature concerning the effect of age on third party kin recognition:

Author	**Finding**
Alvergne, A., Faurie, C., & Raymond, M. (2007)	The global resemblance of children to their parents tended to increase with the age of girls, not significantly for boys.
Alvergne, A., Oda, R., Faurie, C., Matsumoto-Oda, A., Durand, V., & Raymond, M. (2009)	Experiment 1: No effect
	Experiment 2: The scores obtained by French judges were not associated with the age of the child (p = 0.11). However, the scores were negatively associated with the age of the mother (p = 0.03). The scores obtained by Senegalese judges were

	not associated with the age of the mother ($p = 0.94$). However, their scores were associated with the age of the child ($p < 0.01$). For Senegalese judges living in France, age had no effect.
Brédart, S., & French, R. M. (1999)	For boys only, better at matching 5-year-olds than younger boys to parent
Christenfeld, N. J., & Hill, E. A. (1995)	1 year olds look more like father than mother, at no other age children can be matched to parents
Kaminski, G., Dridi, S., Graff, C., & Gentaz, E. (2009)	No effect
Maloney, L. T., & Dal Martello, M. F. (2006)	No effect
Nesse, R. M., Silverman, A., & Bortz, A. (1990)	No effect of age of child (no indication about effect of age of parent)
DeBruine, L. M., Smith, F. G., Jones, B. C., Roberts, S. C., Petrie, M., & Spector, T. D. (2009)	Effect of age in similarity rating task (but confounded with sex of stimuli, as average age difference between opposite-sex pairs was greater than the average age difference between same-sex

1.4. The Effects of Participant Sex and Age

In this chapter, we will take a closer look at the effect of two properties of participants which have been intensely debated in the literature, namely the effect of sex of participants and the effect of age of participants.

1.4.1. The Effect of Sex of Participants

Similar to the reasoning for why there might be an effect of sex of stimuli, there might be an effect of sex of participant when judging kinship. As previously discussed, it might be beneficial or detrimental for the child to look like the biological father depending on the possibility of infidelity (Daly & Wilson, 1996; French et al., 2000). And rather than children actually resembling their fathers more than their mothers, it seems that increased resemblance is a socially reinforced concept mostly driven by relatives disproportionally attributing the physical appearance of the child to the father (Alvergne et al., 2007; Daly & Wilson, 1982; McLain, Setters, Moulton, & Pratt, 2000; Regalski & Gaulin, 1993). This reinforced believe of self-resemblance was linked to increased (self-reported) likelihood of parental investment in some studies (Platek et al., 2003, 2004) and could therefore be beneficial to the child.

In a nutshell, there are three possible effects the sex of participants could have on kinship judgments:

1) There is a main effect of sex, whereby men are better at judging kinship than women. One explanation for this may be that based on paternity uncertainty they need to be able to identify their own offspring and other men’s offspring to invest their resources only in their own child and own genetic future
2) There is a main effect of sex, whereby men are worse at judging kinship than women. This may be because they believe relatives commenting on the child’s resemblance to the father, even though there is no actual increased phenotypic resemblance. A decreased sensitivity to facial resemblance could lead to the acceptance of that bias.
3) There is no effect of sex of participant, and men and women are equally accurate at detecting kinship.

A few studies looked at the effect of sex of participant (see table 3), with most of them finding no effect of sex of participant (Alvergne, Oda, Faurie, Matsumoto-Oda, Durand, & Raymond, 2009; Alvergne, Perreau, Mazur, Mueller, Raymond, 2014; Kaminski, Gentaz, & Mazens, 2012; Porter, Cernoch, & Balogh, 1984). Bressan and Grassi (2004) found that men give higher similarity ratings than women in general but are not more accurate, yet this effect of men giving higher resemblance ratings was not actually significant with a p-value of .058. In another study, Bressan and Dal Martello (2002) did find a significant effect of sex of participant, whereby men gave higher resemblance ratings than women in general ($F(1, 58)$ 4.02, $p = .049$). Nesse, Silverman and Bortz (1990) found that men were better at detecting the parent of sons rather than daughters, and that women were better at detecting the parent of daughters than sons. This would suggest a bias towards being able to process and judge kinship in own-sex faces more accurately than other-sex faces, however, there was no interaction between the sex of participant and the sex of parent, hence an advantage of judging own-sex faces seems unlikely. Furthermore, Bressan and Dal Martello (2002) found in the first part of their study that when analysing men and women's judgments separately, a sex difference was noted when participants thought that the faces shown were related: women rated the child as more similar to the mother (female-superiority index), while men rated the child to resemble the father and mother to the same extent. When participants thought that the faces shown were unrelated there was no sex difference in kinship judgments. In the second part of this study the researchers did not give participants any indication of whether the stimuli were related or not, and in this condition, there was no effect of sex found. This suggests that women's responses are only biased when they believe that the stimuli shown to them are related. In the third part of this study they found that there is only a sex difference in resemblance ratings when judges are shown male-male stimuli pairs. There was no difference in resemblance rating for any other pair constellation.

To conclude, these findings suggest that sex of participant does not play a crucial role in kinship detection, even if it might marginally bias judgments in specific situations.

Table 3: Summary of all findings in the literature concerning the effect of sex of participant on third party kin recognition:

Author	**Finding**
Alvergne, A., Oda, R., Faurie, C., Matsumoto-Oda, A., Durand, V., & Raymond, M. (2009)	No effect
	No effect
Alvergne, A., Perreau, F., Mazur, A., Mueller, U., & Raymond, M. (2014)	No effect
Bressan, P., & Dal Martello, M. F. (2002)	Marginally significant= males give slightly higher resemblance ratings, F(1, 58) 4.02, p= .049 Women rated the child as more similar to the mother (female-superiority index), men rated the child as resembling the father and mother to the same extent. When participants thought that the faces shown to them are unrelated there was no sex difference.
	No effect
	Difference between female and male raters for male-male stimuli pairs *t*(78)

	3.68, *p* .0001, no difference for other pairs
Bressan, P., & Grassi, M. (2004)	Male judges gave slightly higher ratings for similarity than females (mean +- S.E.M. =4.31 +- 0.14 vs 3.8 +- 0.11, F(1,78)=3.69, p=.058). But no difference in accuracy
Kaminski, G., Gentaz, E., & Mazens, K. (2012)	No effect
Nesse, R. M., Silverman, A., & Bortz, A. (1990)	No main effect, interaction sex rater - child = men are better at judging relatedness of sons than daughters, and women are better at judging the relatedness of daughters than sons; no interaction sex rater - parent
Porter, R. H., Cernoch, J. M., & Balogh, R. D. (1984)	No effect

1.4.2. The Effect of Age of Participants

A small number of studies included the effect of age of the judge in their analyses. Facial recognition abilities improve with age and exposure to faces as the cognitive ability of configural processing facilitates the processing of relational information between facial features (see Farah, Wilson, Drain, & Tanaka, 1998). However, the effect of age on kinship detection is unclear as the results from the few studies investigating this issue are mixed (see table 4). Two studies report no effect of age on the ability to detect kin (Alvergne, Oda, Faurie, Matsumoto-Oda, Durand, & Raymond, 2009; Alvergne, Perreau, Mazur, Mueler, &

Raymond, 2014). One study found an effect of age in their Senegalese group of judges but did not indicate what this effect was (Alvergne, Oda, Faurie, Matsumoto-Oda, Durand, & Raymond, 2009). Kaminski, Gentaz and Mazens (2012) found an effect of age of judge, whereby kinship detection marginally increased with increasing age (p=.07), which would support the theory that facial processing improves with age. However, this effect was only found in one condition of the study. Moreover, it has been established that kin recognition and facial recognition are two different processes, with kin recognition using features rather than configuration information that is used in facial identification (Alvergne, Oda, Faurie, Matsumoto-Oda, Durand, & Raymond, 2009; Alvergne, Perreau, Mazur, Mueler, & Raymond, 2014). Hence it is not surprising that age has an impact on face recognition but not necessarily on kin recognition also.

Table 4: Summary of all findings in the literature concerning the effect of age of participant on third party kin recognition:

Author	**Finding**
Alvergne, A., Oda, R., Faurie, C., Matsumoto-Oda, A., Durand, V., & Raymond, M. (2009)	No effect
Alvergne, A., Oda, R., Faurie, C., Matsumoto-Oda, A., Durand, V., & Raymond, M. (2009)	Effect of age (p< .001) for Senegalese judges only (unknown what direction)
Alvergne, A., Perreau, F., Mazur, A., Mueller, U., & Raymond, M. (2014)	No effect
Kaminski, G., Gentaz, E., & Mazens, K. (2012)	Marginally getting better with increased age (p=0.07); effect only found for adult comparison panel

1.5. Open science

None of the extant research on third party kin recognition has been pre-registered, but it is important to note that most of these studies were conducted before pre-registrations were widely introduced. None of the studies mentioned have shared their data or analysis plan on open access platforms such as the Open Science Framework to increase the visibility and reproducibility of the research, but it is again worth mentioning that this was not common practice or even not available when most of these studies were conducted.

I am committed to open and reproducible science; hence all my studies are pre-registered and all our data and code are available online at the Open Science Framework. Moreover, we pre-print our manuscripts to make them available to everyone and publish in open access journals.

- **Experiment 1:** Contribution of shape and surface reflectance information to kinship detection in 3D face images
  - Open Science Framework project with pre-registration: osf.io/7ftxd
  - Pre-print: psyarxiv.com/7b56y/
- **Experiment 2:** Facial expressions influence kin recognition accuracy
  - Open Science Framework project with pre-registration: osf.io/58ewu/
  - Pre-print: Fast turn-around times from writing to publication in open journal, hence it was already accessible to everyone as soon as possible
- **Experiment 3:** Birth order does not affect ability to detect kin
  - Open Science Framework project with pre-registration: osf.io/h43ep/
  - Pre-print: psyarxiv.com/d2vy5/
- **Experiment 4:** Methods comparison in third party kin recognition; or how everyone finds a different answer to the same question
  - Open Science Framework project with pre-registration: osf.io/a3t8x/
  - Pre-print: No pre-print yet

1.6. What factors influence third party kinship recognition?

The past chapter has introduced various factors that influence kinship judgments, and how these factors might have influenced studies on third party kinship judgments. The next four experimental chapters will address some of the issues identified in the extant literature while determining what factors influence third party kinship judgments.

All of the following studies use highly standardized and high-quality stimuli.

In Chapter 2, I demonstrate that the face holds important kinship cues, with both shape and surface reflectance contributing to kinship detection and enabling the rater to draw correct conclusions about the relatedness of complete strangers.

In Chapter 3, I demonstrate that facial expressions can impact kinship detection rate, whereby a neutral face increases the likelihood of correctly identifying related and unrelated pairs, while a smiling facial expression actually decreases the likelihood of correctly identifying related and unrelated pairs.

In Chapter 4, I shift the focus from factors pertaining the stimuli to factors that are pertinent to the rater. Specifically, here I demonstrate that birth order of the rater does not impact their ability to judge relatedness from face images.

In Chapter 5, I address a wider issue that needs to be addressed in the literature, namely the exact methods used in the previous literature. I demonstrate that this is a factor that can crucially influence the outcomes of a study and therefore lead to incorrect conclusions based on purely methodological effects.

These following four experimental investigate important questions about what factors influence third party kinship judgments, yet they are only a tentative start in trying to understand the exact nature of third party kin recognition.

Chapter 2:

Experiment 1

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

Vanessa Fasolt*[1], Iris J Holzleitner[1], Anthony J Lee[2], Kieran J O'Shea[1], Lisa M DeBruine[1]

[1] Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

[2] Division of Psychology, University of Stirling, Stirling, UK

*Corresponding author: Vanessa Fasolt

Published: Fasolt, V., Holzleitner, I. J., Lee, A. J., O'Shea, K. J., & DeBruine, L. M. (2019). Contribution of shape and surface reflectance information to kinship detection in 3D face images. *Journal of Vision*, *19*(12), 9-9.

2.1. Author Contribution

Contributor Role	Role Definition	Initials
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims.	VF, LD
Methodology	Development or design of methodology; creation of models.	VF, LD
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.	VF, LD, IH, AL
Validation	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.	//
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.	VF, LD
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.	VF
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.	VF, LD, IH, AL, KO
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse.	//
Writing – Original Draft Preparation	Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).	VF
Writing – Review &	Preparation, creation and/or presentation of the published work by those from the original research group,	VF, LD

Editing	specifically critical review, commentary or revision – including pre- or post-publication stages.	
Visualization	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.	VF, LD
Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.	//
Project Administration	Management and coordination responsibility for the research activity planning and execution.	LD, VF
Funding Acquisition	Acquisition of the financial support for the project leading to this publication.	LD

2.2. Abstract

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 colour). Using 3D facial images, 195 participants were asked to judge the relatedness of one hundred 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, pre-registration, code and data are available on the Open Science Framework (osf.io/7ftxd).

2.3. 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 (using, e.g., 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, inter-ocular 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; R. D. Whitehead, Re, Xiao, Ozakinci, & Perrett, 2012), redness has been positively linked to skin vascularisation 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 colour can be an informative cue of kinship, as eye colour 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 colour 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 colour alone in allocentric kin recognition, as both eye colour and shape were obscured. Still, observing a decrease in accuracy suggests that the eye region is to some extent an informative cue to kinship which 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 colour. 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.

2.4. Methods

The methods and analyses for this study were pre-registered 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 pre-registration. Additional non-preregistered analyses are clearly marked and improved visualisations of findings have been added.

2.4.1. Stimuli

Face images were collected from children visiting a local science centre, 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 90cm 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 colour

3D surface images using six standard digital cameras (Canon EOS100D; lenses: Canon EF 50 mm f/1.8 STM). Two remote-controlled flash units (Elinchrom D-Lite RX 2) 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) and Blender (Blender Foundation) 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 standardised photographs, from which all background information was removed.

The use of highly standardised 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 centre 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 2000 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.

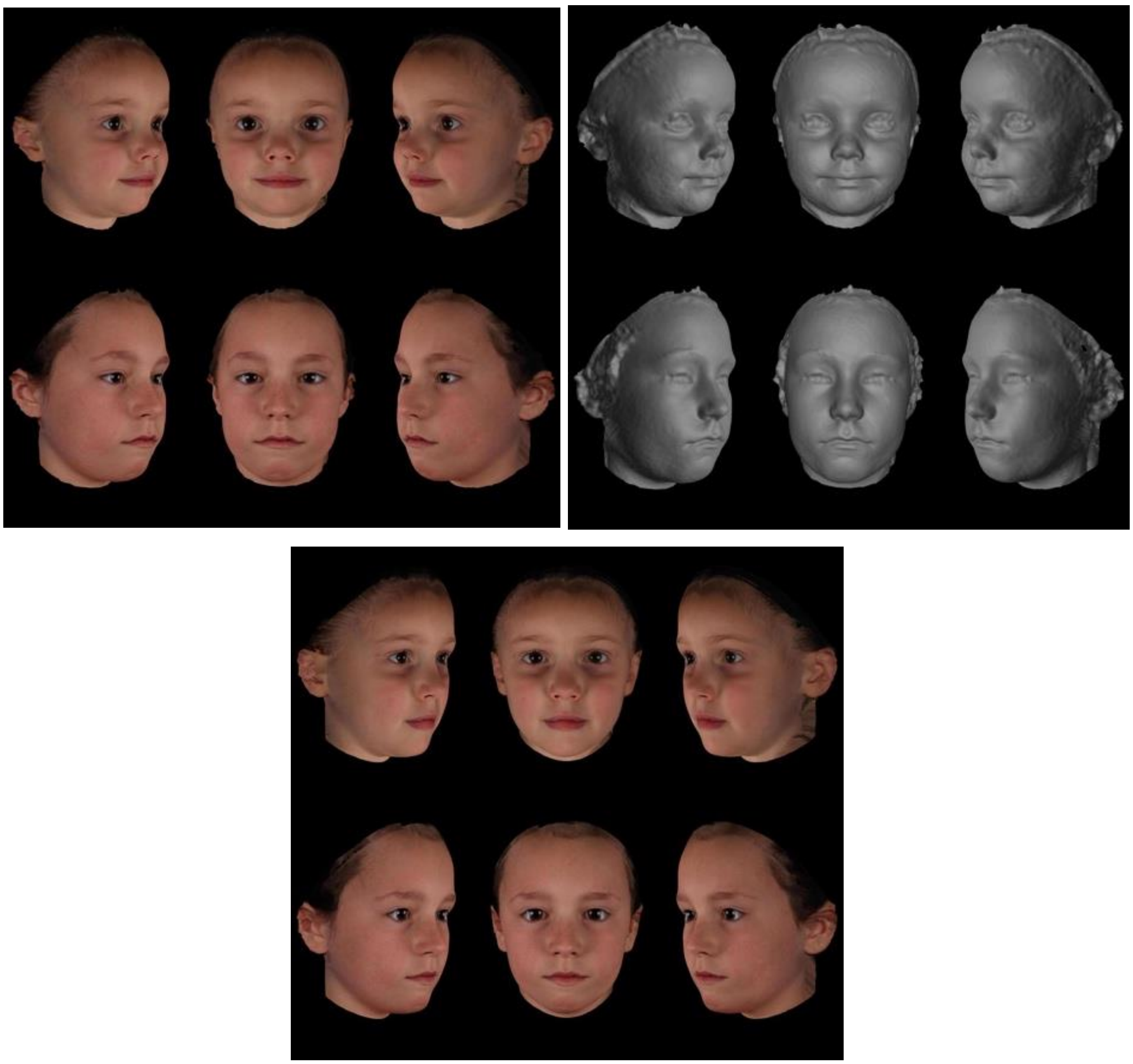
This matching procedure is crucial as it ensures that there is no interdependencies of stimuli within the set, as this could result in judgment biases. For example, most studies in the field use individuals from one family as both experimental and control stimuli, hence the same faces are seen in multiple 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 degrees, frontal view and +40 degrees, see Figure 3).

Figure 3. 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).



2.4.2. 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 labelled 'unrelated' or 'related' without any time restrictions.

2.4.3. 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 pre-registration, 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/. Including 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 analysed. Six raters (mean age = 30.46; SD = 5.18) did not indicate their gender. Most raters identified as white (155 out of 195 raters).

2.4.4. 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) and 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 generalise beyond the specific set of stimuli used, while analyses aggregating over stimuli do not generalise beyond the specific raters. These limitations are overcome in a mixed model analysis where responses are not aggregated.

2.5. Results

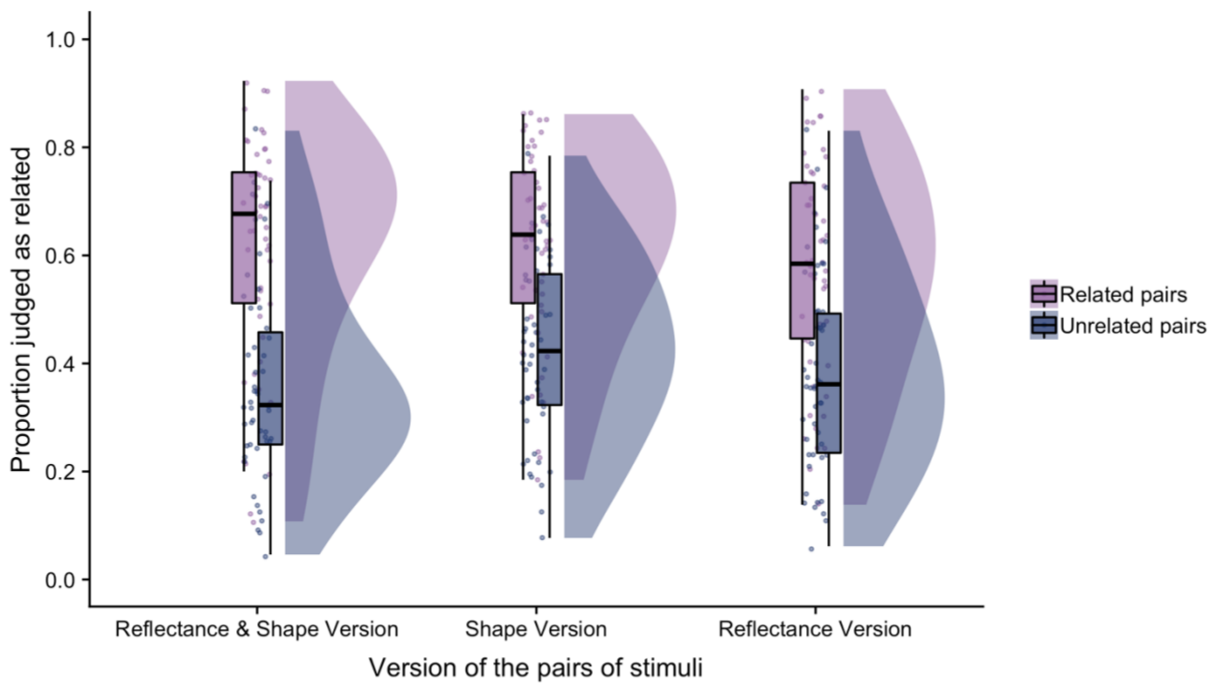
Supporting hypothesis 1, we found a main effect of relatedness (ß=0.96, SE=0.17, z=5.73, $p < .001$), whereby actually related pairs were 2.61 times more likely to be judged as related than unrelated pairs (see Figure 2).

Hypothesis 2 was partially supported by our results (see Figure 4). As predicted, there was a significant positive interaction between relatedness and shape information (ß=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 (ß=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 5).

Table 5. Results from main analysis

Effect	***Estimate (ß)***	***SE***	***z***	***p***	***odds ratio***
Intercept	0.00	0.10	-0.03	0.973	1.00
Relatedness	0.96	0.17	5.73	< .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

Figure 4. The effects of stimulus version and actual relatedness on average kinship judgments (0 - 'unrelated judgment', 1 -'related judgment'). The boxplots, points and distributions represent the average relatedness score for each individual stimulus pair. The boxplots are showing the median, first and third quartile, and the lower (Q1 - 1.5*IQR) and upper (Q3 + 1.5*IQR) extreme relatedness score for related (pink) and unrelated (blue) pairs. The kernel density distributions 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.



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 pre-registration. 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 6).

Table 6. 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.

Version	*Related Pairs*	*Unrelated Pairs*	*Estimate*	*se*	*z*	*p*	*odds ratio*
Reflectance & Shape Version	61.7%	36.2%	1.25	0.21	6.08	< .001	3.49
Reflectance Version	57.2%	38.6%	0.95	0.20	4.75	< .001	2.59
Shape Version	61.7%	42.4%	0.98	0.18	5.35	< .001	2.66

Following Dal Martello and Maloney (2006), we conducted a signal detection analysis obtaining estimates of sensitivity *d'* and likelihood criteria *ß*, which allowed us to further examine performance rates in the three different versions of the stimuli (Green, Swets, & others, 1966). Performance accuracy in all three versions was above chance, which was indicated by a *d'* value being significantly bigger than 0 (see Table 7). 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* ($z = -3.558$, $p < .001$) and skin *reflectance version* ($z = -4.022$, $p < .001$) compared to the *reflectance and shape version.* Performance rates in the *shape version* and the *reflectance version* did not differ from each other ($z = -0.464$, $p = 0.643$).

Table 7. The d' estimate and the likelihood criterion ß 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.

Version	***d***	***d_SD***	***beta***	***beta_SD***	***z***	***p***
Shape Version	0.491	0.032	0.974	0.008	15.557	< .001
Reflectance Version	0.470	0.032	1.025	0.008	14.649	< .001
Reflectance & Shape Version	0.652	0.032	1.019	0.011	20.278	< .001

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 et al., 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= .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.

2.6. 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, DeBruine, & Maloney, 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, Stark, Walsh, Jardine, & Martin, 1981; Djordjevic, Zhurov, & Richmond, 2016; Frisancho, Wainwright, & Way, 1981; Kim et al., 2013; Tsagkrasoulis, Hysi, Spector, & Montana, 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 minimise 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 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 modelling 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 colour or eye shape is used as kinship cue, as both are heritable (Larsson, Pedersen, & Stattin, 2003; Tsagkrasoulis, Hysi, Spector, & Montana, 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 near 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.

Chapter 3:

Experiment 2

Facial expressions influence kin recognition accuracy

Vanessa Fasolt*[1], Iris J Holzleitner[1], Anthony J Lee[2], Kieran J O'Shea[1], Lisa M DeBruine[1]

[1] Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

[2] Division of Psychology, University of Stirling, Stirling, UK

*Corresponding author: Vanessa Fasolt

Published: 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.

3.1. Author Contribution

Contributor Role	Role Definition	Initials
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims.	VF, LD
Methodology	Development or design of methodology; creation of models.	VF, LD
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.	VF, LD, IH, AL
Validation	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.	//
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.	VF, LD
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.	VF
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.	VF, LD, IH, AL, KO
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse.	//
Writing – Original Draft Preparation	Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).	VF
Writing – Review &	Preparation, creation and/or presentation of the published work by those from the original research group,	VF, LD

Editing	specifically critical review, commentary or revision – including pre- or post-publication stages.	
Visualization	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.	VF, LD
Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.	//
Project Administration	Management and coordination responsibility for the research activity planning and execution.	LD, VF
Funding Acquisition	Acquisition of the financial support for the project leading to this publication.	LD

3.2. Abstract

Kinship informs the allocation of pro-social and sexual behaviour. In addition to the ability to detect kin who are directly related to the observer, humans are also able to detect relatedness among others who are not related to themselves based on facial cues of relatedness. However, it is unclear what exact facial cues inform these kinship judgments. Facial expression might be one candidate, as it has been shown that a computer kin-detection algorithm can match relatives accurately when the stimuli are smiling. The current study investigated whether a smiling facial expression increases the accuracy of judging relatedness compared to a neutral facial expression in human raters. The stimuli were images of 50 sibling pairs and 50 unrelated pairs (aged 3-17 years) matched for age, ethnicity and sex. The stimuli included both neutral and smiling versions of each individual. Raters (N=77) were asked to judge whether the presented pairs were related or not in one of two counterbalanced versions of the study, where the same stimuli were never presented as both smiling and neutral to the same rater, and the expression within the pair was always the same. Binary relatedness judgments were analysed using binomial logistic mixed regression. Contrary to expectations, smiling decreased the accuracy of relatedness judgments compared to a neutral facial expression. When shown with a smiling expression compared to a neutral one, related pairs were judged to be related less often, while unrelated pairs were judged to be related more often. Evidence that the upper face is mostly used for kinship judgments suggests that smiles could distort or distract from other, more reliable cues of kinship. Pre-registration, data and code available at https://osf.io/58ewu/.

3.3. Introduction

Humans, along with other animals, possess the ability to distinguish between kin and non-kin, which is integral to the development of social, sexual and parental behaviours (Chapais & Berman, 2004; Hepper, 2005; Lieberman, Tooby, & Cosmides, 2007). The ability to detect kin allows individuals to favour their relatives by displaying prosocial behaviour (Hamilton, 1964) and optimise their reproductive behaviour by avoiding inbreeding (Bateson, 1983). One cue used for kin recognition is visual processing of physical similarities, or *phenotype matching* (for a review, see Penn & Frommen, 2010). Research shows that those who share facial similarities with the observer are favoured in social contexts (see DeBruine, Jones, Little, & Perrett, 2008 for a review). For example, studies have shown that in economic games, raters displayed increased levels of cooperation and trust with players whose faces were more similar to their own (DeBruine, 2002; Krupp, DeBruine, & Barclay, 2008). Similarly, experimentally increased facial resemblance results in increased intentions about investment in children (DeBruine, 2004; Platek et al., 2003), while perceptions of facial resemblance between siblings predict altruistic behaviours and emotional closeness (Lewis, 2011).

In addition to the ability to detect kin who are directly related to the observer, humans also demonstrate the capacity to detect relatedness among others who are not related to themselves. This ability is referred to as *allocentric kin recognition* and has been illustrated repeatedly in previous literature, for both parent-child pairs (Alvergne, Perreau, Mazur, Mueller, & Raymond, 2014; Bressan & Dal Martello, 2002; Bressan & Grassi, 2004; Nesse, Silverman, & Bortz, 1990) and sibling pairs (Dal Martello, DeBruine, & Maloney, 2015; DeBruine et al., 2009; Maloney & Dal Martello, 2006).

Limited research has looked into which facial cues are actually used when making kinship judgments. Dal Martello and Maloney (2006) conducted two experiments to determine where kin recognition signals are in the human face. In both experiments, observers were shown pairs of photographs of children’s faces and asked to judge sibship. The first experiment measured performance when either the upper or lower half of the face were masked, and the second experiment measured performance when either the mouth or eye region were masked. They found that kin judgment accuracy deteriorated significantly when the upper half of the face was masked, but found no difference in kin recognition when the lower face was masked. This suggests that cues used for kinship judgments are situated primarily in the upper half of a child's face. These findings highlight that different areas of the face are of different importance when making judgements about relatedness.

Following research finding that the right and left cerebral hemispheres are differentially involved in the perception of facial emotions and face identities (Butler et al., 2005; Coolican, Eskes, McMullen, & Lecky, 2008; Megreya & Havard, 2011; Rhodes, 1985), lateralisation of the face as a kinship cue has also been considered. However, Dal Martello and Maloney (2010) concluded that neither the left side of the face or the right side was superior in revealing information of kinship and that symmetry cues were not utilised when evaluating kinship. Dal Martello, DeBruine & Maloney (2015) also did not find an effect of inversion or rotation on kinship judgment accuracy, although face inversion disrupts other perceptions such as identity and expression. The results of the above studies suggest that the process of kinship detection is specialised and differs from the way in which other features of the face are processed.

Based on the above evidence, it is unclear what role facial cues of emotions might play in kinship judgments. Some evidence suggests that a smiling facial expression aids some facial judgments, as for example nationality is identified correctly more frequently when the stimuli's facial expression is smiling rather than neutral (Marsh, Elfenbein, & Ambady, 2003, 2007). Consequently, it might be that facial expressions can be cues to kinship, or enhance kinship recognition. However, facial expressions might be processed differently from cues to kinship and hinder kinship recognition. One study somewhat addressed this question, using a computer verification task to assess kinship of short videos of faces showing a dynamic, spontaneous smile (Dibeklioglu, Ali Salah, & Gevers, 2013). The computer verification task achieved a kinship detection accuracy of 73%, which is slightly superior to human kinship detection rates. Most human kinship detection studies have used stimuli with neutral expressions, so it is unclear what effect a smiling facial expression would have on kinship detection accuracy. Moreover, the computer verification task did not compare its accuracy levels for smiling faces to accuracy levels for neutral faces, therefore very little can be said about whether a smiling facial expression influences kinship recognition at all.

Nevertheless, based on this successful computer verification task of smiling kin and Marsh, Elfenbein and Ambady's (2003,2007) findings that smiles aid nationality identification, smiles might be a helpful cue to kinship also. Smiles might partially be a contextual cue of kinship, with smiles within a family being more similar than smiles of strangers.

In light of the above, the current investigation is the first to explore the effect of facial expression on human raters' ability to recognise kin. This will help to provide further information about which factors can influence allocentric kin recognition. We hypothesised that relatedness will have a main effect, whereas raters are more likely to judge related pairs as related, and that a smiling facial expression will increase the accuracy of this judgment compared to a neutral expression.

3.4. Methods

The methods for this study were pre-registered on the Open Science Framework at https://osf.io/ujnfp/.The planned analysis script is available at this site, as well as details about the hypotheses, stimuli and procedure. All procedures below follow this pre-registration exactly. The final data and analysis including improved visualisations and additional analyses suggested by reviewers can be found at https://osf.io/ggc79/. Any not-preregistered analysis is pointed out clearly.

3.4.1. Stimuli

Stimuli were collected from children visiting the Glasgow Science Centre who volunteered to take part in a study of facial cues of family relatedness. Parental consent and child assent were obtained for each child to use their face photograph in studies of family resemblance detection. Children were photographed with a smiling expression and then a neutral expression looking straight at the camera with hair pulled back and any glasses, scarves, and hats removed. The specific procedures for image collection are available at https://osf.io/bvtnj.

From a set of approximately 1500 images of individuals of varying ages, sex and relatedness, we algorithmically chose the maximum number of sibling pairs fitting a number of criteria. Both siblings were required to be genetically related and non-twin full siblings under the age of 18. We also required that an age-matched (within 1 year), ethnicity-matched, and sex-matched foil image was 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. We are not able to exclude the possibility that stimuli might be distantly related without our knowledge.

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, meaning that at least one pair was born within 12 months of each other without being twins.

3.4.2. Procedure

Recruitment of raters was done online through social media (e.g., Facebook) and social bookmarking sites. The study itself was completed online at faceresearch.org and lasted around 10 minutes.

Raters were randomly assigned to one of two counterbalanced versions of the study. Each rater was presented with 100 stimuli pairs, which were presented in a random order. Half of these stimuli pairs were shown smiling and half with a neutral expression, which ensured that raters rated both smiling and neutral faces. Raters were, however, never shown the same pair with both expressions, as the pairs that were shown smiling in one version of the study were shown neutral in the other version. 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." They were shown one pair of child faces at a time and chose their answer by clicking on buttons labelled "unrelated" or "related" without any time restrictions. We do not know whether any of the raters were familiar with any of the individuals shown during the study, however, recruitment for data collection and recruitment for the online study were done separately and on separate platforms. Photographs were mainly taken of local families in the local science centre, while raters from all over the world took part in the online study, making it unlikely, but not impossible, that they would know a small number of the individuals shown.

3.4.3. Raters

The kinship task was started by 81 people; we excluded 4 raters who did not rate all 100 stimuli, and were therefore left with 77 raters for analyses. After the exclusions, the distribution of raters looked as follows: 40 raters completed version A of the study and 37 raters completed version B of the study.

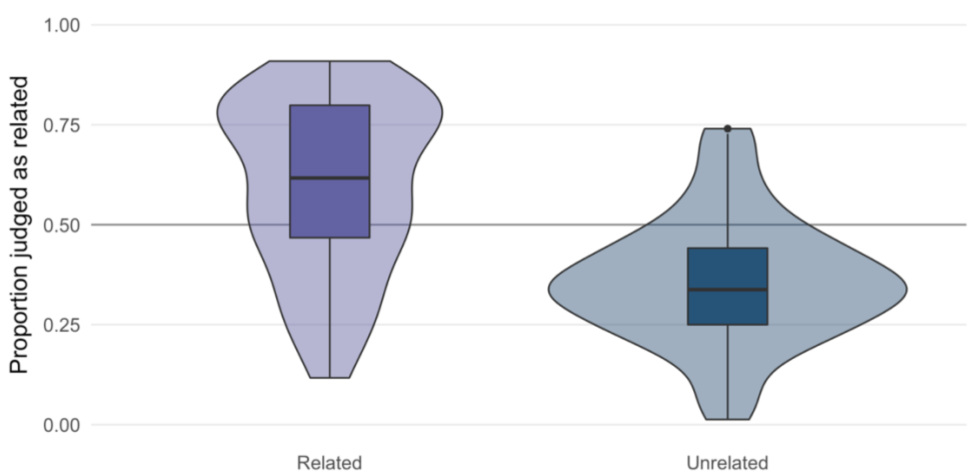
The responses from the two versions were analysed together in order to being able to compare the performance of the raters within pairs of stimuli. Overall, the responses from 28 men (mean age = 26.89; SD = 12.5) and 49 women (mean age = 26.15; SD = 11.27) were analysed.

3.5. Results

We used a logistic mixed model to predict relatedness judgments from actual relatedness (effect coded as related = +0.5 and unrelated = -0.5), facial expression (effect coded as smiling = +0.5 and neutral = -0.5) and the interaction between facial expression and relatedness. We included the rater ID and stimulus ID as random effects and specified our random 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).

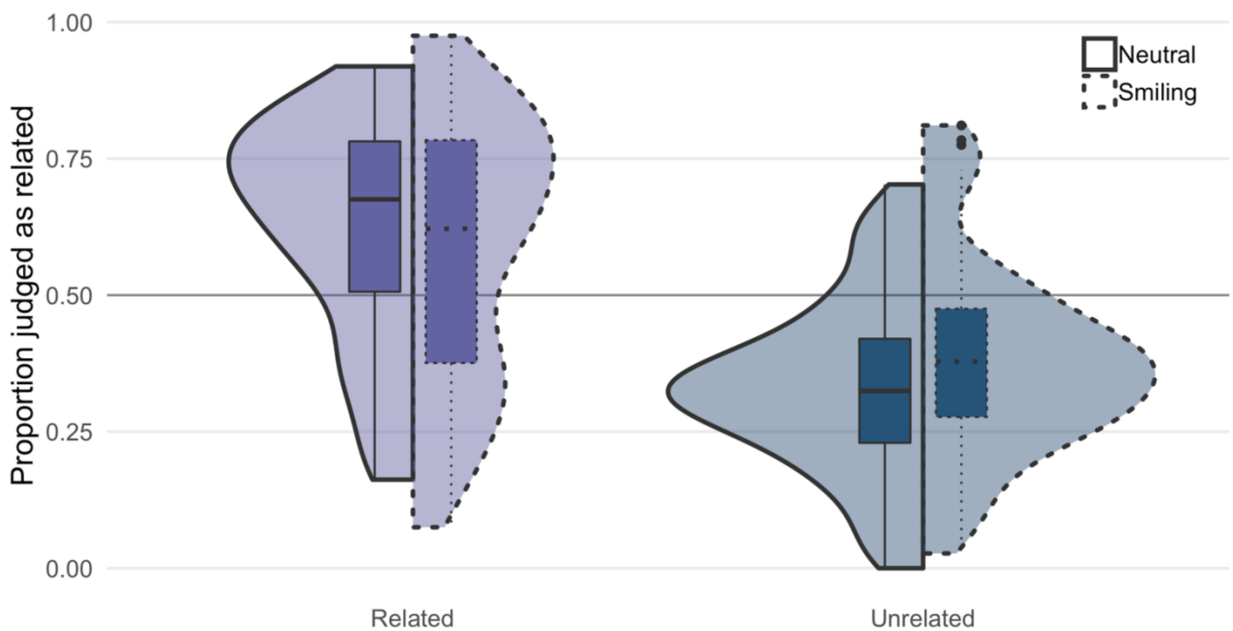
As predicted, the analysis revealed a main effect of relatedness (*ß*=1.19, *SE*=0.19, *z*=6.09, *p* < .001, *odds ratio*=3.29), whereby related pairs were 3.29 times more likely to be judged as related and unrelated pairs were 3.29 times more likely to be judged as unrelated (see Figure 5). Both correct related judgments for related pairs (*ß*=0.48, *SE*=0.16, *z*=2.97, *p* = .003, *odds ratio*=1.62) and correct unrelated judgments for unrelated pairs (*ß*=-0.71, *SE*=0.14, *z*=-5.22, *p* < .001, *odds ratio*=0.49) were significantly above chance (not pre-registered hypotheses/analyses).

Figure 5. The main effect of relatedness on proportion of face pairs judged as related.



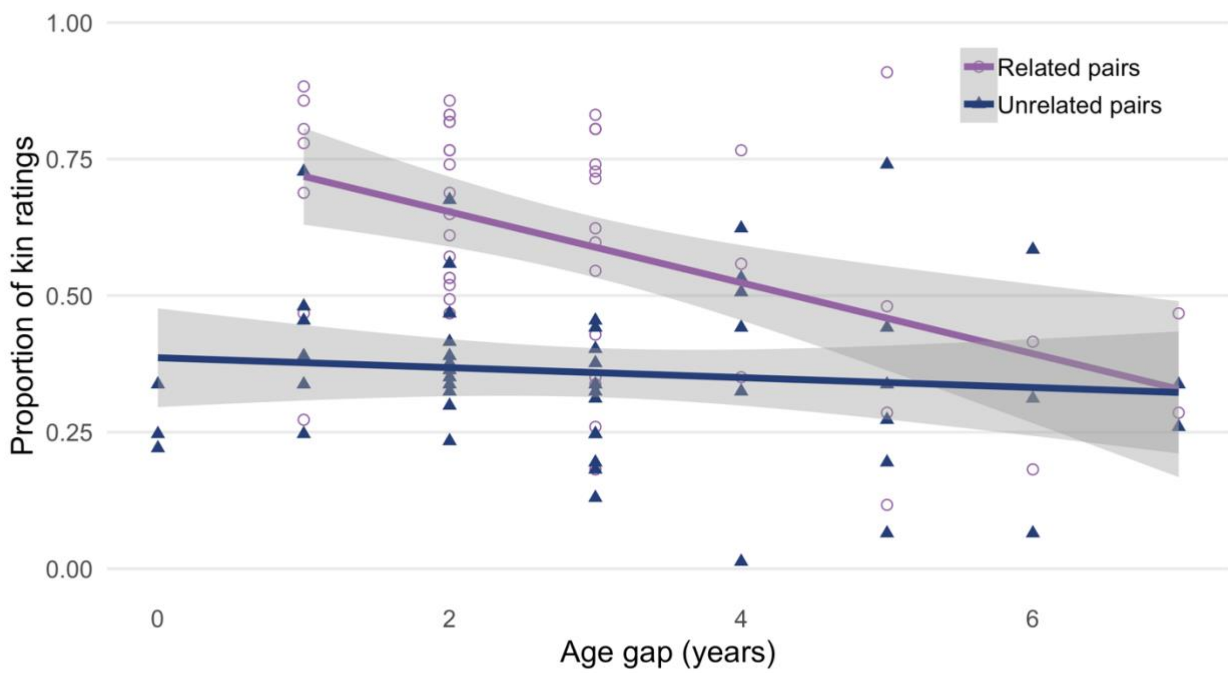
There was no main effect of facial expression (*ß*=0.04, *SE*=0.08, *z*=0.56, *p* = 0.573, *odds ratio*=1.04), but there was an interaction between facial expression and relatedness (*ß*=-0.42, *SE*=0.15, *z*=-2.69, *p* = 0.007, *odds ratio*= 0.66), whereby smiling related pairs were judged to be related less often than neutral pairs, while smiling unrelated pairs were judge to be related more often than neutral unrelated pairs (See Figure 6). This shows that a smiling facial expression decreases kinship judgment accuracy, contradicting our initial hypothesis.

Figure 6. The interaction between relatedness and facial expression on proportion of face pairs judged as related.



We ran an exploratory (not pre-registered) analysis looking at possible effects of age gap within a pair on relatedness judgments. We repeated the analysis above, adding age gap as an additional factor. We found a significant interaction between age gap and relatedness (*ß*=-0.25, *SE*=0.11, *z=-2.21*, p = 0.027, *odds ratio=0.*78), whereby related pairs with a bigger age gap were less likely to be judged as related (See Figure 7). This analysis showed the same significant interaction between relatedness and expression as the pre-registered analysis (*ß*=-0.40, *SE*=0.15, *z=-2.69*, p = 0.007, *odds ratio=0.*67).

Figure 7. The interaction between relatedness and the age gap within a pair on proportion of face pairs judged as related.



3.6. Discussion

In summary, we found that raters are able to discriminate siblings from unrelated pairs with some accuracy, which is consistent with previous literature (Alvergne et al., 2014; Bressan & Dal Martello, 2002; Bressan & Grassi, 2004; Dal Martello et al., 2015; DeBruine et al., 2009; Maloney & Dal Martello, 2006). We also found that facial expression does influence kinship judgments, whereby a smiling facial expression decreases kinship judgment accuracy. In other words, related pairs were judged to be related less often when smiling than with a neutral expression, and unrelated pairs were judged to be related more often when smiling than with a neutral expression.

Marsh, Elfenbein and Ambady's (2003, 2007) found that a smiling facial expression aids the identification of an individual's nationality, yet, in the current study smiles do not aid the identification of kinship but hinder it. This could also suggest that smiles are partially a cue of common cultural background, but not a contextual cue of kinship, at least not in non-adult individuals, or that smiles as a possible cultural mannerism mask kinship cues.

One possible explanation for the finding that a smiling facial expression hinders kin recognition accuracy could be that conflicting mechanisms are employed when faces are processed for emotions and kinship. Previous research finds that the upper half of the face carries more kinship information than the lower half of the face (Dal Martello & Maloney, 2006). However, when processing facial expressions of emotions observers focus on the mouth region (Schyns, Bonnar, & Gosselin, 2002). Consequently, it is possible that an emotional expression could act as a distraction from processing facial cues that are informative of relatedness, as the mouth area is being attended to in order to process these expressions. Moreover, the lower part of the face undergoes radical shape and bone structure changes from childhood to adulthood (Kohn, 1991) and might therefore not be a reliable kinship cue, at least not in children. The current study used photographs of children who were between the ages of 3 and 17 years, hence a widely varying age group in terms of development of lower face characteristics. We found that a bigger age gap reduced the likelihood of related pairs to be judged as related. This could mean that kinship is harder to detect in related pairs when siblings are at different stages of facial development. The previously mentioned research finding that the upper half of the face carries more kinship information than the lower half of the face was also conducted with non-adult stimuli (Dal Martello & Maloney, 2006). Consequently, we cannot readily conclude that facial expression in all cases decreases kinship judgment accuracy, as our results might be specific to kin with non-adult facial characteristics.

Moreover, it is unclear how dynamic spontaneous smiles as observed during natural interactions would impact kin recognition. The stimuli in the current study are static and the individuals in the photographs were instructed to smile, rather than capturing a natural smile. Hence, we cannot determine if dynamic spontaneous smiles have the same effect on kinship detection accuracy as static smiles. Idiosyncrasies in emotional expressions, which might be heritable, may be more clearly observable in dynamic spontaneous smiles.

The current study provides further insight into our understanding and the nature of kinship detection. Our findings show that observers can identify sibling pairs and unrelated pairs at levels above chance, which is in line with previous research. Yet, a bigger age gap within related pairs reduced the likelihood of siblings being judged as related. Moreover, a smiling facial expression decreased the accuracy of judging relatedness compared to a neutral facial expression. This finding could be explained in light of previous research showing that the upper half of the face holds more information about relatedness than the lower half. The study would benefit from being replicated with adult stimuli, to account for facial changes due to growth and dynamic stimuli displaying natural smiles.

Chapter 4:

Experiment 3

Birth order does not affect ability to detect kin

Vanessa Fasolt*[1], Iris J Holzleitner[1], Anthony J Lee[2], Kieran J O'Shea[1], Lisa M DeBruine[1]

[1] Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

[2] Division of Psychology, University of Stirling, Stirling, UK

*Corresponding author: Vanessa Fasolt

Published: Fasolt, V., Holzleitner, I. J., Lee, A. J., O'Shea, K. J., & DeBruine, L. M. (2019). Birth Order Does Not Affect Ability to Detect Kin. *Collabra: Psychology*, *5*(1).

4.1. Author Contribution

Contributor Role	Role Definition	Initials
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims.	VF, LD
Methodology	Development or design of methodology; creation of models.	VF, LD
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.	VF, LD, IH, AL
Validation	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.	//
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.	VF, LD
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.	VF
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.	VF, LD, IH, AL, KO
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse.	//
Writing – Original Draft Preparation	Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).	VF
Writing – Review &	Preparation, creation and/or presentation of the published work by those from the original research group,	VF, LD

Editing	specifically critical review, commentary or revision – including pre- or post-publication stages.	
Visualization	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.	VF, LD
Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.	//
Project Administration	Management and coordination responsibility for the research activity planning and execution.	LD, VF
Funding Acquisition	Acquisition of the financial support for the project leading to this publication.	LD

4.2. Abstract

Previous studies suggest that birth order affects kinship detection ability. Kaminski et al. (2010) argued that firstborns use contextual cues (e.g. maternal perinatal association) to assess kinship in their own family, leading to a disadvantage in assessing kinship from facial cues alone in strangers. In contrast, laterborns do not have the contextual cue of maternal perinatal association and hence rely more on facial cues, leading to an advantage in detecting kin from facial cues alone. However, Alvergne et al. (2010) found no evidence in support of such a birthorder effect. The current study aimed to replicate previous studies with better suited methods to determine the effect of birth order on kin recognition. 109 raters viewed 132 pairs of photographs of children (aged 3-17 years), and indicated whether each pair was related or unrelated. Half of the pairs were sibling pairs and half were unrelated child pairs that were age- and gender- matched to the related pairs. No image was shown more than once, related pairs were not known to be related to any other image in the study, and individuals from unrelated pairs were not known to be related to any other image. We used binomial logistic mixed effects modelling to predict kinship judgments from relatedness and birth order (with image pair and rater as random factors). Relatedness was the main factor driving kinship judgments; related child-pairs were more than twice as likely as unrelated pairs to be judged as kin. Kinship judgment accuracy was unaffected by rater birth order. These findings indicate that laterborns did not have an advantage in detecting child sibling pairs. Pre-registration, data, code, and preprint available at osf.io/h43ep

4.3. Introduction

Kinship is crucial to biological theories of social behaviour, as kinship influences altruistic and reproductive behaviour. Inclusive fitness theory argues that pro-social behaviour is increased towards kin (Hamilton, 1964). Sexual interest, however, is decreased towards close kin to achieve optimal outbreeding (Bateson, 1983), as mating with close kin can result in increased risk of autosomal recessive genetic disorders and miscarriages (Bittles, 2001).

But how do we recognise our kin in the first place? Two main categories of cues, namely phenotypic cues such as vocal, facial and odour resemblance and contextual cues such as maternal perinatal association (intensive maternal care of a sibling after their birth) and co-residence are involved in kin recognition (reviewed in Penn & Frommen, 2010). Maternal perinatal association (Lieberman, Tooby, & Cosmides, 2007) and co-residence (Lieberman, Tooby, & Cosmides, 2003) are correlated with increased pro-social behaviour and increased incest avoidance towards that sibling. Facial resemblance has been reported to influence behaviour in similar ways (see DeBruine, Jones, Little, & Perrett, 2008 for a review), as increased facial self-resemblance increased contributions and trust in economic games (DeBruine, 2002; Krupp, DeBruine, & Barclay, 2008), and self-resembling same-sex faces were found to be more trustworthy and attractive (DeBruine, 2004, 2005). Yet, in line with incest avoidance, facial self-resemblance had a negative effect on attractiveness perceptions of opposite-sex faces in a short-term relationship context, where sexual appeal is the main incentive (DeBruine, 2005). This effect was bigger for women with brothers (especially younger brothers) than women without brothers, with an increasing number of brothers decreasing the perceived attractiveness of unknown self-resembling male faces (DeBruine et al., 2011). Perceptions of trustworthiness were, however, independent of the woman having brothers or not. This suggests that contextual cues, especially maternal perinatal association, are influential cues shaping sexual and pro-social behaviour throughout life (Lieberman et al., 2003, 2007).

Moreover, detecting kinship is not confined to one’s own kin. People are also reliably able to detect kinship among others, which, to a certain extent, enables us to expect and modify behaviours accordingly. Research on third party kin recognition focuses on the physical information that contributes to accurate kinship detection. For instance, Maloney and Dal Martello (2006) found that perceived facial similarity serves almost exclusively as a cue to kinship. Furthermore, studies have shown that people mostly rely on facial features situated in the upper half of the face when making kinship judgments (Alvergne, Perreau, Mazur, Mueller, & Raymond, 2014; Dal Martello & Maloney, 2006). While Dal Martello and Maloney (2006) found that the lower face was still used for kinship judgments when assessing parent-child pairs, Alvergne et al. (2014) found that the lower part of the face did not contain any paternity cues specifically. Moreover, they found that the presence of specific useful information is more important than the number of cues provided and that paternity can be detected even when the features of the face are mixed up (Alvergne et al., 2014). This notion of spatially localised cues being more informative of kinship than holistic cues is supported by Dal Martello, DeBruine and Maloney's (2015) study showing that facial inversion does not affect kinship judgments. Additionally, Dal Martello and Maloney (2010) found that both the left hemi-face and the right hemi-face inform kinship judgments equally, and importantly, that information from the left and right hemi-face is redundant, meaning that given one, no additional kinship information is available from seeing a full face.

However, less research has looked at individual differences in the accuracy with which kinship is detected. Kinship detection accuracy is consistent across cultures, with participants showing no difference in the ability to identify parent-child pairs from their own or another ethnicity (Alvergne et al., 2009; Kaminski, Ravary, Graff, & Gentaz, 2010). Even 5-year-olds can accurately detect parent-child pairs, with no difference between child and adult performances for neonate comparison trials (Kaminski, Gentaz, & Mazens, 2012). At the age of 9, children are also able to distinguish between relevant and irrelevant facial features for kinship detection, i.e. same head orientation or open/closed mouth or eyes (Kaminski, Berger, Jolly, & Mazens, 2013).

A couple of studies found that men gave generally higher similarity ratings than women, but accuracy did not differ between the sexes (Bressan & Dal Martello, 2002; Bressan & Grassi, 2004). An early study by Nesse, Silverman and Bortz (1990) found that men are better at judging relatedness of sons than daughters, and women are better at judging the relatedness of daughters than sons in a similarity task. Nesse, Silverman and Bortz (1990) also looked at the number of children and the number of siblings participants had, but did not find an effect of these factors on accuracy levels. It is important to note that these studies relied on similarity ratings rather than direct kinship judgments, which are highly overlapping (Maloney & Dal Martello, 2006) but not necessarily synonymous (DeBruine et al., 2009).

Bressan and Dal Pos (2012) found that fathers report higher facial resemblance between unfamiliar face pairs than non-fathers, mothers and non-mothers, but that fathers are not more accurate at detecting relatedness than others. This suggests that facial resemblance perception could be biased in fathers, possibly to reinforce paternity beliefs and hence guarantee investment in offspring.

Kaminski, Ravary, Graff and Gentaz (2010) also found a difference in raters' ability to detect kin. In a series of experiments they asked participants to match parent-child pairs, or judge the relatedness of face pairs of varying degrees of relatedness, and found that laterborns outperform firstborns in kinship detection accuracy in both tasks. They found this effect in participants from Taiwan and France, with Taiwanese raters accurately matching Caucasian parent-child pairs, and in child and adult raters. Kaminski et al. (2010) argue that firstborns use facial cues combined with contextual cues (e.g., maternal perinatal association) to assess kinship in their own family, leading to a disadvantage in assessing kinship from facial cues only in unknown faces. In contrast, laterborns do not have the contextual cue of maternal perinatal association and hence rely more on facial cues, leading to an advantage in detecting kin from facial cues alone. However, Alvergne et al. (2010) used a near-identical experimental paradigm and did not replicate this effect of birth order when raters had to determine parent-child pairs.

In light of the above, this study aimed to clarify the role of birth order on kinship detection accuracy. In an attempt to clarify the effect of having older and younger siblings as a child on kin detection, we showed raters stimuli which consisted exclusively of unknown child siblings, as this is arguably the category of kin firstborns and laterborns use differing kin detection strategies on when growing up. In addition, we used colour pictures instead of black and white pictures, and masked images to exclude hair, ears and background to focus on facial cues alone. This avoids variation in global characteristics of the photos, such as posture, as it has been shown that such global characteristics can influence kinship recognition (Kaminski et al., 2013; Vokey, Rendall, Tangen, Parr, & Waal, 2004). Another reason for masking images was to ensure kinship judgments would be exclusively based on facial cues, rather than extraneous kinship information such as a shared hair style. Furthermore, we used a guessing rather than a matching paradigm, which means that raters saw one pair of faces for each trial, rather than a target face and multiple potential matches. This ensured that the relatedness judgments for each pair were based on a given pairs' similarity, rather than being based on comparing a number of possible matches for similarity. Moreover, the guessing task explicitly asked raters whether they thought a pair was related or not, while a matching task implies that there must be a related pair within the set of presented faces. A number of previous studies have used the same methodology as presented in this paper (Dal Martello et al., 2015; Dal Martello & Maloney, 2006; DeBruine et al., 2009; Kaminski et al., 2010; Maloney & Dal Martello, 2006; Nesse et al., 1990). Lastly, we used a binomial logistic mixed model in our analysis to predict relatedness judgments from stimulus pairs' actual relatedness, raters' birth order, and their interaction. We included rater ID and Stimulus ID as random effects rather than fixed effects. This allowed us to account for variation among both raters and stimuli without having to aggregate over one of these groups, which can limit the generalisability of findings beyond the used stimulus/rater set and may inflate false-positive rates. This also means that our dependent variable (DV) was coded as related/unrelated choice, rather than an accuracy score aggregated over stimuli as in Kaminski et al. (2010). Consequently, the effect of interest in the current study is the interacting effect of birth order and actual relatedness on raters' kinship choices rather than a main effect of birth order on overall kinship choices independent of actual relatedness.

Following Kaminski et al. (2010), we hypothesized that
1) There would be a main effect of relatedness, whereby related pairs would be judged as related more often than unrelated pairs.
2) There would be a two-way interaction of relatedness and birth order, whereby the accuracy of relatedness judgments would be higher for laterborns than firstborns.

4.4. Methods

The methods for this study were pre-registered on the Open Science Framework. The analysis script and final data set, as well as details about hypotheses, stimuli (including examples), procedure, and exclusion criteria are all available at osf.io/h43ep.

4.4.1. Stimuli

Stimuli were collected from children visiting a local science centre 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 online studies of family resemblance detection (an example consent form can be found on the OSF).

Children were photographed with a neutral expression looking straight at the camera with hair pulled back and any glasses, scarves, and hats removed. The specific procedures for image collection are available at osf.io/6g7ze.

From a set of approximately 2000 images of individuals of varying ages, sex and relatedness, we algorithmically chose the maximum number of sibling pairs fitting a number of criteria. Both siblings were required to be genetically related and non-twin full siblings under the age of 18. We also required that an age-matched (within 1 year), ethnicity-matched, and sex-matched foil image was 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 are not known to be related to any other individual in the set, while all individuals in unrelated pairs, too, are not known to be related to any individuals in the set. We also required that the algorithm returned an equal number of brother pairs, sister pairs and brother-sister pairs.

This produced 66 sibling pairs and 66 matched unrelated pairs. In each group, 22 pairs were both male, 22 pairs were both female, and 22 pairs were male and female. The individuals' age ranged from 3 to 17 years (mean age = 9.51 years, SD = 2.89 years) and the age difference between individuals in a pair ranged from 0 to 7 years (mean = 2.7 years, SD = 1.56 years). All included children pairs were white.

4.4.2. Procedure

Recruitment of participants was done online through social media (e.g., Facebook) and social bookmarking sites. The study itself was completed by participants online at faceresearch.org using their own computer. Participants were not compensated for their participation, apart from Psychology first-year students at the University of Glasgow, who were offered participation credits for their time.

Participants were told that they would view 132 pairs of faces, some of which were siblings and some of which were unrelated. They were informed that they were to judge whether the pairs were in fact "related" or "unrelated", and that subsequently there would be a short questionnaire about their own family composition (e.g.how many siblings they have). For the actual experimental task, they were shown one pair of child faces at a time and chose their answer by clicking on buttons labelled "unrelated" or "related".

In the questionnaire, participants were asked to indicate how many full siblings they had (from the same mother and father as the participant). The answer was chosen from a drop-down menu ranging from 0 (no siblings) to 10 (10 or more siblings). Participants also provided further information on each of their siblings (e.g., the number of younger/older/same-aged brothers or sisters they have). Information about other types of siblings such as half siblings, adopted siblings and stepsiblings was also gathered but not analysed in this study.

4.4.3. Raters

The study was started by 288 people. Participants who did not rate all 132 stimuli (n = 60) or did not complete the questionnaire (n = 18) were excluded from analyses. After these initial exclusions, we followed a categorisation of raters implemented in Kaminski et al. (2010), i.e. we only included raters with a maximum of two full siblings. This left us with a pool of 109 raters that fit the categorisation criteria and completed all tasks. Raters with one or two younger siblings were categorised as firstborns, while raters with one or two older siblings were categorised as laterborns. Raters with both one younger and one older sibling were also categorised as laterborns.

A power calculation during pre-registration indicated that with 100 participants (50 firstborn/50 laterborn), we would have 93% power to detect an interaction between birth order and relatedness with estimate ≅ 0.27 (odds ratio ≅ 1.3) at 5% alpha. We overshot this recruitment target and included all 109 eligible raters in the main analysis. The analysis and results based on the 100 pre-registered participants can be found in the supplemental materials. There are no differences in results between the two analyses. The laterborns group was made up of 48 raters with only older siblings and 11 raters with an older and a younger sibling. Firstborns (n= 50) only had younger siblings.

In more detail, we excluded

- Participants who did not complete the sibling questionnaire;
- Participants who had more than two full siblings (Kaminski et al., 2010);
- Participants who had “non-full” siblings (e.g. half-, step-, and adopted siblings) – to ensure that the participants were not exposed to maternal perinatal association (intensive

maternal care of a sibling after their birth) through “non-full” siblings, or lived with a “non-full” sibling;

- “Only” children - as they do not have siblings and we were interested in the influence of siblings on kinship judgment accuracy;
- Twins who did not have any other “full” (younger or older) siblings – as birth order in twins is not related to observation of maternal perinatal association.

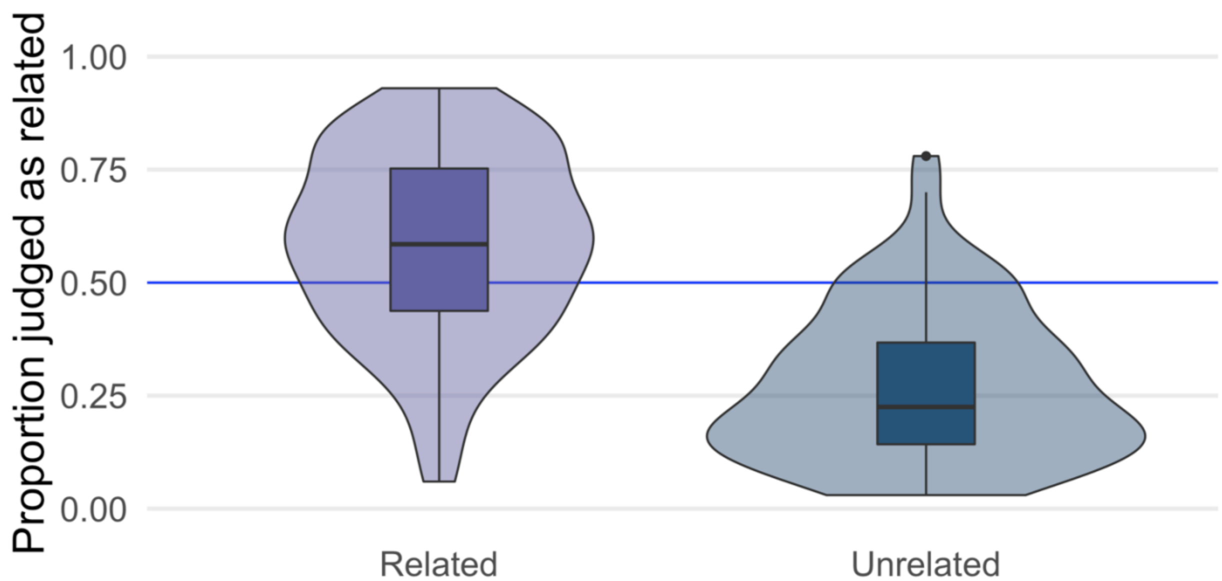
After filtering, the responses from 25 laterborn men (mean age = 29.22 years; SD = 12.7 years), 33 laterborn women (mean age = 25.8 years; SD = 9.96 years) and 1 laterborn of unspecified gender (age = 23 years) were analysed along with 18 firstborn men (mean age = 26.33 years; SD = 4.24 years), 31 firstborn women (mean age = 30.25 years; SD = 14.81 years) and 1 firstborn of unspecified gender (age = 17.1 years). Raters were predominantly white (89 out of 109 raters). Data from the excluded raters can be found in the data file used for the analysis, with the exclusion criteria being clearly marked in the analysis code (both available at osf.io/h43ep).

4.5. Results

We used a binomial logistic mixed model to predict relatedness judgments from actual relatedness (effect-coded as related = +0.5 and unrelated = -0.5), birth order (effect-coded as firstborns = +0.5 and laterborns = -0.5) and the interaction between birth order and relatedness in the kinship task. 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 1.1.19 (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest version 3.0.1 (Kuznetsova, Brockhoff, & Christensen, 2016).

The analysis revealed a main effect of relatedness (*ß*=1.68, *95% CI* [1.32;2.05], *SE*=0.19, *z*=9.07, *p* < .001, *odds ratio*=5.37, *95% CI* [3.74;7.74]), whereby actual related pairs were 5.37 times more likely to be judged as related than unrelated pairs (see Figure 8).

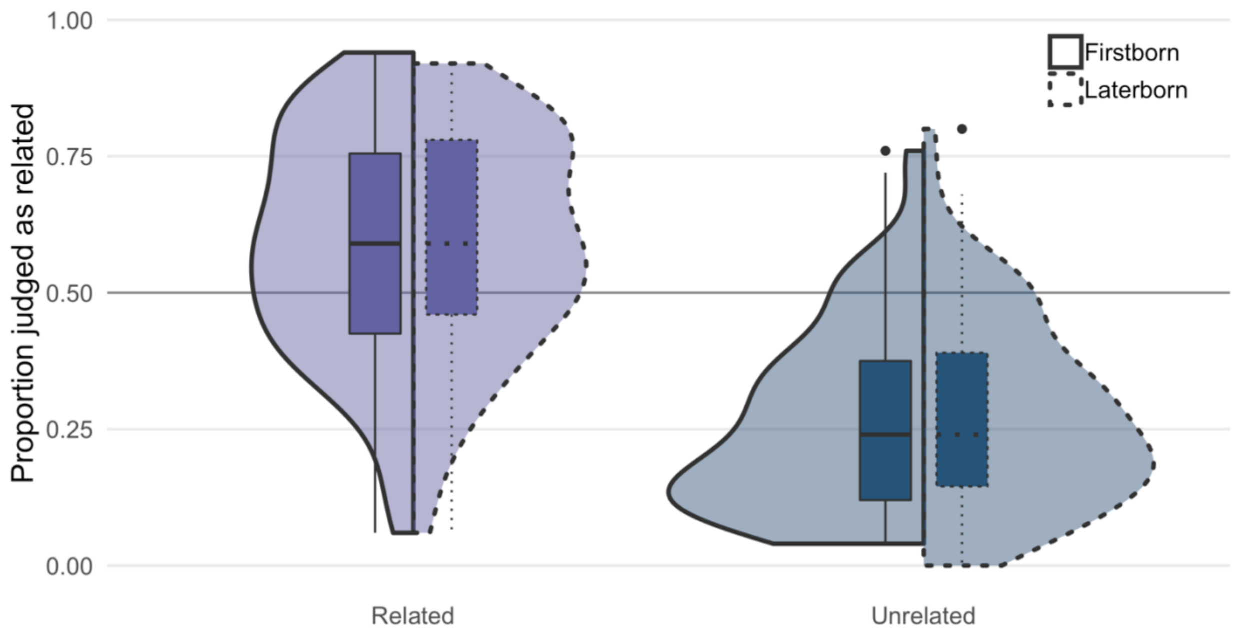
Figure 8. The main effect of relatedness on proportion of face pairs judged as related.



There was no main effect of birth order (*ß*=-0.07, *95% CI* [-0.33;0.19], *SE*=0.13, *z*=-0.54, *p* = 0.59, *odds ratio*=0.93, *95% CI* [0.72;1.2]) and no interaction between birth order and relatedness (*ß*=0.11, *95% CI* [-0.12;0.33], *SE*=0.12, *z*=0.91, *p* = 0.363, *odds ratio*=1.12, *95% CI* [0.89;1.39]), see Figure 9.

In fact, when looking at the non-significant difference between firstborns and laterborns (not pre-registered), firstborns tended to be more accurate in their kinship judgments (*ß*=1.75, *95% CI* [1.35; 2.15], *SE*=0.2, *z*=8.65, *p* < .001, *odds ratio*=5.75, *95% CI* [3.88; 8.57]) than laterborns (*ß*=1.62, *95% CI* [1.25; 1.99], *SE*=0.19, *z*=8.65, *p* < .001, *odds ratio*=5.05, *95% CI* [3.50; 7.30]), opposite to the prediction by Kaminski et al. (2010).

Fig 9. The interaction between relatedness and birth order on proportion of face pairs judged as related.



4.6. Discussion

In summary, we found that raters are able to identify who is related and who is unrelated when shown only facial information of children, with no further context information. This is a robust finding in the literature. We did not find that birth order, namely whether raters were firstborns or laterborns, influenced the accuracy of kinship judgments of children. Our results are consistent with Alvergne et al. (2010) who also found no effect of birth order when matching parents and children.

Our results are inconsistent with the finding by Kaminski et al. (2010) that laterborns have an advantage in detecting parent-child pairs and kin of varying degrees of relatedness. This failure to replicate Kaminski et al.'s (2010) could be a result of using different stimuli. That is, we used exclusively child pairs while Kaminski and colleagues used pairs that differed in their degree of relatedness, with only a subset being siblings, of which the age was unknown. However, as Kaminski et al. (2010) argued that an advantage in kinship detection accuracy is based on birth order (i.e. having different constellations of siblings as a child), identifying child siblings is arguably a better test of this hypothesis. The current study could be repeated with other degrees of relatedness (e.g., parent-neonate pairs, grandparent-grandchild pairs, aunt/uncle-niece/nephew pairs etc.) to see whether this advantage in detecting kin is in fact limited to other kin constellations, which in turn could mean that the explanation as to why there is an advantage based on birth order has not fully been understood yet. Moreover, Kaminski et al.'s (2010) definition of laterborns included individuals who had both an older and a younger sibling, hence the laterborn might have witnessed maternal perinatal association with a younger sibling. In our data set, 11 of the 59 "laterborns" had both an older and younger sibling. This could mean that we are simply not picking up the effect of birth order due to categorisation issues. To investigate this, we conducted an exploratory analysis in which we only included laterborns with one or two older siblings, excluding raters with both an older and a younger sibling. This exclusion criterion resulted in 48 laterborns with only older siblings. Re-analysis did not change the results: birth order still had no main effect on kinship judgment accuracy (*ß*=-0.06, *SE*=0.14, *z*=-0.44, p = 0.659, *odds ratio*=0.94) and there was no significant interaction between birth order and actual relatedness (*ß*=0.12, *SE*=0.12, *z*=0.98, p = 0.327, *odds ratio*=1.13). To conclude, we find that raters are able to identify related and unrelated pairs of children, a finding consistent with the majority of research on third party kin recognition. We did not find that birth order of the rater, namely being a firstborn or a laterborn, influences kinship judgment accuracy when judging these pairs of children, which is in line with Alvergne et al. (2010) and inconsistent with Kaminski et al. (2010), who found that laterborns have an advantage when identifying kin of different degrees of relatedness.

Chapter 5:

Experiment 4

Methods comparison in third party kin recognition; or how everyone finds a different answer to the same question.

Vanessa Fasolt*[1], Iris J Holzleitner[1], Anthony J Lee[2], Kieran J O'Shea[1], Lisa M DeBruine[1]

[1] Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

[2] Division of Psychology, University of Stirling, Stirling, UK

*Corresponding author: Vanessa Fasolt

5.1. Author Contribution

Contributor Role	Role Definition	Initials
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims.	VF, LD
Methodology	Development or design of methodology; creation of models.	VF, LD
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.	VF, LD, IH, AL
Validation	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.	//
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.	VF, LD
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.	VF
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.	VF, LD, IH, AL, KO
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse.	//
Writing – Original Draft Preparation	Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).	VF
Writing – Review &	Preparation, creation and/or presentation of the published work by those from the original research group,	VF, LD

Editing	specifically critical review, commentary or revision – including pre- or post-publication stages.	
Visualization	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.	VF, LD
Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.	//
Project Administration	Management and coordination responsibility for the research activity planning and execution.	LD, VF
Funding Acquisition	Acquisition of the financial support for the project leading to this publication.	LD

5.2. Abstract

Research on third party kin recognition has consistently found that humans can reliably judge relatedness among strangers when presented with face photographs alone. However, contrasting results have been found when looking at the effect of sex and age of the portrayed individuals on kinship judgments. This discrepancy could partially be due to the use of different methods. To explore this issue, we conducted a study implementing three commonly used methods (i.e., kinship judgment, similarity rating, matching paradigm) and directly compared the performance of participants across these methods, using the same highly-controlled stimulus set. We found that while responses on all three tasks were correlated, performance varied significantly across the tasks. Participants in the kinship judgment task were most accurate at detecting unrelated pairs, participants in the matching task were most accurate at detecting related pairs, and participants in the similarity rating task were equally good at detecting related and unrelated pairs. Furthermore, when looking at the effect sex and age of the portrayed individuals had on performance, we found that stimuli sex only had a main effect in the kinship judgment paradigm. Raters judged same-sex pairs to be related more often than opposite-sex pairs, independent of actual relatedness. In the matching task, there was an interaction between stimuli sex and stimuli age, where a larger age difference between stimuli decreased relatedness judgments for same-sex pairs, but marginally increased relatedness judgments for opposite-sex pairs. Our results suggest that different answers to the same question can be found, depending on which method is used. This highlights the need for standardised methods in the field to allow for generalizable conclusions. Pre-registration, data and code are available on the Open Science Framework osf.io/a3t8x/.

5.3. Introduction

Kinship crucially influences social behaviour by increasing pro-social behaviours towards kin (for a review, see Bressan & Kramer, 2015). One way of identifying kin is through phenotype matching (for a review, see Penn & Frommen, 2010), namely comparing facial characteristics between oneself and a potential family member, or in the case of allocentric kin recognition, comparing facial characteristics between strangers to judge relatedness of any possible constellation of people.

Research on allocentric kin recognition has shown that individuals are able to detect kin among strangers from facial characteristics alone (e.g., Alvergne, Faurie, & Raymond, 2007; Alvergne et al., 2009; Alvergne, Perreau, Mazur, Mueller, & Raymond, 2014; Bressan & Dal Martello, 2002; Bressan & Grassi, 2004; Brédart & French, 1999; Dal Martello, DeBruine, & Maloney, 2015; Dal Martello & Maloney, 2006; Dal Martello & Maloney, 2010a; DeBruine et al., 2009; French, Brédart, Huart, & Labiouse, 2000; Kaminski, Méary, Mermillod, & Gentaz, 2010; Maloney & Dal Martello, 2006; Nesse, Silverman, & Bortz, 1990). Yet, this appears to be the only robust finding in the literature, as different studies have reported contradicting effects of variables such as sex of stimuli and age of stimuli on kin judgements.

For example, sex of stimuli has been suggested to be an important factor in detecting parent-child pairs. On one hand, it has been hypothesized that negative effects of paternity uncertainty on care behavior have led to an increased facial resemblance of children to their father (Platek et al., 2003, 2004; Volk & Quinsey, 2007). However, some research has found no evidence for a difference in men and women's preference for and investment in self-resembling child faces (DeBruine, 2004), or has even reported the opposite pattern, with a greater preference for self-resembling children in women (Bressan, Bertamini, Nalli, & Zanutto, 2009). On the other hand, it could also be the case that paternal resemblance could be disadvantageous and costly for the child when infidelity occurs (Daly & Wilson, 1996; French et al., 2000). Studies based on interviews with relatives and observations of family interactions with newborns found that a belief of child-father resemblance is primarily established and nurtured by relatives commenting on such a resemblance rather than an actual strong phenotypic resemblance (Alvergne et al., 2007; Daly & Wilson, 1982; McLain, Setters, Moulton, & Pratt, 2000; Regalski & Gaulin, 1993).

Similarly, the results from studies investigating the effect of sex of children and their parents are contradictory. One study has found that one-year-old children resemble their fathers more than their mothers (Christenfeld & Hill, 1995), but this finding has never been replicated (Brédart & French, 1999; French et al., 2000). One study found that children resemble their mothers more than their fathers (McLain et al., 2000), while two other studies found that boys bear greater resemblance to their fathers and girls to their mothers (Alvergne et al., 2009; Kaminski et al., 2010). Yet another study found that children look most similar to women in general, with girls resembling mothers and other women more than their fathers or other men, and boys resembling both parents equally, but unrelated women more so than men (Bressan & Dal Martello, 2002). Only one study has looked at sibling- as opposed to parent-child pairs. DeBruine et al. (2009) found that unrelated same-sex pairs received higher similarity ratings than unrelated opposite-sex pairs, while sex composition had no effect on similarity ratings of actual sibling pairs. This suggests that the role of sexually dimorphic facial cues on kinship judgments and similarity judgments is not fully understood yet.

Age of the stimuli has also been suggested to influence kin recognition, but again findings have been contradictory. Some studies found that age does not affect kin recognition (Kaminski et al., 2009; Maloney & Dal Martello, 2006; Nesse et al., 1990). In contrast, Alvergne, Faurie and Raymond (2007) found that newborn boys resemble their mothers more than their fathers, but between the ages of two and three years an inversion occurs, and boys start to resemble their fathers more so than their mothers. For girls, this inversion does not occur; they resemble their mothers more than their fathers at any age. Brédart and French (1999) found that raters were better at matching five-year-old boys to their parents than younger boys, while there was no such age effect for girls. Furthermore, Christenfeld and Hill (1995) found that one-year old children resemble their fathers more than their mothers, with older children not being accurately matched to their parents at all. For siblings, DeBruine et al. (2009) age difference in opposite -sex pairs was larger than that in same-sex pairs, i.e. age and sex composition of stimuli pairs were confounded.

Considering the above, a crucial question is how could these numerous studies have found such different results? The answer, in part, might be to do with the use of different methods across studies. One significant difference between studies can be found in the quality of the stimuli; very few used standardized image sets. Some image sets consisted of photographs that had been sent in by relatives (e.g., Brédart & French, 1999; Bressan & Dal Pos, 2012; Christenfeld & Hill, 1995; Kaminski et al., 2009), which could be problematic as picture selection might have been biased (e.g., pictures could have been chosen precisely because of their unusually high resemblance). Moreover, quality and properties of the pictures themselves may vary, allowing factors other than resemblance to influence kinship judgements. Numerous studies did use pictures taken by researchers (e.g., Dal Martello & Maloney, 2006; Dal Martello & Maloney, 2010b; DeBruine et al., 2009; Maloney & Dal Martello, 2006; Nesse et al., 1990; Oda, Matsumoto-Oda, & Kurashima, 2002), although standardization did still vary between sets, and some studies used a mix of photographs which were sent in by families and also taken by researchers (Bressan & Grassi, 2004). Another difference between studies is that some used black and white photographs, while others used colour photographs, which could lead to different results as Kaminski et al. (2010) found that kinship judgments were more accurate with black and white photographs. Also, aspect ratio of stimulus images is often not controlled for, but can crucially bias raters' perceptions (Vokey, Rendall, Tangen, Parr, & Waal, 2004).

Another difference between studies is the exact experimental paradigm they used. All studies cited above used a variation of three methods, namely 1) asking how similar a pair is, 2) asking whether a pair is related or not or 3) asking participants to match a target face to a real relative out of a set of possible matches. While it is assumed that these tasks measure the same construct, this is not necessarily the case. For instance, similarity studies are not explicitly asking questions about kinship - while it is clear that facial similarity informs kinship judgments (Maloney & Dal Martello, 2006), they may not be necessarily synonymous (DeBruine et al., 2009). Also, studies using the matching task used varying numbers of stimuli; for example, Arantes and Berg (2012) displayed two possible matches per target face, while Bressan and Dal Pos (2012) displayed twelve possible matches per target face. This might crucially influence kinship judgments as more reference information is available the more faces are displayed.

In light of the above, we conducted a study directly investigating the effect of study method and the effect of sex and age of the stimuli with a large set of standardised stimuli. We tested several hypotheses that were pre-registered on the Open Science Framework osf.io/p3btx, some being confirmatory after conducting a pilot study and others being exploratory as no clear predictions emerged from the existing literature.

1) Confirmatory Hypotheses

Based on the results of a pilot study, we proposed the following hypotheses:

1a) Findings from all three methodologies are correlated with each other.

1b) Participants will perform more accurately in the matching task than the similarity rating and relatedness judgment tasks, due to being able to directly compare related and unrelated options.

2) Exploratory Hypotheses

Tentative hypotheses regarding the effect of sex and age are based on previous research (DeBruine et al., 2009) that found that sex and age differences between stimuli influence similarity ratings (pairs with a sex difference or a bigger age difference were rated as less similar than same-sex or more similarly aged pairs), but not kinship judgments. This would suggest that similarity ratings convey information that is not present in kinship judgment.

2a) Sex differences negatively influence accuracy in the similarity rating task, but not in the kinship judgment task or matching task.

2b) Age differences negatively influence accuracy in the similarity rating task, but not in the kinship judgment or matching task.

5.4. Methods

This study was pre-registered on the Open Science Framework at osf.io/p3btx, including hypotheses, methods, and analyses. All procedures and analyses reported in this manuscript follow this pre-registration.

5.4.1. Stimuli

Photographs were taken of families visiting the Glasgow Science Centre as part of a study investigating facial cues of kinship. Both parents and children consented to the use of their pictures in further kinship studies. Children were photographed looking straight at the camera with a neutral facial expression and their hair held back with a headband. Any glasses, scarves, and hats were removed for the picture. The specific procedures for image collection are available at osf.io/bvtnj.

From a set of approximately 2000 images of individuals, we algorithmically chose the maximum number of sibling pairs fitting a number of criteria. Both siblings were required to be genetically related and non-twin full siblings under the age of 18. We also required that an age-matched (within 1 year), ethnicity-matched, and sex-matched foil image was 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. We also required the algorithm to return equal numbers of sister pairs, brother pairs and sister-brother pairs.

This procedure resulted in 66 sibling pairs and 66 matched unrelated pairs, hence 132 pairs in total. In each group, 22 pairs were both male, 22 pairs were both female, and 22 pairs were male and female. The individuals ranged from 3 to 17 years old (mean age = 9.51, SD = 2.89) and the age difference between individuals in a pair ranged from 0 to 7 (mean = 2.7, SD = 1.56).

5.4.2. Procedure

Participants completed the experiment online at faceresearch.org. Raters were recruited through social media and social bookmarking sites.

Participants were randomly assigned to one of three tasks, namely either the kinship judgment task, the similarity rating task or the matching task. Raters were told that they would view 132 pairs of faces in the kinship task and the similarity task, and 66 sets of faces in the matching task. In the kinship task, raters were asked to determine whether each pair was related or not, by clicking a corresponding “related” or “unrelated” button presented above the pairs of faces. In the similarity task, raters were asked to indicate how similar they thought the pair looked on a scale from 0 (not very similar) to 10 (very similar). In the matching task, raters were asked to choose which one of two faces could be the sibling of a target face by clicking on either the left or right child’s face. We created four counterbalanced versions of the matching task, which means that each individual of the related and unrelated pairs was at one point the centre image to help reduce any possible bias.

5.4.3. Raters

The kinship task was started by 82 people, the similarity task by 97 people, and the matching task by 189 people. The number of participant starting the matching task was higher than the starting numbers of the other tasks as the randomization for assigning participants in equal numbers to each task did not work initially. Raters who did not rate all 132 stimuli pairs in the similarity task and the kinship task were excluded from analyses (29 for the kinship task; 47 for the similarity task), and 47 raters who did not complete all 66 trials in the matching task were also excluded from the analyses. Following the procedure specified in the pre-registration, the first 50 raters from each task were selected, meaning that the data of 150 raters was used in the analyses. For the four counterbalanced versions of the matching task this meant specifically that the first 12 or 13 raters from each version were included in the final data set of 50 raters for the matching task.

After filtering, the responses from 35 men, 99 women and 16 raters who did not indicate their sex were included in the analysis. The mean age of the remaining raters was 25.4 years (standard deviation = 10.3 years).

5.5. Results

Analyses were conducted in the programming software R version 3.5.0 (R Core Team, 2017) in conjunction with lme4 version 1.1.19 (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest version 3.0.1 (Kuznetsova, Brockhoff, & Christensen, 2016). All analyses and complete data files can be found at osf.io/a3t8x/.

Testing Correlations Among Studies

Hypothesis 1a) predicted that the ability to detect relatedness would be correlated among all study methods.

In order to test this hypothesis, we conducted three Pearson’s product-moment correlations to correlate two methods at a time and found that all methods were significantly correlated with each other (all $p < .001$), with correlations ranging from R=0.7 between the similarity and matching task, R=0.71 between the kinship and matching task and R=0.93 between the kinship and similarity task. This shows that the ability to detect kin was correlated between tasks, supporting Hypothesis 1.

Testing Differences Between Studies

Hypothesis 1b) predicted that participants would perform more accurately in the matching task than the similarity rating and relatedness judgment tasks.

Thresholding

As both matching and relatedness judgment tasks retuned binary responses, we needed to convert the 0-10 similarity ratings into binary responses for a direct comparison across tasks. This was achieved by first calculating the mean percentage of each 0-10 rating for related and unrelated pairs separately. Then, we calculated the log likelihood ratio by dividing the related by the unrelated mean percentage calculated before. This data was then entered into a linear model to obtain the *beta coefficient* (4.66), which indicated at which point the similarity ratings could be divided into binary ratings. This meant that ratings over 4.66 were recoded as ‘related ratings’ and ratings under 4.66 were recoded as ‘unrelated ratings’ in the subsequent threshold analysis. This thresholding procedure followed Maloney and Dal Martello (2006).

Main Analysis

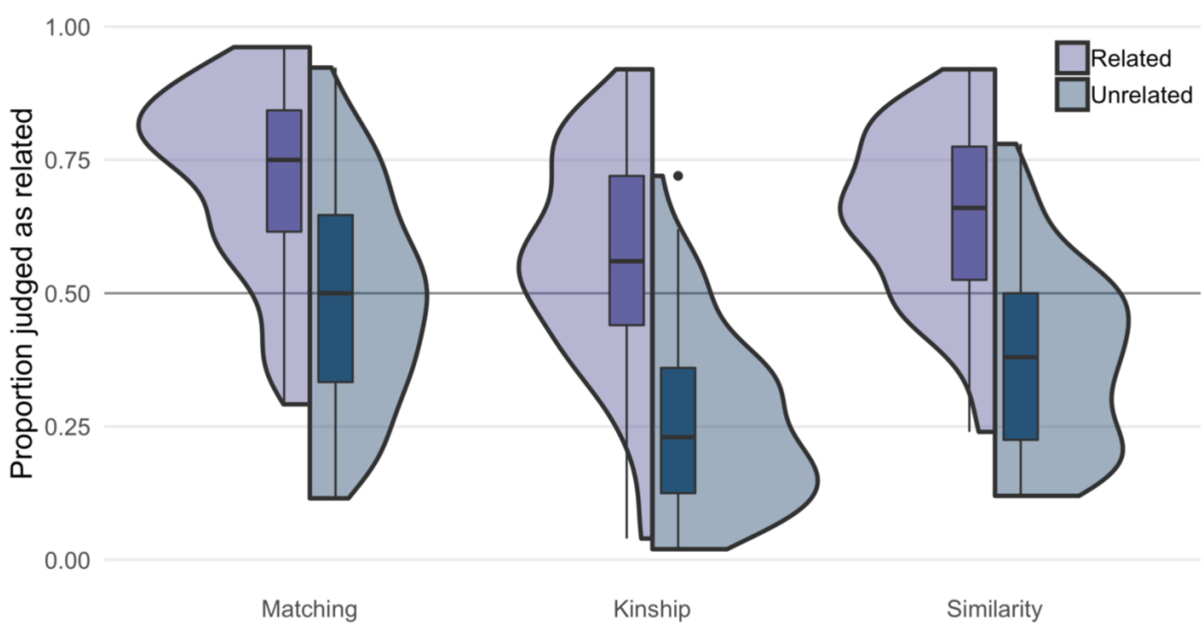
The now binary dependent variable allowed us to directly compare kinship detection between all three tasks. We used a generalized linear mixed model to predict relatedness judgments from actual relatedness (effect-coded as related = +0.5 and unrelated = −0.5), study type (dummy-coded with matching task as reference, kinship task and similarity task) and possible two-way interactions. Rater ID and stimuli ID were entered as random effects and random slopes were specified maximally (Barr, Levy, Scheepers, & Tily, 2013).

Table 8. Results from Main analysis

	Estimate	***SE***	***z.value***	***p-value***	***Odds ratio***
Matching Task	*0.51*	*0.14*	*3.68*	*p < .001*	*1.67*
Relatedness	*1.15*	*0.18*	*6.32*	*p < .001*	*3.16*
Kinship Task	*−1.06*	*0.14*	*−7.79*	*p < .001*	*0.35*
Similarity Task	*−0.43*	*0.15*	*−2.83*	*p = 0.005*	*0.65*
Kinship * Relatedness	*0.48*	*0.18*	*2.63*	*p = 0.009*	*1.62*
Similarity * Relatedness	*0.45*	*0.19*	*2.40*	*p = 0.016*	*1.57*

With the matching task as the reference task in this model, we found that raters' responses in both the similarity task and the kinship task were significantly different from the responses in the matching task, independent of actual relatedness. Moreover, the significant interactions between the similarity task and actual relatedness, and between the kinship task an actual relatedness indicated that kinship detection rates are different in these two tasks compared to the matching task (See Figure 10).

Figure 10. The interaction between relatedness and study task on proportion of face pairs judged as related. Note that for trials showing unrelated pairs, raters in the matching task should show no preference as both possible choices are unrelated.



Individual Study Analyses

Next, we conducted separate analyses for each study method to look more closely at potential interactions between study task and relatedness. We performed three generalized linear mixed models, one for each task, with relatedness as fixed effect (effect-coded as related = +0.5 and unrelated = −0.5), rater ID and stimulus ID as random effects, and maximally specifying our random slopes.

These analyses showed that actual relatedness had a main effect in each task. In the similarity task, related pairs were 5.16 times more likely to be judged related than unrelated pairs ($ß$=1.64, SE=0.18, z=8.88, p<.001), and in the kinship task related pairs were 5.37 times more likely to be judged related than unrelated pairs ($ß$=1.68, SE=0.17, z=9.65, p < .001). In the matching task, related pairs were 2.97 times more likely to be judged related than unrelated pairs ($ß$=1.09, SE=0.16, z=6.73, p<.001). These lower odds were due to unrelated pairs being judged as related at chance level rather than accurately, as there was no accurate choice (See Figure 10).

Indeed, when we looked at task performance including only related pairs (matching task as reference), we found that performance in the kinship task was significantly different from the performance in the matching task ($ß$=−0.98, SE=0.18, z=−5.39, p<.001, *odds ratio*= 0.38), whereby raters in the matching task were significantly better at detecting related pairs than raters in the kinship task. However, there was no significant difference in performance between raters in the matching task and the similarity task ($ß$=−0.3, SE=0.19, z=−1.53, p=0.126, *odds ratio*= 0.74). Raters in the similarity task were also significantly more accurate at identifying related pairs that raters in the kinship task ($ß$=0.68, SE=0.16, z=4.18, p<.001, *odds ratio*= 1.97) (see Figure 1). When looking at unrelated pairs (kinship task as reference as matching task is chance), we found that performance in the similarity task was significantly different from the performance in the kinship task ($ß$=0.54, SE=0.16, z=3.26, p=0.001, *odds ratio*= 1.72), whereby raters in the kinship task were significantly better at detecting unrelated pairs than raters in the similarity task. As expected, performance in the matching task was significantly different from the performance in the kinship task ($ß$=1.16, SE=0.17, z=6.71, p<.001, *odds ratio*= 3.19), as raters could not accurately detect unrelated pairs in the matching task as result of the methodology (See Figure 10).

Exploratory Analysis

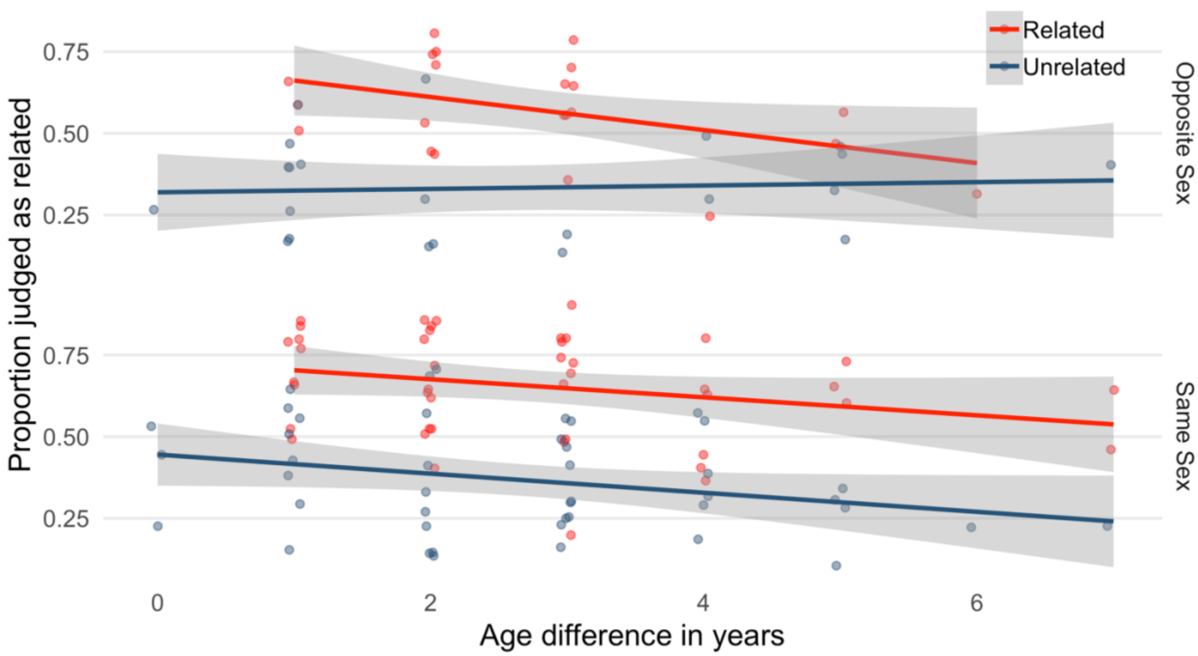
Following the exploratory hypotheses, we wanted to investigate and compare the effects of sex and age on judgments across the different methods used in the literature.

We used a generalized linear mixed model to predict relatedness judgments from actual relatedness (effect-coded as related = +0.5 and unrelated = −0.5), study type (matching task as reference, kinship task and similarity task), sex composition within a pair (effect-coded as same sex = +0.5 and opposite sex = −.05), age difference within a pair (continuous), and any possible interactions. Rater ID and stimuli ID were entered as random effects and random slopes were specified maximally (Barr et al., 2013). The matching task was set as reference study method.

This exploratory analysis revealed the same main effect of relatedness (*ß*=1.06, *SE*=0.08, *z*=12.62, $p < .001$, *odds ratio*= 2.89), kinship task (*ß*=−1.05, *SE*=0.14, *z*=−7.54, $p < .001$, *odds ratio*= 0.35) and similarity task (*ß*=−0.47, *SE*=0.16, *z*=−3.00, $p = 0.003$, *odds ratio*= 0.63) as the main analysis. Moreover, identical to the main analysis, we found an interaction between relatedness and the kinship task (*ß*=0.48, *SE*=0.11, *z*=4.55, $p < .001$, *odds ratio*= 1.62) and an interaction between relatedness and the similarity task (*ß*=0.43, *SE*=0.11, *z*=4.03, $p < .001$, *odds ratio*= 1.54).

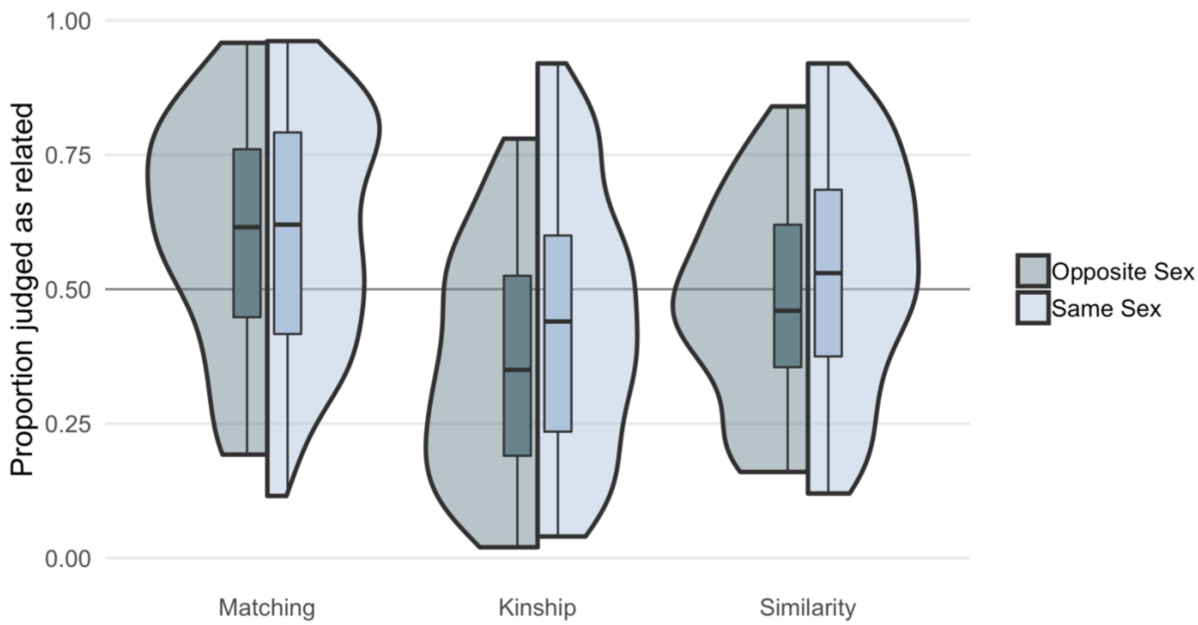
In regards to the age and sex of stimuli, there was a main effect of age difference within pairs (*ß*=−0.16, *SE*=0.05, *z*=−3.37, $p < .001$, *odds ratio*= 0.85), whereby a larger age gap within a pair decreased relatedness judgments (*R*= -0.06, $p < .001$). This main effect of age gap was qualified by an interaction with relatedness (*ß*=−0.19, *SE*=0.06, *z*=−3.30, $p < .001$, *odds ratio*= 0.83), whereby this effect of age gap was bigger for related pairs (*R*=−0.1, $p < .001$) than unrelated pairs (*R*= −0.05, $p < .001$). This two-way interaction was qualified by a three-way interaction between age difference, relatedness and sex composition (*ß*=0.31, *SE*=0.12, *z*=2.63, $p = 0.009$, *odds ratio*= 1.36), whereby an increased age gap decreases relatedness judgments for everyone (all $p < .001$) but unrelated opposite sex pairs (*R*= 0.02, $p = 0.314$) (see Figure 11).

Figure 11. The three-way interaction between age difference, relatedness and sex composition within pairs on proportion of face pairs judged as related.



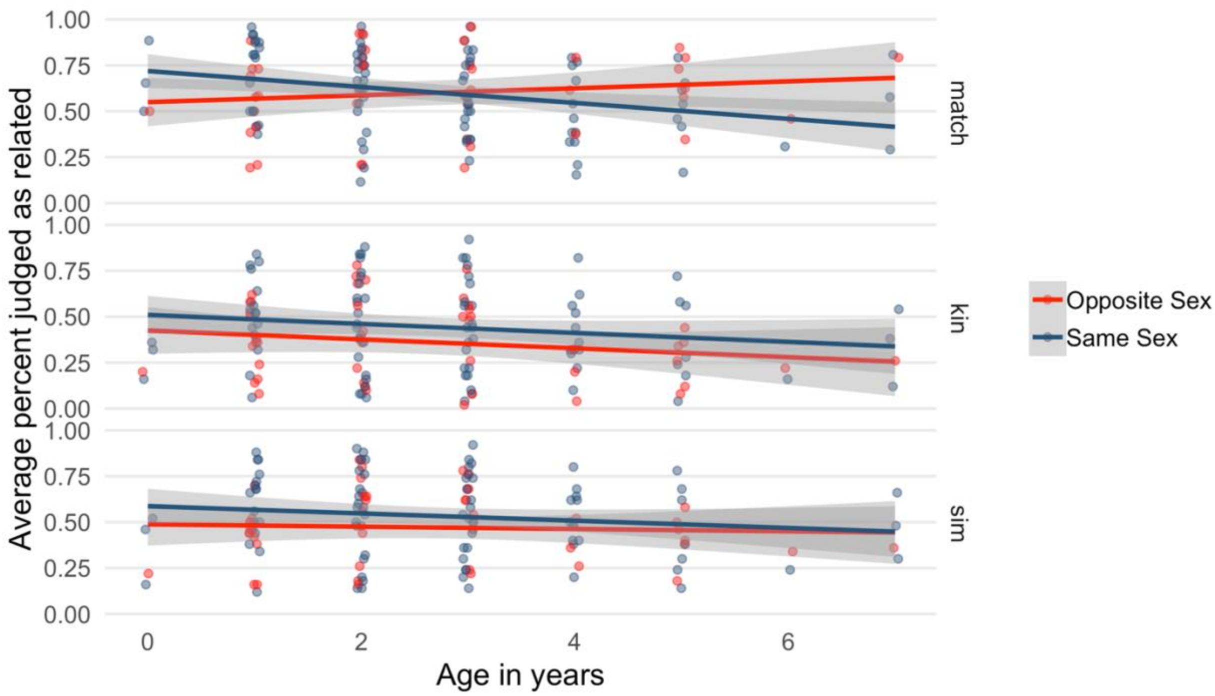
Furthermore, there was a significant interaction between the kinship task and sex composition within a pair (*ß*=0.43, *SE*=0.18, *z*=2.34, *p* = 0.019, *odds ratio*= 1.54). This indicated that the interaction between the kinship task and sex composition was different from an interaction between the reference matching task and sex composition. There was no significant interaction between sex composition and the similarity task (*ß*=0.36, *SE*=0.19, *z*=1.86, *p* = 0.063, *odds ratio*= 1.43), which indicates that sex composition influenced performance in the similarity task and matching task comparably (see Fig. 12).

Figure 12. The interaction between study task and sex composition within pairs on proportion of face pairs judged as related.



The significant two-way interaction between sex composition and the kinship task was further qualified by a significant three-way interaction between the kinship task, sex composition and age difference within a pair (*ß*=0.3, *SE*=0.11, *z*=2.79, *p* = 0.005, *odds ratio*= 1.35). Also, the non-significant two-way interaction between sex composition and the similarity task was further qualified by a significant three-way interaction between the similarity task, sex composition and age difference (*ß*=0.23, *SE*=0.11, *z*=2.09, *p* = 0.037, *odds ratio*= 1.26). This suggests that the sex composition and age difference within a pair had a significantly different effect on the kinship task and the similarity task than on the matching task (see Fig. 13).

Figure 13. The three-way interaction between study task, sex composition and age difference within pairs on proportion of face pairs judged as related.



The exploratory analysis did not find any other significant result (all p > 0.502).

Individual Study Analyses

In light of this, we conducted three additional separate analyses for each task to test the effect sex composition and age difference had on the respective tasks.

We found that sex composition had a main effect on the kinship task (*ß*=0.41, *SE*=0.18, *z*=2.32, *p* = 0.02, *odds ratio*= 1.51), whereby same-sex pairs were judged to be related more often than opposite-sex pairs, independent of actual relatedness. There was no effect of sex composition in the similarity task (*ß*=0.33, *SE*=0.17, *z*=1.94, *p* = 0.052, *odds ratio*= 1.39) or in the matching task (*ß*=0.01, *SE*=0.16, *z*=0.03, *p* = 0.973, *odds ratio*= 1.01).

Age difference did not have a main effect on any of the tasks (kinship task: *ß*=0, *SE*=0.04, *z*=0.04, *p* = 0.966, *odds ratio*= 1.00; similarity task: *ß*=0.01, *SE*=0.04, *z*=0.13, *p* = 0.895, *odds ratio*= 1.01; matching task: *ß*=−0.07, *SE*=0.04, *z*=−1.68, *p* = 0.094, *odds ratio*= 0.93).

There was however, a significant interaction between sex composition and age difference in the matching task (*ß*=−0.21, *SE*=0.08, *z*=−2.49, p = 0.013, *odds ratio*= 0.81), whereby an increased age difference decreased relatedness judgments for same-sex pairs (*R*=−0.13, $p <$ .001), but increased relatedness judgments for opposite-sex pairs (*R*= 0.06, p = 0.051; see Figure 4). This two-way interaction was not significant in the kinship task (*ß*=0.11, *SE*=0.08, *z*=1.4, p = 0.16, *odds ratio*= 1.12) or the similarity task (*ß*=−0.07, *SE*=0.08, *z*=−0.86, p = 0.393, *odds ratio*= 0.93).

5.6. Discussion

In summary, we found that while results from all three methods were correlated, performance varied significantly across the different tasks. In the kinship task, raters were highly accurate in detecting unrelated pairs but performed not much above chance in detecting related pairs; in the similarity task, raters were similarly accurate in detecting related and unrelated pairs, and in the matching task, raters were highly accurate in detecting related pairs. Raters in the matching task and the similarity tasks were equally good at detecting related pairs. Raters in the kinship task were significantly worse at detecting related pairs, yet, significantly the best at detecting unrelated pairs. This suggests that the same related or unrelated pairs will be judged differently depending on the task that is used. Moreover, independent of task, an increased age difference within a pair decreased kinship judgments for all but unrelated opposite-sex pairs, where it had no effect. Furthermore, when looking at the effect of sex difference and age difference on the performance in the different tasks specifically, we found that sex composition had only a main effect in the kinship task. Raters judged same-sex pairs to be related more often than opposite-sex pairs, independent of actual relatedness. In the matching task, there was an interaction between sex difference and age difference, which meant that an increased age difference decreased relatedness judgments for same-sex pairs, but marginally increased relatedness judgments for opposite-sex pairs.

We were able to confirm only one of our hypotheses, namely that performance of raters across tasks is correlated. We did not find that participants performed most accurately in the matching task, but instead found that raters were equally good at detecting related pairs in the matching task and the kinship task. Detection of unrelated pairs was as expected at chance level; both choices were unrelated to the target face, and raters should therefore not have a preference. None of our exploratory hypotheses were supported.

The current study used photographs of individuals between the ages of 3 and 17 years, depicting siblings or matched control pairs. On the one hand, this allowed us to limit the effect of facial development on morphological similarity and kinship, as the individuals in the current study were much closer in age than parent-child pairs would have been. On the other hand, this allowed us to investigate the effect of sex and age without biases introduced by (contradicting) theories about benefits and drawbacks of paternal resemblance. Yet, it will be important to carry out the same study with parent-child pairs to be able to conclude how tasks specifically influence results for that kin type, as it might be different due to cognitive biases.

In conclusion, depending on what method is being used, results can vary widely and different answers to the same question can be found. Despite performance rates being correlated between tasks, performance distributions differed significantly. That is, the specific task used may skew any conclusion on third party kin recognition more generally. Moreover, when looking at more specific effects of sex and age of the portrayed individuals, task choice can again crucially influence the results. This highlights the need of standardised methods in the field to allow for generalizable conclusions. The use of high-quality stimuli is also critical to ensure results are not influenced by features of the photographs rather than actual facial resemblance.

Alternatively, future studies may show that results vary even when highly standardised stimulus sets and methods are employed, which would suggest that there is no generalisable rule to kin resemblance beyond general increased resemblance.

Chapter 6:

General Discussion

6.1. Summary of Experimental Findings

In summary, across four studies we found that raters were able to reliably identify related and unrelated child sibling pairs and parent-child pairs, which is also a robust finding across the literature in general (e.g., 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). Hence, third party kinship recognition seems to be a universal skill.

Subsequently, each study addressed a different area of third party kinship recognition that had received no attention or that needed further strategic investigating.

Chapter 2 investigated the contribution of shape and surface reflectance information to kinship detection in 3D face images of children. This study was based on research suggesting that facial morphology and skin texture/tone are heritable (Clark, Stark, Walsh, Jardine, & Martin, 1981; Djordjevic, Zhurov, & Richmond, 2016; Frisancho, Wainwright, & Way, 1981; Kim et al., 2013; Tsagkrasoulis, Hysi, Spector, & Montana, 2017; Weinberg, Parsons, Marazita, & Maher, 2013; Williams-Blangero & Blangero, 1991) and studies looking at what areas of the face inform kin recognition (Alvergne et al., 2014; Dal Martello & Maloney, 2006). Yet it was the first study investigating how we use them to make subjective decision about third party kinship from face images. It was also the first study to have both shape information and surface reflectance information presented separately. We found that raters were able to accurately judge kinship when they were only presented with shape information, as well as when they were only presented with only surface reflectance information. This suggests that both cues independently carry kinship information that identifies kinship pairs as being related, and unrelated pairs as being unrelated. Moreover, individual shape and reflectance information are optimally combined to make kinship judgments. This means that when we are judging faces, we are using both shape and surface reflectance information to make decisions about relatedness. The experimental design did not allow us to conclusively distinguish between whether kinship judgments were based on face similarities due to genetic or shared environmental sources.

Chapter 3 investigated the effect of facial expression of children on third party kin recognition. This study was based on a computer algorithm study, which reported that kinship detection is above chance when smiling faces are given to a computer algorithm. Yet, this study did not include any human raters or a comparison task with neutral faces. Anecdotally, it is very common to compare the facial expressions of family members and some research suggests that a smiling facial expression aids the identification of an individual's nationality (Marsh, Elfenbein & Ambady, 2003, 2007). Nonetheless, we found that a smiling facial expression decreased kinship judgment accuracy compared to a neutral facial expression. This might be related to the conflicting mechanisms that are employed when faces are processed for emotions and kinship: When processing emotions, the lower half of the face and specifically the mouth region are observed (Schyns, Bonnar, & Gosselin, 2002), whereas when processing kinship cues, the upper half of the face is in focus (Dal Martello & Maloney, 2006). Therefore, processing the emotion information from the lower half of the face could distract from the upper half of the face carrying more kinship information. Additionally, the shape of the face is distorted while smiling, potentially decreasing the availability of kinship cues. It is also important to note that we found an overall decrease in relatedness judgments with an increased age gap between the two siblings in a pair, which could suggest that it is harder to detect related pairs when siblings are at different stages of facial development.

Chapter 4 investigated the effect of birth order of the rater on third party kin recognition. This study was based on research that claimed that firstborns are less accurate at judging kinship than laterborns, as firstborns are relying on contextual cues such as perinatal association that laterborns do not have to identify their own kin and they will therefore not be as experienced as laterborns in using facial appearance to inform kinship judgments (Kaminski et al., 2010). According to this theory, laterborns mainly use facial resemblance to identify their own kin and are therefore more experienced in the use of this physical cue to detect kinship also in other unknown pairs. It is of course possible that laterborns rely more on the use of facial resemblance than firstborns, yet, laterborns still access a number of other contextual cues that indicate kinship to their siblings, such as cohabiting. Consequently, I decided to investigate this observation further and replicate the original study with an increased number of high quality stimuli and raters. We did not find that birth order has any effect on the ability to judge kinship, a result that is in line with another study that failed to replicate this effect of birth order (Alvergne et al., 2010).

Chapter 5 investigated the methods used in the third party kin recognition literature. This study was motivated by the fact that widely varying methods have been used in the literature and the fact that studies investigating the same factors have often found contradictory results, which might be the consequence of using diverging methods. Overall, we found that the three most common methods were correlated when looking at overall accuracy in the tasks. However, there are differences in detection accuracy depending on relatedness between tasks. Raters in the matching task and the similarity task were significantly better at detecting related pairs than raters in the kinship task. Inversely, raters in the kinship task were significantly better at detecting unrelated pairs than raters in the similarity task. This would suggest that using different methods in the literature could certainly lead to conflicting results, as the same related or unrelated pairs will be judged differently depending on the task that is used. In addition, when looking at the contentious effect of sex of and age of stimuli, we found that task choice can influence the results. In the kinship task only, we found a significant main effect of sex composition of the pair, whereby same-sex pairs were judged to be related more often than opposite-sex pairs, irrespective of actual relatedness. In the matching task only, we found an interaction between sex difference and age difference, whereby an increased age difference decreased relatedness judgments for same-sex pairs, but marginally increased relatedness judgments for opposite-sex pairs.

These four studies presented in this thesis investigate and replicate existing claims and observations, such as the effect of birth order and the effect of sex and age of stimuli on kin detection. This research also expands the literature with new ideas and creates avenues for future research, such as the effect of facial expressions and the contribution of shape and surface reflectance information to kinship recognition. In a nutshell, all four pieces of research confirm that humans are capable of judging kinship not only among their own kin but also more broadly when just presented with facial information of complete strangers. Kinship is crucial to biological theories of social and sexual behavior; hence this line of research could provide important insight into the wider implications of kinship detection beyond the nuclear family.

11.2. Limitations

One limitation of my PhD research is that the ethnicity of the stimuli we were able to collect was almost exclusively white, hence we are not able to present and represent diverse ethnicities in our research. All our photographs were collected in the Glasgow Science Centre, therefore we were always in the same place within Glasgow. Glasgow itself has a diverse community, but not to the extent where we would have been able to represent them in our photographs in equal numbers. Also, as we never moved location we did not access areas that have a higher ratio of non-white ethnicities, which would have helped to increase the diversity of our photographs. The reasoning behind staying in the Glasgow Science Centre rather than moving around different locations to take photographs was based on the practical difficulties of moving all of the heavy and bulky equipment we used to take highly standardized pictures. We did not just use one camera in any room with any lighting, but we had a highly restricted and regulated set up which allowed us to take standardized pictures of thousands of people over a few weeks. Therefore, moving the equipment on a more regular basis was just not feasible unfortunately. This also meant that any collaborations with face researchers in other countries was impossible, as we could not provide them the same equipment we had. This also excluded utilizing existing face photographs from previous research that depicted different ethnicities, as we wanted to make sure that all our photographs were of the exact same quality and standard. Therefore, this means that our research can only really inform conclusions on third party kin recognition in white faces. We do not have any reason to believe that these insights are not generalizable to kin recognition in faces of other ethnicities, but this should be tested in the future.

Another limitation of our research was that we were not able to collect a high number of photographs of different family constellations. As mentioned before, as we were exclusively based in the Glasgow Science Centre, we mostly encountered family units that were comprised of the parents and the children. Some families also had a grandparent with them, and some had aunts and cousins present as well, but unfortunately that was definitely the exception rather than the norm. Therefore, our research was mostly based on siblings or parent- child constellations. It would have been fantastic to investigate further the effect of degree of kinship on kinship detection, but we simply did not have enough grandparents, aunts, uncles or cousins to conduct any meaningful analyses and, as mentioned above, we are not able to use existing photographs from other researchers to increase numbers. This is definitely one of the limitations to our stimuli set that should be addressed to be able to conduct broader kinship research.

A further limitation was that we were only recruiting raters that were 16 years old or older and therefore did not have a chance to investigate how accurate children are at detecting third party kinship, and how this skill changes over time. We only recruited raters that were 16 years or older as we conducted all our studies online, therefore the raters needed to have access to a computer and to our website, and they also had to be able to complete the tasks using the computer. Recruiting children as raters would have required a different approach to our testing, for example bringing them into the lab and walking them through every trial, asking them what they thought and noting down the response. This was unfortunately outwith the scope of my doctoral work.

A massive limitation, and I am sure every PhD candidate feels the same way, was time. At the beginning, a PhD seems like a long time to pursue all your research dreams, but the years flew by and even though I was able to realise a lot of my study ideas myself or through collaborations with other researchers (not included in this thesis) and I am very proud of them, there are still so many things I wish I was able to investigate and clarify. One of the major things I wish I had the time to achieve was to replicate my own studies, as replications are crucial to establish robust findings and theories, especially in a field with such contradictory findings.

11.3. Future Directions/Closing Remarks

There is only a limited amount of research and studies done on the topic of third party kin recognition, which means that we are missing a key piece to our understanding of altruistic and sexual behavior. It would be of great value if more researchers would dedicate their time to this topic to understand the intricate mechanism that is behind our ability to detect not only our own kin but also other kin relations just from a face photograph. However, it is crucial that the current literature is critically examined rather than just expanding the literature by introducing new ideas. This could be achieved by revisiting some of the ideas introduced already and reproducing and rebuilding our understanding and assumptions on the topic. We cannot keep building studies on fallacious knowledge and vague concepts, but need to first establish a strong base from which further research can expand and thrive. Reproducing previous findings is especially crucial in this field as the findings have been very mixed and it is still unclear why, even though I made an attempt at clarifying some of these uncertainties during my doctoral research. Furthermore, it is pivotal that studies in this field take into special consideration the validity and quality of all components of their research. For example, if the stimuli used in a study are of poor quality, include background information, or are reused throughout the study, then this could lead to results that are based on information extrinsic to the face, and hence invalidate the findings completely. The same goes for methodology, we have identified in this thesis that the method used in a study significantly influences the findings, again possibly invalidating the finding as it is not based on facial recognition of kinship cues but rather based on other factors introduced by the study components. The use of these different methods also makes it impossible to directly compare different studies, and as there are not that many published this is a critical issue.

Throughout this thesis I have also talked about pre-registrations, preprints, open access and sharing data that all full under the umbrella of open science practices. I strongly believe that by working together, sharing our work, making all data and analyses accessible, we can produce better science. It increases the accountability of the researcher, it amplifies transparency and it solicits an open dialogue between researchers. We can work together to produce a stronger piece of research that is then available to anyone with an interest in the topic. Moreover, over time statistical models have reached new levels of sophistication, hence if we could take all that data that is available from all these different studies done over the past 30 years and run new and improved statistical analyses on them, we would be able to do valuable work with existing data. Obviously, most of the concepts mentioned here have not been around for that long, hence it is not a surprise that studies from 30 years ago do not have their data, analysis, etc. openly available on the internet. However, moving forward, new studies do need to consider being part of the open science movement, as otherwise we will not be able to reach a stage of open and free discussion. I am also aware that I have mixed a few concepts in here, with open science being an advocate for given everyone access to knowledge, whereas pre-registrations might be more focused on holding research accountable. Pre-registrations discourage fishing for results as the analysis is pre-set, no matter whether that will yield a significant result or not. This is further supported by openly sharing data and final analysis plans after the study was conducted, as anyone can go and look at the analysis and the data, increasing the accountability of the researcher and decreasing the likelihood of dishonest or substandard analyses.

To conclude, moving forward, third party kin recognition needs further methodological research in order to ensure the validity of stimuli and measures. This research will need to build itself up from the beginning, however, as the foundations in the literature so far are not strong enough to support it. These studies need to aim to replicate and rebuild a field of research that has been neglected and that requires stronger, higher quality and consistent research. By employing standardized methods, high quality stimuli and appropriate analyses, we will be able to learn so much more about our ability to recognize kin and how this impacts our daily lives. Lastly, moving forward it is crucial that these studies will be dedicated to open science, pre-registering their hypothesis and analysis, preprinting their findings and making their data and analysis openly available to the public to stimulate debate and discussions that do not discriminate and recognize reliable and honest science.

Appendices

Appendix 1:

Supplemental Material for 'Birthorder does not affect ability to detect kin', Chapter 9

These are the supplemental materials for the manuscript "Birthorder does not affect kin recognition". Here we run the exact analysis with the pre-registered 100 participants only rather than the 109 participants that were eligible after overshooting our recruitment target. Limiting our analysis to only the pre-registered number of raters does not change the results of the analysis.

Raters

The laterborns group was made up of 41 raters with only older siblings and 9 raters with an older and a younger sibling. Firstborns (n= 50) only had younger siblings.

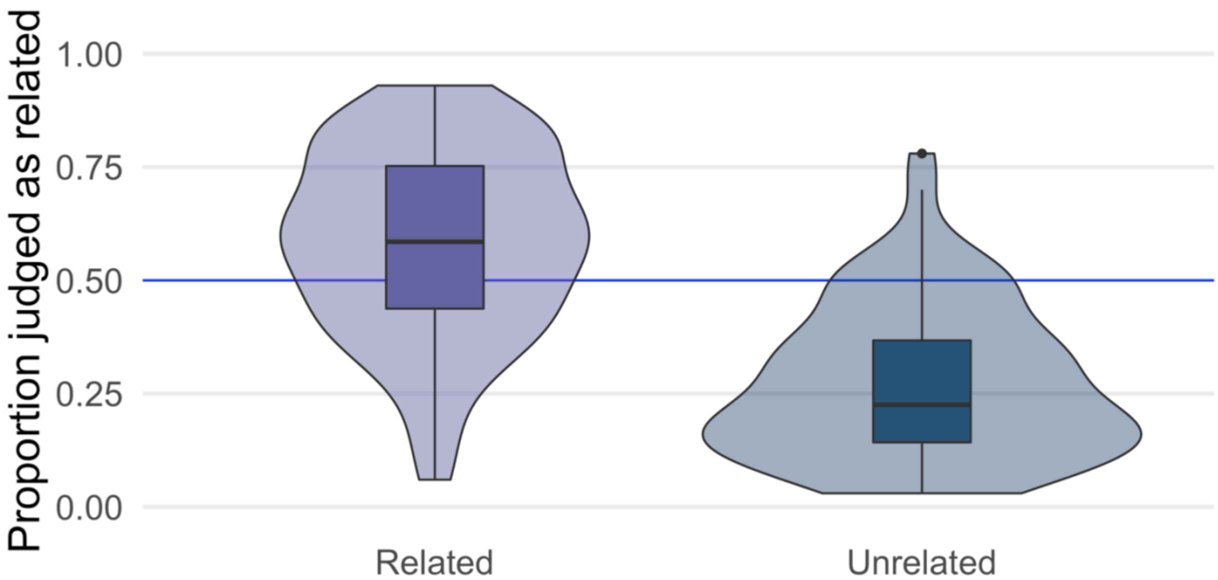
After filtering, the responses from 21 laterborn men (mean age = 30.13 years; SD = 13.96 years), 28 laterborn women (mean age = 25.95 years; SD = 10.28 years) and 1 laterborn of unspecified gender (age = 23 years) were analysed along with 18 firstborn men (mean age = 26.33 years; SD = 4.36 years), 31 firstborn women (mean age = 30.25 years; SD = 15.05 years) and 1 firstborn of unspecified gender (age = 17.1 years). Raters were predominantly white (81 out of 100 raters). Data from the excluded raters can be found in the data file used for the analysis, with the exclusion criteria being clearly marked in the analysis code (both available at osf.io/h43ep).

Results

We used a binomial logistic mixed model to predict relatedness judgments from actual relatedness (effect coded as related = +0.5 and unrelated = -0.5), birth order (effect coded as firstborns = +0.5 and laterborns = -0.5) and the interaction between birth order and relatedness in the kinship task. We included the rater ID and stimulus ID as random effects and specified our slopes maximally.

The analysis revealed a main effect of relatedness (*ß*=1.69, *95% CI* [1.32;2.06], *SE*=0.19, *z*=8.93, $p < .001$, *odds ratio*=5.42, *95% CI* [3.75;7.88]), whereby actual related pairs were 5.42 times more likely to be judged as related than unrelated pairs (see Figure 14).

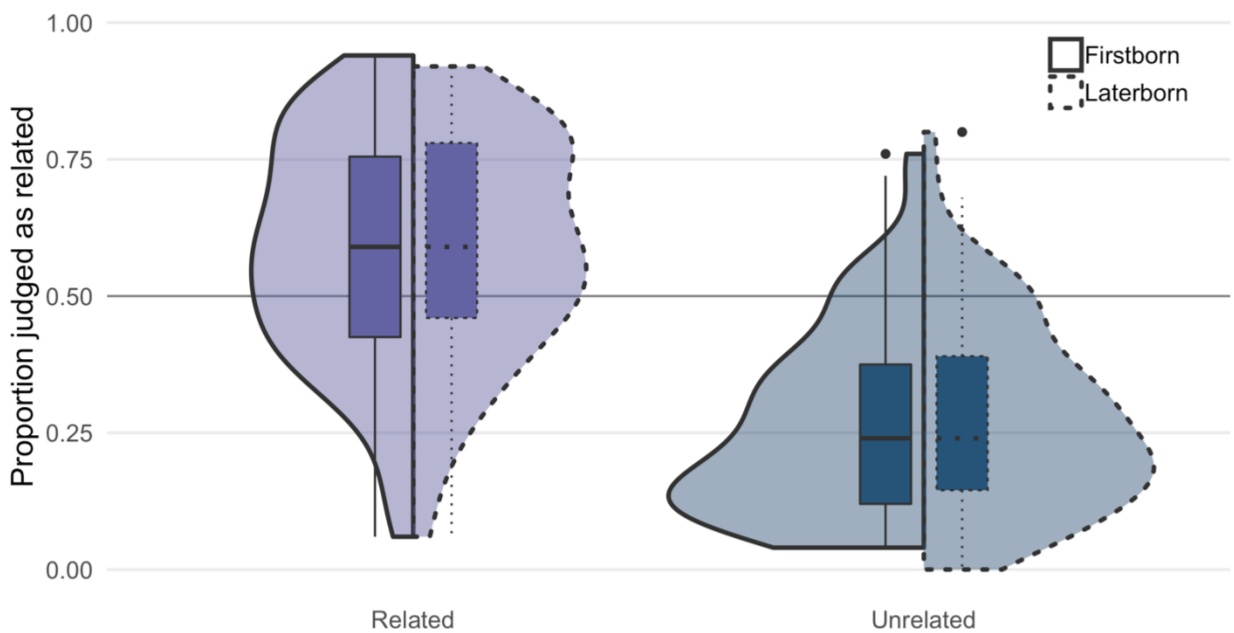
Figure 14. The main effect of relatedness on proportion judged as related.



There was no main effect of birth order (*ß*=-0.05, *95% CI* [-0.32;0.22], *SE*=0.14, *z*=-0.37, *p* = 0.708, *odds ratio*=0.95, *95% CI* [0.72;1.24]) and no interaction between birth order and relatedness (*ß*=0.09, *95% CI* [-0.15;0.34], *SE*=0.12, *z*=0.76, *p* = 0.449, *odds ratio*=1.09, *95% CI* [0.86;1.4]), see Figure 15.

In fact, when we look at the non-significant difference between firstborns and laterborns (not pre-registered), we see that firstborns tended to be more accurate in their kinship judgments (*ß*=1.75, *95% CI* [1.35; 2.15], *SE*=0.2, *z*=8.65, *p* < .001, *odds ratio*=5.75, *95% CI* [3.88; 8.57]) than laterborns (*ß*=1.64, *95% CI* [1.26; 2.02], *SE*=0.2, *z*=8.39, *p* < .001, *odds ratio*=5.16, *95% CI* [3.51; 7.54]), opposite to the prediction by Kaminski et al. (2010).

Figure 15. The interaction between relatedness and birth order on proportion of face pairs judged as related.



Appendix 2:

Study Information

Title of project: How do humans recognize kin?

Investigator: Dr Lisa DeBruine

You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others, if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of this study?

We are interested in how people determine who feels "related" to them. We might just rely on what we're told, or we may also use situational clues like whether and when you lived in the same house. Physical cues, like family resemblance, may also shape these feelings. However, we don't know exactly how we perceive family resemblance in the face, so we want to take 3D pictures of a lot of relatives to see what parts of our faces can show family relatedness.

What will happen to me if I take part?

You will sit down and have a picture taken of your face with six cameras at the same time. It's that easy!

What will happen to my face picture then?

You can decide exactly how you want us to use your face picture. We could just use it to build our computer models of how family resemblance is expressed. This will help us to make more realistic "virtual relatives" for our research.

You could also decide to let us show your picture to people in the lab or online in studies about family resemblance. People would try to guess who is related and we would see if our virtual relatives are mistaken for real relatives as often as real relatives are identified.

Finally, you could decide to let us use your picture for face research in general and to illustrate our research in scientific papers and even in the popular media.

What will happen to the study results?

In accordance with good research practice, they will be kept securely for a minimum of 10 years in our secure data archive.

What are the possible risks/side effects of taking part?

There are no known or foreseeable risks or side effects.

What are the possible benefits of taking part?

There will be no direct benefit to you, apart from contributing to our understanding of human kinship.

What happens at the end of the study?

The results of this study may be published in a journal or used for teaching purposes. The results may also be presented at scientific meetings or in talks at academic institutions. Results will always be presented in such a way that data from individual volunteers cannot be identified.

Confidentiality - who will have access to the data?

The data will be stored on a secure computer network. Your face will be identified only by a numeric code and this will be kept separate from your name on your consent form. Only members of our research team have access to this data.

Can I ask questions about the research project?

You can email us at info@faceresearch.org with any questions.

Can I withdraw from the study?

Your participation to this study is voluntary and you may withdraw from the research at any time and for any reason, without explaining why.

Will I receive a financial compensation?

You will receive no compensation.

This research study has been approved by the Ethics Committee of the College of Science and Engineering at University of Glasgow (Ethics Application No: D1424324021876).

Contact details

Name: Dr Lisa DeBruine

Address: Institute of Neuroscience and Psychology
58 Hillhead Street, Glasgow G12 8QB
Telephone: 0141 330 5351
Email: lisa.debruine@glasgow.ac.uk

Appendix 3:

Consent Form

Face Research Lab

Image Consent Form

University of Glasgow

Here we are collecting face images to use in our research. Computer graphic methods allow us to manufacture composite faces by averaging the shape, colour and texture from a sample of faces (see examples of these synthetic faces below). If you agree, we can also use your images to help make computer-graphic faces that let us investigate agreement and individual differences in social attitudes to facial cues.

There are a few points we would like to emphasise to you at this time:

- You can withdraw from the project at any point and without penalty
- You can contact the experimenters at any time to request that your images be destroyed
- All images are anonymous; your name or identifying information (besides your facial appearance) will never be attached

Below are examples of 5 individual faces (left) and their composite face image (right)
(Note that no single identity is recognisable in these synthetic composite faces)

Please indicate below in which ways you consent for us to use your face photographs by initialling the boxes below

	YES, I consent	**NO, I don't consent**
I consent to have my photographs taken for this project and analysed by researchers or included in composite images in which my identity is not recognisable		
I consent for my photographs to be shown to participants in laboratory studies in their original or altered forms		
I consent for my photographs to be used in web-based studies in their original or altered forms and to illustrate research (e.g. in scientific journals, news media or presentations)		

I have read and understand the above and give my informed consent to those parts indicated.

Name (please print): ______________________________

Signature: Date:

Please contact info@faceresearch.org for further information

ID: ______________________________

Appendix 4:

Spreadsheet of studies

The following pages list all major studies done concerning third party kin recognition

Reference	Alvergne, A., Faurie, C., & Raymond, M. (2007). Differential facial resemblance of young children to their parents: who do children look like more?. Evolution and Human behavior, 28(2), 135-144.
Resemblance detection	Yes
Participant number	209
Participant sex	mixed (89f/120m)
Participant age	no info
Total Stimuli Number	332(4 age groups of children with approx 21 children stimuli each)
Stimuli ethnicity	French
Stimuli sex	mixed
Stimuli age	4 age groups= 1-3 days/8m-1y10m/ 2y-3y6m/4y-6y 2m
Stimuli relatedness	parent-child
Stimuli	colour photographs taken by experimenter, backgrounds black
Subject Design	between= age classes
Method	match parent-child
Method Details	1 child & 3 parents (1 true parent)
Results	Detection kinship significantly above chance

Reference	Alvergne, A., Oda, R., Faurie, C., Matsumoto-Oda, A., Durand, V., & Raymond, M. (2009). Cross-cultural perceptions of facial resemblance between kin. Journal of Vision, 9(6), 23-23.
Resemblance detection	Yes
Participant number	362 (170 French & 192 Senegalese)
Participant sex	mixed (196f/166m)
Participant age	mean= 29y
Total Stimuli Number	73
Stimuli ethnicity	French & Senegalese
Stimuli sex	mixed (36f/35m)
Stimuli age	2y-6y, adults unknown
Stimuli relatedness	parent-child
Stimuli	color pictures, backgrounds in white for Senegalese stimuli, black background for French stimuli
Subject Design	within
Method	match parent-child
Method Details	1 child & 3 parents (1 true parent)
Results	Detection kinship above chance regardless of origin (.53, .53, .45,.46)

Reference	Alvergne, A., Oda, R., Faurie, C., Matsumoto-Oda, A., Durand, V., & Raymond, M. (2009). Cross-cultural perceptions of facial resemblance between kin. Journal of Vision, 9(6), 23-23.
Resemblance detection	Yes
Participant number	235 (80 french, 81 Senegalese, 74 Senegalese living in France)
Participant sex	mixed (133f/102m)
Participant age	mean ages between 26y-32.4
Total Stimuli Number	41
Stimuli ethnicity	French & Senegalese
Stimuli sex	mixed (22f/19m)
Stimuli age	4y-6y, adults unknown
Stimuli relatedness	parent-child
Stimuli	color pictures, backgrounds in white for Senegalese stimuli, black background for French stimuli
Subject Design	within
Method	match parent-child
Method Details	1 child & 3 parents (1 true parent)
Results	Detection kinship above chance regardless of origin

Reference	**Resemblance detection**	**Participant number**	**Participant sex**	**Participant age**	**Total Stimuli Number**	**Stimuli ethnicity**	**Stimuli sex**	**Stimuli age**	**Stimuli relatedness**	**Stimuli**	**Subject Design**	**Method**	**Method Details**	**Results**
Alvergne, A., Perreau, F., Mazur, A., Mueller, U., & Raymond, M. (2014). Identification of visual paternity cues in humans. Biology letters, 10(4), 20140063.	YES 67% paternal detection	271	male (146), female (125)	in supplementary excel file which wasn't accessible	54	White American	male	21-26y (mean= 22.7, s.d.= +-1.4), photos of fathers/sons at same age	father and sons	Black and White potraits from West Point Academy	between, 4 conditions: 1- orginal face, 2- upper and lower part, 3 - inner and external, 4- mixed faces	Match father-son pair	1 son & 3 fathers; 6 stimuli created: original, upper part, lower part, inner/ external features, mixed composite	67% of the pairs accurately detected, EXCEPT when only lower face shown (here not significant detection)

Reference	Arantes, J., & Berg, M. E. (2012). Kinship recognition by unrelated observers depends on implicit and explicit cognition. Evolutionary Psychology, 10(2), 147470491201000204.
Resemblance detection	Yes, 62.5% mother daughter pairs detected
Participant number	50
Participant sex	Male
Participant age	mean= 22 years, SD= 2.97, range= 17-35
Total Stimuli Number	40
Stimuli ethnicity	Portuguese
Stimuli sex	female
Stimuli age	daughters mean= 23y, range= 21-29; mothers mean=48.5y, range= 44-55
Stimuli relatedness	mother-daughter
Stimuli	Color passport photos for daughters, similar photographs of biological mothers provided by daughter. Photographs digitized and backgrounds rendered similar (grey)
Subject Design	Mixed - 2 groups of raters seeing the same stimuli. 20 trials, mothers shown twice, once as real and once as control
Method	Match mother-daughter
Method Details	1 daughter & 2 mothers (1 true)
Results	65,2% correct matches

Brédart, S., & French, R. M. (1999). Do babies resemble their fathers more than their mothers? A failure to replicate Christenfeld and Hill (1995). Evolution and Human Behavior, 20(2), 129-135.	**Reference**
Yes, 7-14% above chance	**Resemblance detection**
180	**Participant number**
mixed (half/half)	**Participant sex**
18-30y (median=28)	**Participant age**
140	**Total Stimuli Number**
Caucasian	**Stimuli ethnicity**
mixed	**Stimuli sex**
same child at 1y, 3y, 5y & both parents at child's age of 1y	**Stimuli age**
parent-child	**Stimuli relatedness**
photographs provided by families, scanned as 256 gray levels, 300DPI, size= 5x4cm, no glasses, beards, moustaches	**Stimuli**
Between. 6 blocks, randomly assigned	**Subject Design**
match parent-child pair	**Method**
1 child & 3 adults; 28 sets	**Method Details**
detection 7-14% above chance	**Results**

Reference	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.
Resemblance detection	Yes
Participant number	60
Participant sex	mixed (half/half)
Participant age	19-65y (median=38)
Total Stimuli Number	30
Stimuli ethnicity	Italian
Stimuli sex	mixed (half/half)
Stimuli age	8 years old, parent unknown
Stimuli relatedness	parent-child
Stimuli	Outdoor color photographs taken by researchers with different backgrounds, shoulders up
Subject Design	Within
Method	Degree facial similarity scale 0-10
Method Details	Parent-child pairs; Photoalbum, 40 pages, 40 parent-child pairs (20 related/20 non-related) : 2 labels (related/not related)
Results	Judged to look more like actual parent(M=5.12) than unrelated adult (M=4.22) $F(1,58)=132.83$, $p<.0001$; label bigger effect than biol relatedness

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.	**Reference**
Yes	**Resemblance detection**
60	**Participant number**
mixed (half/half)	**Participant sex**
18-70years (median=36)	**Participant age**
31	**Total Stimuli Number**
Italian	**Stimuli ethnicity**
mixed (half/half)	**Stimuli sex**
9 years old, parent unknown	**Stimuli age**
parent-child	**Stimuli relatedness**
Outdoor color photographs taken by researchers with different backgrounds, shoulders up	**Stimuli**
Within	**Subject Design**
Degree facial similarity scale 0-10	**Method**
Parent-child pairs; no labels	**Method Details**
Detection parent (M=5.21) over unrelated adult (M=4.21) $F(1,58)=81.84$, $p<.0001$	**Results**

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.	**Reference**
Yes	**Resemblance detection**
80	**Participant number**
mixed (half/half)	**Participant sex**
19-62y (median=30)	**Participant age**
32	**Total Stimuli Number**
Italian	**Stimuli ethnicity**
mixed (half/half)	**Stimuli sex**
10 years old, parent unknown	**Stimuli age**
parent-child	**Stimuli relatedness**
Outdoor color photographs taken by researchers with different backgrounds, shoulders up	**Stimuli**
Within	**Subject Design**
Degree facial similarity scale 0-10	**Method**
Parent-child pairs; no labels, were told that all related, 40 pairs (20 related, 20 NEW unrelated)	**Method Details**
Detection parent (M=5.46) over unrelated adult (M=4.48) $F(1,78)=31.27$, $p<.0001$	**Results**

Reference	Bressan, P., & Dal Pos, S. (2012). Fathers see stronger family resemblances than non-fathers in unrelated children's faces. Archives of sexual behavior, 41(6), 1423-1430.
Resemblance detection	Fathers gave higher similarity ratings than non fathers or mothers/non mothers
Participant number	140
Participant sex	mixed (half/half)
Participant age	18-70y (median=35)
Total Stimuli Number	60 (12 children at three ages and their 24 parents)
Stimuli ethnicity	Italian?
Stimuli sex	mixed (half/half)
Stimuli age	1y (median=14months, range=8-24months), 8y(median=8, range=6-10), 16y (range=13-21). Parents unknown
Stimuli relatedness	parent-child, child at different ages
Stimuli	most photos came from family albums of friends of one of the authors, some taken especially for the study, cropped from the shoulders up and printed on 9x12cm paper
Subject Design	between, 2 groups, 2 albums dividing the 108 stimuli pairs
Method	Degree facial similarity scale 0-10
Method Details	Some parent-child pairs, some child at different ages pairs, raters told that all pairs either parent-child pairs or same child at different ages
Results	Significant diff between non-fathers and fathers only when children 16, fathers give higher resemblance ratings. mothers vs non mothers no diff in ratings. Fathers also higher resemblance ratings for children at different ages

Bressan, P., & Dal Pos, S. (2012). Fathers see stronger family resemblances than non-fathers in unrelated children' s faces. Archives of sexual behavior, 41(6), 1423-1430.	**Reference**
Fathers not better at matching parent-child than others	**Resemblance detection**
140	**Participant number**
mixed (half/half)	**Participant sex**
18-70y (median=31)	**Participant age**
same stimuli as exp one, but only use part. 48 (12 children at 2 ages - 1y and 16y - and their parents)	**Total Stimuli Number**
Italian?	**Stimuli ethnicity**
mixed (half/half)	**Stimuli sex**
1y, 16y and parents unknown	**Stimuli age**
parent-child	**Stimuli relatedness**
most photos came from family albums of friends of one of the authors, some taken especially for the study, cropped from the shoulders up and printed on 9x12cm paper	**Stimuli**
within	**Subject Design**
match child-parent pair, can select same parent more than once regardless of previous choices	**Method**
1 child and 12 fathers/12 mothers. 2 stimulus boards, each all 12 fathers or 12 mothers taped ont it. An album with all the children at 2 ages, one child per page.	**Method Details**
fathers vs non fathers did not differ in matching ability, neither did mother vs non mothers. Ence fathers not actually better, just cognitive bias higher resemblance ratings in previous study	**Results**

Reference	Bressan, P., & Dal Pos, S. (2012). Fathers see stronger family resemblances than non-fathers in unrelated children's faces. Archives of sexual behavior, 41(6), 1423-1430.
Resemblance detection	Men giving higher resemblance ratings to 22 y.o. than women. This is data from an unpublished masters thesis, not much information
Participant number	
Participant sex	
Participant age	
Total Stimuli Number	
Stimuli ethnicity	
Stimuli sex	
Stimuli age	22
Stimuli relatedness	parent-child
Stimuli	
Subject Design	
Method	Facial similarity
Method Details	
Results	Men higher resemblance ratings than women, only when relatedness is mentioned in task though, if no reference made then no difference men/women

Bressan, P., & Grassi, M. (2004). Parental resemblance in 1-year-olds and the Gaussian curve. Evolution and Human Behavior, 25(3), 133-141.	**Reference**
Yes	**Resemblance detection**
80	**Participant number**
mixed (half/half)	**Participant sex**
18-60y	**Participant age**
120	**Total Stimuli Number**
Italian	**Stimuli ethnicity**
mixed (22 girl infants, 18 boy infants, 80 mixed parents)	**Stimuli sex**
children 2-18 months, parents unknown	**Stimuli age**
parent-child	**Stimuli relatedness**
Some from family album, some taken in lab using different backgrounds	**Stimuli**
between, 4 groups of 20 (10f/10m)	**Subject Design**
Degree facial similarity scale 0-10	**Method**
1 child & 3 adults; Photo album, 20 sets	**Method Details**
detection 1.37 times higher than chance	**Results**

Bressan, P., & Grassi, M. (2004). Parental resemblance in 1-year-olds and the Gaussian curve. Evolution and Human Behavior, 25(3), 133-141.	**Reference**
Yes	**Resemblance detection**
	Participant number
	Participant sex
	Participant age
	Total Stimuli Number
	Stimuli ethnicity
	Stimuli sex
	Stimuli age
	Stimuli relatedness
	Stimuli
	Subject Design
Match parent-child	**Method**
	Method Details
detection 1.47 times higher than chance	**Results**

Christenfeld, N. J., & Hill, E. A. (1995). Whose baby are you. *Nature*, *378*(6558), 669-669	**Reference**
Mixed, only 1y old boys to fathers	**Resemblance detection**
122	**Participant number**
no info	**Participant sex**
no info	**Participant age**
unclear, 24 families providing a maximum of 3 pictures of child, 48 parents providing a maximum of 2 pictures; minimum of 72 base stimuli	**Total Stimuli Number**
Caucasian	**Stimuli ethnicity**
mixed (half/half)	**Stimuli sex**
children at 1y, 10y, 20y; parents at child's age of 1y & a recent one, 12/24 children old enough to provide 3 pictures	**Stimuli age**
parent-child	**Stimuli relatedness**
photos provided by families	**Stimuli**
between, 18-21 participants rated each possible match	**Subject Design**
Degree facial similarity scale 0-10	**Method**
1 child (12 girls/12 boys) & 3 adults; child at 3 age points; parents at 2 age points	**Method Details**
Detection that same person at different age points above chance <55%; Detection of familial resemblance below chance apart from 1y old child matched to father above chance <49.2%	**Results**

Dal Martello, M. F., & Maloney, L. T. (2006). Where are kin recognition signals in the human face?. Journal of Vision, 6(12), 2-2.	**Reference**
Yes	**Resemblance detection**
109	**Participant number**
mixed (53 males/55 females) -> numbers don't add up to 109	**Participant sex**
19-36y (median=22)	**Participant age**
60	**Total Stimuli Number**
Italian, Caucasian	**Stimuli ethnicity**
mixed	**Stimuli sex**
17 months - 15 yrs	**Stimuli age**
siblings	**Stimuli relatedness**
color photographs, taken by experimenter under controlled lighting conditions, depicting child from neck up; replaced background with uniform dark gray field	**Stimuli**
between, 3 conditions (Full Face/Lower Half Masked/Upper Half Masked)	**Subject Design**
Related" or "Unrelated" Judgment; Told half siblings, half unrelated	**Method**
Sibling pairs; 30 Pairs (15 related/15 unrelated); within each group of 15= 5 boy pairs, 5 girl pairs and 5 boy-girl pairs	**Method Details**
Detection siblings above chance for Full Face condition & Lower Half Masked & Upper Half Masked; Upper Half Masked significantly worse but still able	**Results**

Reference	Dal Martello, M. F., & Maloney, L. T. (2006). Where are kin recognition signals in the human face?. Journal of Vision, 6(12), 2-2.
Resemblance detection	Yes
Participant number	111
Participant sex	mixed (39 males/72 females)
Participant age	20-31y (median=23)
Total Stimuli Number	60
Stimuli ethnicity	Italian, Caucasian
Stimuli sex	mixed
Stimuli age	18 months - 15 yrs
Stimuli relatedness	siblings
Stimuli	color photographs, taken by experimenter under controlled lighting conditions, depicting child from neck up; replaced background with uniform dark gray field
Subject Design	between, 3 conditions (Full Face/Eyes Masked/Mouth Masked), 37 people in each condition
Method	Related" or "Unrelated" Judgment; Told half siblings, half unrelated
Method Details	Sibling pairs; 30 Pairs (15 rel/15 unrel); within each group of 15= 5 boy pairs, 5 girl pairs and 5 boy-girl pairs
Results	Detection siblings above chance for all conditions; Eyes Masked worse than Mouth Masked but still able

Reference	Dal Martello, M. F., & Maloney, L. T. (2010). Lateralization of kin recognition signals in the human face. Journal of vision, 10(8), 9-9.
Resemblance detection	Yes
Participant number	124
Participant sex	mixed (65 males/59 females)
Participant age	19-36y (median=22.5)
Total Stimuli Number	60
Stimuli ethnicity	Italian, Caucasian
Stimuli sex	mixed
Stimuli age	17 months - 15 yrs
Stimuli relatedness	siblings
Stimuli	color photographs, taken by experimenter under controlled lighting conditions, depicting child from neck up; replaced background with uniform dark gray field
Subject Design	between, 3 conditions (Full Face/Left Hemi Masked/Right Hemi Masked)
Method	Related" or "Unrelated" Judgment; Told half siblings, half unrelated
Method Details	Sibling pairs; 30 Pairs (15 related/15 unrelated); within each group of 15= 5 boy pairs, 5 girl pairs and 5 boy-girl pairs
Results	Detection siblings above chance for Full Face condition & Left Hemi Masked & Right Hemi masked; No significant diff between conditions either

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.	**Reference**
Yes	**Resemblance detection**
118 (divided into 4 groups)	**Participant number**
mixed (17f/13m; 27f/7m; 23f/4m; 24f/3m)	**Participant sex**
mean age=20.6y; 22.2y; 20.8y; 20.5y)	**Participant age**
104	**Total Stimuli Number**
European ethnicity	**Stimuli ethnicity**
mixed	**Stimuli sex**
28y-46y (mean=37.9y)	**Stimuli age**
two imagesets: non-identical twins; siblings & their **matched** unrelated pairs	**Stimuli relatedness**
color, full face, taken by experimenter, grey background; unmasked faces and masked (hair, ears & neck removed) stimuli	**Stimuli**
Between, 4 conditions (unmasked/unmasked & related or not/similarity rating)	**Subject Design**
BOTH Degree Facial Similarity / "Related" or "Unrelated" Judgment	**Method**
Sibling Pairs	**Method Details**
Similarity ratings and kinship judgments highly correlated but not synonymous	**Results**

Reference	Resemblance detection	Participant number	Participant sex	Participant age	Total Stimuli Number	Stimuli ethnicity	Stimuli sex	Stimuli age	Stimuli relatedness	Stimuli	Subject Design	Method	Method Details	Results
Dibeklioglu, H., Ali Salah, A., & Gevers, T. (2013). Like father, like son: Facial expression dynamics for kinship verification. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1497-1504).	Computer verification	n/a	n/a	n/a	152		mixed	8y-74y	parent-child	Videos		tracking landmarks, temporal completed local binary pattern desscriptors (CLBP), classified by support vector machines (SVM)	n/a	73% Computer verification

Reference	Kaminski, G., Dridi, S., Graff, C., & Gentaz, E. (2009). Human ability to detect kinship in strangers' faces: effects of the degree of relatedness. Proceedings of the Royal Society of London B: Biological Sciences, 276(1670), 3193-
Resemblance detection	yes
Participant number	59
Participant sex	mixed (54f/5m)
Participant age	18-35y (mean=21.6)
Total Stimuli Number	236
Stimuli ethnicity	Caucasian
Stimuli sex	mixed
Stimuli age	no info
Stimuli relatedness	families (including siblings, parents, aunts, uncles, grandparents, cousins)
Stimuli	Family photos, cut head including ears, excluding neck and hair, black and white
Subject Design	between, 2 sets
Method	Related" or "Unrelated" Judgment within 20 seconds
Method Details	Pairs; 2 sets, 60 and 58 stimuli respectively (half related, half unrelated)
Results	Detection kinship above chance, no matter what relatedness category; no reaction time difference between related/unrelated pairs

Reference	Kaminski, G., Gentaz, E., & Mazens, K. (2012). Development of children's ability to detect kinship through facial resemblance. Animal cognition, 15(3), 421-427.
Resemblance detection	Mixed
Participant number	143
Participant sex	mixed
Participant age	6 age groups: 5y-11y
Total Stimuli Number	128
Stimuli ethnicity	Caucasian
Stimuli sex	mixed (half/half)
Stimuli age	Mean age 101h for neonates; adult unknown
Stimuli relatedness	parent-child
Stimuli	Photos taken in nursery, faces pasted on grey background
Subject Design	Within
Method	match the child-parent pair
Method Details	1 neonate & 3 adult OR 1 adult & 3 neonates; 32 panels
Results	38.8%, significant p<0.001 when match 1 parent ; BUT match the 1 neonate NOT significant

Reference	Kaminski, G., Méary, D., Mermillod, M., & Gentaz, E. (2010). Perceptual factors affecting the ability to assess facial resemblance between parents and newborns in humans. Perception, 39(6), 807-818.
Resemblance detection	Yes
Participant number	90
Participant sex	mixed (66f/24m)
Participant age	mean= 21.5y
Total Stimuli Number	129
Stimuli ethnicity	Caucasian
Stimuli sex	mixed (half/half)
Stimuli age	Mean age 101h for neonates; adult unknown
Stimuli relatedness	parent-child
Stimuli	Photos taken in nursery, faces pasted on grey background
Subject Design	Between; 2 groups= same items, but 1 group saw them in black and white, other group in colour
Method	match the child-parent pair after 25s exposure, in 5s interval between panels
Method Details	1 neonate & 3 adult OR 1 adult & 3 neonates; 32 panels shown for 25secs
Results	Detection above chance (Z=1.8, P=0.035); significant kinship detection for 34.4% of participants; significant kinship detection of 18 out of 32 items

Reference	Kaminski, G., Méary, D., Mermillod, M., & Gentaz, E. (2010). Perceptual factors affecting the ability to assess facial resemblance between parents and newborns in humans. Perception, 39(6), 807-818.
Resemblance detection	
Participant number	67
Participant sex	mixed (94%women)
Participant age	mean=19y
Total Stimuli Number	64 neonates for gender distinctiveness task; 96 adult&neonates for attractiveness task
Stimuli ethnicity	Caucasian
Stimuli sex	mixed (half/half)
Stimuli age	Mean age 101h for neonates; adult unknown
Stimuli relatedness	unrelated
Stimuli	Photos taken in nursery, faces pasted on grey background
Subject Design	within
Method	Guess gender of neonate; Choose the face they preferred out of 3
Method Details	See each neonate individually for 5 secs for gender distinctiveness task; see 3 adults/neonates for 25 secs for attractiveness task
Results	Correct kinship identification decreased with increasing attractiveness and increased with increasing gender distinctiveness

Reference	Maloney, L. T., & Dal Martello, M. F. (2006). Kin recognition and the perceived facial similarity of children. Journal of Vision, 6(10), 4-4.
Resemblance detection	Similarity ratings inform kinship judgment
Participant number	64
Participant sex	mixed (42f/22m)
Participant age	19-33y, median=23
Total Stimuli Number	60
Stimuli ethnicity	Italian, Caucasian
Stimuli sex	mixed (half/half)
Stimuli age	17months to 15y
Stimuli relatedness	siblings
Stimuli	colour photos taken by experimenter from neck to top of head, dark grey field background
Subject Design	between; 2 conditions "similarity"/"kinship"
Method	In similarity task participants rated facial similarity of pair (0-10), given no info on relatedness; in kinship task 'Related" or "Unrelated" Judgment; Told half siblings, half unrelated
Method Details	Sibling pair; 30 (15related/15 unrelated): 10boy-boy, 10girl-girl, 10boy-girl
Results	Children judged to be more similar more likely to be classified as siblings= a rating of 4.79 corresponds to even odds that the children are related

McLain, D. K., Setters, D., Moulton, M. P., & Pratt, A. E. (2000). Ascription of resemblance of newborns by parents and nonrelatives. Evolution and Human behavior, 21(1), 11-23.	**Reference**
Yes, mean=3.67 (3.33 for significance)	**Resemblance detection**
160	**Participant number**
mixed (100f/60m)	**Participant sex**
mean= 20y	**Participant age**
70	**Total Stimuli Number**
States, ethnicity not specified	**Stimuli ethnicity**
mixed	**Stimuli sex**
1-3 day old newborns; adults unknown	**Stimuli age**
parent-child	**Stimuli relatedness**
Taken in hospital, face against pillow for mother, white background for father	**Stimuli**
within	**Subject Design**
match a father and mother to neonate	**Method**
1 neonate & 6 adults (3f&3m); 10 stimuli boards	**Method Details**
Detection kinship above chance , t-test; 3.67 mean (3.33 for significance)	**Results**

Reference	McLain, D. K., Setters, D., Moulton, M. P., & Pratt, A. E. (2000). Ascription of resemblance of newborns by parents and nonrelatives. Evolution and Human behavior, 21(1), 11-23.
Resemblance detection	Yes, mean=3.69 (3.33 for significance)
Participant number	156
Participant sex	females
Participant age	mean= 20y
Total Stimuli Number	
Stimuli ethnicity	
Stimuli sex	
Stimuli age	
Stimuli relatedness	
Stimuli	
Subject Design	
Method	match child-parent pair
Method Details	1 neonate & 3 adults; 20 stimulus boards
Results	Detection kinship above chance , t-test; 3.69 mean

Reference	Nesse, R. M., Silverman, A., & Bortz, A. (1990). Sex differences in ability to recognize family resemblance. Ethology and Sociobiology, 11(1), 11-21.
Resemblance detection	Yes, mean of 14.78 out of 24 detected
Participant number	200
Participant sex	mixed (108f/92m)
Participant age	15-78y (mean=30)
Total Stimuli Number	48
Stimuli ethnicity	Caucasian
Stimuli sex	mixed
Stimuli age	children 6months-18 yr; adults unknown
Stimuli relatedness	parent-child
Stimuli	Colour photographs taken by experimenter, cropped so only neck and head visible
Subject Design	within
Method	Related" or "Unrelated" Judgment; no info given about relatedness
Method Details	Parent-child pairs; 24 items (half related/unrelated) & 4 subsections: Mother-daughter/Mother-son/Father-daughter/Father-son
Results	Detection kinship above chance, mean of 14.78 out of 24 items correct detection

Reference	Oda, R., Matsumoto-Oda, A., & Kurashima, O. (2002). Facial resemblance of Japanese children to their parents. Journal of Ethology, 20(2), 81-85.
Resemblance detection	Assessing resemblance to father, no difference between children's sex (only in one unrealistic condition)
Participant number	209
Participant sex	mixed (56f/153m)
Participant age	18-24y
Total Stimuli Number	117
Stimuli ethnicity	Japanese
Stimuli sex	mixed (17 boys/21girls) & parents
Stimuli age	Children 3y-6y; adults unknown
Stimuli relatedness	parent-child
Stimuli	black and white photos taken my examiner, just inside of face, no ears, hair; shaded around the face; neck included
Subject Design	between; 3 conditions= "No-indication"/"Sex-indicated" (Sex indicated)/ "Sex-reversed" (opposite sex indicated)
Method	Determine which parent the child resembles
Method Details	1 child & both real parents; projected for 15s, then 15s interval
Results	Only for girls significant difference between sex-indication condition vs sex-reversed condition when assessing resemblance to father, in that sex-reversed decreased the association with father

Oda, R., Matsumoto-Oda, A., & Kurashima, O. (2005). Effects of belief in genetic relatedness on resemblance judgments by Japanese raters. Evolution and Human Behavior, 26(5), 441-450.	**Reference**
Parent child resemblance and labels, both labels and relatedness independently from each other influence similarity ratings	**Resemblance detection**
60	**Participant number**
mixed (half/half)	**Participant sex**
18-25 years (median age: 20y)	**Participant age**
84 photos, 28 children and their 56 parents	**Total Stimuli Number**
Japanese	**Stimuli ethnicity**
mixed (half/half)	**Stimuli sex**
Children 4-5 years old, parents unknown	**Stimuli age**
parent-child	**Stimuli relatedness**
black and white photos taken in school of children and parents by experimenter, just inside of face, no ears, hair; shaded around the face; neck included	**Stimuli**
between, 2 counterbalanced versions so that each shown pair in one group labelled as related and in other as unrelated	**Subject Design**
Similarity Rating 1-7	**Method**
124 Parent-child pairs, 2 labels: related or unrelated	**Method Details**
Significant effect of relatedness and label on resemblance ratings, with the effect size being larger for label (d=0.74) than genetic relatedness (d=0.47)	**Results**

Reference	Oda, R., Matsumoto-Oda, A., & Kurashima, O. (2005). Effects of belief in genetic relatedness on resemblance judgments by Japanese raters. Evolution and Human Behavior, 26(5), 441-450.
Resemblance detection	Normal parent child resemblance
Participant number	60
Participant sex	mixed (half/half)
Participant age	18-26y (median:20)
Total Stimuli Number	84 photos, 28 children and their 56 parents
Stimuli ethnicity	Japanese
Stimuli sex	mixed (half/half)
Stimuli age	Children 4-5 years old, parents unknown
Stimuli relatedness	parent-child
Stimuli	black and white photos taken in school of children and parents
Subject Design	within, no labels
Method	Similarity Rating 1-7
Method Details	124 Parent-child pairs
Results	Significant effect relatedness

Porter, R. H., Cernoch, J. M., & Balogh, R. D. (1984). Recognition of neonates by facial-visual characteristics. Pediatrics, 74(4), 501-504.	**Reference**
Yes, 36 out of 100 correct (25 chance level)	**Resemblance detection**
50	**Participant number**
no info	**Participant sex**
21-69y	**Participant age**
46	**Total Stimuli Number**
no info	**Stimuli ethnicity**
no info infants; mother	**Stimuli sex**
13h-39h neonates (mean=23.9h); mothers 17-29y (median=23.5)	**Stimuli age**
mother-child	**Stimuli relatedness**
colour photographs, full face view	**Stimuli**
not clear whether see all stimuli or only subset; I think only saw 1 set of pictures each in 2 conditions, so 2 sets in total?	**Subject Design**
Match mother-child pair	**Method**
1 neonate and 4 possible mothers (1real); 1 mother and 4 possible children (1real)	**Method Details**
Detection significantly above chance= 17/50 subjects correctly matched Infant to mother; 19/50 subjects correctly matched mother to infant= so total of 36/100 matches right	**Results**

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., Faurie, C., & Raymond, M. (2007). Differential facial resemblance of young children to their parents: Who do children look like more? *Evolution and Human Behavior*, *28*(2), 135–144. https://doi.org/10.1016/j.evolhumbehav.2006.08.008

Alvergne, A., Faurie, C., & Raymond, M. (2010). Are parents' perceptions of offspring facial resemblance consistent with actual resemblance? Effects on parental investment. *Evolution and Human Behavior*, *31*(1), 7–15. https://doi.org/10.1016/j.evolhumbehav.2009.09.002

Alvergne, A., Huchard, E., Caillaud, D., Charpentier, M. J., Setchell, J. M., Ruppli, C., ... & Raymond, M. (2009). Human ability to recognize kin visually within primates. *International Journal of Primatology*, *30*(1), 199-210.

Alvergne, A., Oda, R., Faurie, C., Matsumoto-Oda, A., Durand, V., & Raymond, M. (2009). Cross-cultural perceptions of facial resemblance between kin. *Journal of Vision*, *9*(6), 23–23. https://doi.org/10.1167/9.6.23

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

Arantes, J., & Berg, M. E. (2012). Kinship recognition by unrelated observers depends on implicit and explicit cognition. *Evolutionary Psychology*, *10*(2), 147470491201000204. https://doi.org/10.1177/147470491201000204

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

Bateson, P. (1983). *Mate Choice*. Cambridge University Press.

Bittles, A. H. (2001). Consanguinity and its relevance to clinical genetics. *Clinical Genetics*, *60*(2), 89–98. https://doi.org/10.1034/j.1399-0004.2001.600201.x

Bower, S., Suomi, S. J., & Paukner, A. (2012). Evidence for kinship information contained in the rhesus macaque (Macaca mulatta) face. *Journal of Comparative Psychology*, *126*(3), 318.

Brédart, S., & French, R. M. (1999). Do babies resemble their fathers more than their mothers? A failure to replicate christenfeld and hill (1995). *Evolution and Human Behavior*, *20*(2), 129–135. https://doi.org/10.1016/S1090-5138(98)00047-6

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., & Dal Pos, S. (2012). Fathers see stronger family resemblances than non-fathers in unrelated children's faces. *Archives of Sexual Behavior*, *41*(6), 1423–1430. https://doi.org/10.1007/s10508-012-9983-x

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

Bressan, P., & Kramer, P. (2015). Human kin detection. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(3), 299–311. https://doi.org/10.1002/wcs.1347

Bressan, P., Bertamini, M., Nalli, A., & Zanutto, A. (2009). Men do not have a stronger preference than women for self-resemblant child faces. *Archives of Sexual Behavior*, *38*(5), 657–664. https://doi.org/10.1007/s10508-008-9350-0

Butler, S., Gilchrist, I., Burt, D., Perrett, D., Jones, E., & Harvey, M. (2005). Are the perceptual biases found in chimeric face processing reflected in eye-movement patterns? *Neuropsychologia*, *43*(1), 52–59. doi: 10.1016/j.neuropsychologia.2004.06.005

Chapais, B., & Berman, C. M. (2004). *Kinship and behavior in primates*. Oxford University Press on Demand.

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

Charpentier, M. J., Harté, M., Ngoubangoye, B., Herbert, A., & Kappeler, P. M. (2017). Visual Discrimination of Kin in Mandrills. *Ethology*, *123*(3), 251-259.

Christenfeld, N., & Hill, E. A. (1995). Whose baby are you? *Nature*, *378*(6558), 669–669. https://doi.org/10.1038/378669a0

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

Coolican, J., Eskes, G. A., McMullen, P. A., & Lecky, E. (2008). Perceptual biases in processing facial identity and emotion. *Brain and Cognition*, *66*(2), 176–187. doi: 10.1016/j.bandc.2007.07.001

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

Dal Martello, M. F., & Maloney, L. T. (2010). Lateralization of kin recognition signals in the human face. *Journal of Vision*, *10*, 1–10. doi: 10.1167/10.8.9

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), 1–11. https://doi.org/10.1167/15.13.5

Daly, M., & Wilson, M. I. (1982). Whom are newborn babies said to resemble? *Ethology and Sociobiology*, *3*(2), 69–78. https://doi.org/10.1016/0162-3095(82)90002-4

Daly, M., & Wilson, M. I. (1996). Violence against stepchildren. *Current Directions in Psychological Science*, *5*(3), 77–80. https://doi.org/10.1111/1467-8721.ep10772793

Dasser, V. (1987). Slides of group members as representations of the real animals (*Macaca fascicularis*). *Ethology*, *76*, 65–73.

DeBruine, L. M. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society of London B: Biological Sciences*, *269*(1498), 1307–1312. doi: 10.1098/rspb.2002.2034

DeBruine, L. M. (2004). Facial resemblance increases the attractiveness of same–sex faces more than other–sex faces. *Proceedings of the Royal Society of London B: Biological Sciences*, *271*(1552), 2085–2090. https://doi.org/10.1098/rspb.2004.2824

DeBruine, L. M. (2004). Resemblance to self increases the appeal of child faces to both men and women. *Evolution and Human Behavior*, *25*(3), 142–154. https://doi.org/10.1016/j.evolhumbehav.2004.03.003

DeBruine, L. M. (2005). Trustworthy but not lust-worthy: Context-specific effects of facial resemblance. *Proceedings of the Royal Society of London B: Biological Sciences*, *272*(1566), 919–922. https://doi.org/10.1098/rspb.2004.3003

DeBruine, L. M., Jones, B. C., Little, A. C., & Perrett, D. I. (2008). Social perception of facial resemblance in humans. *Archives of Sexual Behavior*, *37*(1), 64–77. doi: 10.1007/s10508-007-9266-0

DeBruine, L. M., Jones, B. C., Watkins, C. D., Roberts, S. C., Little, A. C., Smith, F. G., & Quist, M. C. (2011). Opposite-sex siblings decrease attraction, but not prosocial attributions, to self-resembling opposite-sex faces. *Proceedings of the National Academy of Sciences*, *108*(28), 11710–11714. https://doi.org/10.1073/pnas.1105919108

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

Dibeklioglu, H., Ali Salah, A., & Gevers, T. (2013). Like father, like son: Facial expression dynamics for kinship verification. In *Proceedings of the ieee international conference on computer vision* (pp. 1497–1504). doi: 10.1109/ICCV.2013.189

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

Dufour, V., Coleman, M., Campbell, O., Petit, O., & Pascalis, O. (2004). On the species-specificity of face recognition in human adults. *Current Psychology of Cognition*, *22*, 315–333.

Dufour, V., Pascalis, O., & Petit, O. (2006). Face processing limitation to own species in primates: A comparative study in brown capuchins, Tonkean macaques and humans. *Behavioural Processes*, *73*, 107–113. doi:10.1016/j.beproc.2006.04.006

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

French, R. M., Brédart, S., Huart, J., & Labiouse, C. (2000). The resemblance of one-year-old infants to their fathers: Refuting christenfeld & hill. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 22).

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., & others. (1966). *Signal detection theory and psychophysics* (Vol. 1). Wiley New York.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. i. *Journal of Theoretical Biology*, *7*(1), 1–16. doi: 10.1016/0022-5193(64)90039-6

Hepper, P. G. (2005). *Kin recognition*. Cambridge University Press.

http://www.di4d.com/

http://www.meshlab.net/

https://www.blender.org/

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

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*, 1–19. doi:10.1016/0010-0277(91)90045-6

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

Kaminski, G., Berger, C., Jolly, C., & Mazens, K. (2013). Children's consideration of relevant and non-relevant facial features in kinship detection. *LAnnee Psychologique*, *113*(3), 321–334. https://doi.org/10.4074/S0003503313003011

Kaminski, G., Dridi, S., Graff, C., & Gentaz, E. (2009). Human ability to detect kinship in strangers' faces: Effects of the degree of relatedness. *Proceedings of the Royal Society of London B: Biological Sciences*, *276*(1670), 3193–3200. https://doi.org/10.1098/rspb.2009.0677

Kaminski, G., Gentaz, E., & Mazens, K. (2012). Development of children's ability to detect kinship through facial resemblance. *Animal Cognition*, *15*(3), 421–427. https://doi.org/10.1007/s10071-011-0461-y

Kaminski, G., Méary, D., Mermillod, M., & Gentaz, E. (2010). Perceptual factors affecting the ability to assess facial resemblance between parents and newborns in humans. *Perception*, *39*(6), 807–818. https://doi.org/10.1068/p6372

Kaminski, G., Ravary, F., Graff, C., & Gentaz, E. (2010). Firstborns' disadvantage in kinship detection. *Psychological Science*, *21*(12), 1746–1750. https://doi.org/10.1177/0956797610388045

Kazem, A. J., & Widdig, A. (2013). Visual phenotype matching: cues to paternity are present in rhesus macaque faces. *PLoS One*, *8*(2), e55846.

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

Kohn, L. (1991). The role of genetics in craniofacial morphology and growth. *Annual Review of Anthropology*, *20*(1), 261–278.

Krupp, D. B., DeBruine, L. M., & Barclay, P. (2008). A cue of kinship promotes cooperation for the public good. *Evolution and Human Behavior*, *29*(1), 49–55. doi: 10.1016/j.evolhumbehav.2007.08.002

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

Langlois, J. H., Roggman, L. A., Casey, R. J., Ritter, J. M., Rieser-Danner, L. A., & Jenkins, V. Y. (1987). Infant preferences for attractive faces: Rudiments of a stereotype?. *Developmental psychology*, *23*(3), 363.

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

Lewis, D. M. (2011). The sibling uncertainty hypothesis: Facial resemblance as a sibling recognition cue. *Personality and Individual Differences*, *51*(8), 969–974. doi: 10.1016/j.paid.2011.08.002

Lieberman, D., Tooby, J., & Cosmides, L. (2003). Does morality have a biological basis? An empirical test of the factors governing moral sentiments relating to incest. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(1517), 819–826. https://doi.org/10.1098/rspb.2002.2290

Lieberman, D., Tooby, J., & Cosmides, L. (2007). The architecture of human kin detection. *Nature*, *445*(7129), 727. doi: 10.1038/nature05510

Lindsay, D. S., Jack, P. C., & Christian, M. A. (1991). Other-race face perception. *The Journal of Applied Psychology*, *76*, 587–589. doi:10.1037/0021-9010.76.4.587

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

Marlowe, F. (1999). Male care and mating effort among hadza foragers. *Behavioral Ecology and Sociobiology*, *46*(1), 57–64. https://doi.org/10.1007/s002650050592

Marsh, A. A., Elfenbein, H. A., & Ambady, N. (2003). Nonverbal “accents” cultural differences in facial expressions of emotion. *Psychological Science*, *14*(4), 373–376. doi: 10.1111/1467-9280.24461

Marsh, A. A., Elfenbein, H. A., & Ambady, N. (2007). Separated by a common language: Nonverbal accents and cultural stereotypes about americans and australians. *Journal of Cross-Cultural Psychology*, *38*(3), 284–301. doi: 10.1177/0022022107300275

McLain, D. K., Setters, D., Moulton, M. P., & Pratt, A. E. (2000). Ascription of resemblance of newborns by parents and nonrelatives. *Evolution and Human Behavior*, *21*(1), 11–23. https://doi.org/10.1016/S1090-5138(99)00029-X

Megreya, A. M., & Havard, C. (2011). Left face matching bias: Right hemisphere dominance or scanning habits? *Laterality*, *16*(1), 75–92. doi: 10.1080/13576500903213755

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

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition*, *22*, 208–224.

Oda, R., Matsumoto-Oda, A., & Kurashima, O. (2002). Facial resemblance of japanese children to their parents. *Journal of Ethology*, *20*(2), 81–85. https://doi.org/10.1007/s10164-002-0057-8

Parr, L. A., & de Waal, F. B. M. (1999). Visual kin recognition. *Nature*, *399*, 647–648. doi:10.1038/21345

Parr, L. A., Heintz, M., Lonsdorf, E., & Wroblewski, E. (2010). Visual kin recognition in nonhuman primates:(Pan troglodytes and Macaca mulatta): inbreeding avoidance or male distinctiveness?. *Journal of Comparative Psychology*, *124*(4), 343.

Pascalis, O., & Bachevalier, J. (1998). Face recognition in primates: A cross-species study. *Behavioural Processes*, *43*, 87–96. doi:10.1016/S0376-6357(97)00090-9

Penn, D. J., & Frommen, J. G. (2010). Kin recognition: An overview of conceptual issues, mechanisms and evolutionary theory. In *Animal behaviour: Evolution and mechanisms* (pp. 55–85). Springer. https://doi.org/10.1007/978-3-642-02624-9_3

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

Pfefferle, D., Kazem, A. J., Brockhausen, R. R., Ruiz-Lambides, A. V., & Widdig, A. (2014). Monkeys spontaneously discriminate their unfamiliar paternal kin under natural conditions using facial cues. *Current Biology*, *24*(15), 1806-1810.

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

Platek, S. M., Critton, S. R., Burch, R. L., Frederick, D. A., Myers, T. E., & Gallup Jr, G. G. (2003). How much paternal resemblance is enough? Sex differences in hypothetical investment decisions but not in the detection of resemblance. *Evolution and Human Behavior*, *24*(2), 81–87. https://doi.org/10.1016/S1090-5138(02)00117-4

Platek, S. M., Raines, D. M., Gallup Jr, G. G., Mohamed, F. B., Thomson, J. W., Myers, T. E., … others. (2004). Reactions to children's faces: Males are more affected by resemblance than females are, and so are their brains. *Evolution and Human Behavior*, *25*(6), 394–405. https://doi.org/10.1016/j.evolhumbehav.2004.08.007

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/

Regalski, J. M., & Gaulin, S. J. (1993). Whom are mexican infants said to resemble? Monitoring and fostering paternal confidence in the yucatan. *Ethology and Sociobiology*, *14*(2), 97–113. https://doi.org/10.1016/0162-3095(93)90010-F

Rhodes, G. (1985). Lateralized processes in face recognition. *British Journal of Psychology*, *76*(2), 249–271. doi: 10.1111/j.2044-8295.1985.tb01949.x

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, *16*, 440–444.

Schell, A., Rieck, K., Schell, K., Hammerschmidt, K., & Fischer, J. (2011). Adult but not juvenile Barbary macaques spontaneously recognize group members from pictures. *Animal cognition*, *14*(4), 503-509.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*(5), 402–409. doi: 10.1111/1467-9280.00472

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

Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 394–398. doi:10.1073/pnas.0706079105

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

Tanaka, M. (2003). Visual preference by chimpanzees (*Pan troglodytes*) for photos of primates measured by a free choice-order task: Implication for influence of social experience. *Primates*, *44*, 157–165.

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

Tomonaga, M. (1994). How laboratory-raised Japanese monkeys (*Macaca fuscata*) perceive rotated photographs of monkeys: Evidence for an inversion effect in face perception. *Primates*, *35*, 155–165. doi:10.1007/BF02382051

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). https://doi.org/10.1038/srep45885

Vokey, J. R., Rendall, D., Tangen, J. M., Parr, L. A., & Waal, F. de. (2004). Visual kin recognition and family resemblance in chimpanzees (pan troglodytes). *Journal of Comparative Psychology*, *118*(2), 194. https://doi.org/10.1037/0735-7036.118.2.194

Volk, A. A., & Quinsey, V. L. (2007). Parental investment and resemblance: Replications, refinements, and revisions. *Evolutionary Psychology*, *5*(1), 147470490700500. https://doi.org/10.1177/147470490700500101

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 of London B: Biological Sciences*, *270*(Suppl 2), S144-S146.

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

Widdig, A., Nürnberg, P., Krawczak, M., Streich, W. J., & Bercovitch, F. (2002). Affiliation and aggression among adult female rhesus macaques: a genetic analysis of paternal cohorts. *Behaviour*, *139*(2), 371-391.

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

Wittig, R. M., Crockford, C., Seyfarth, R. M., & Cheney, D. L. (2007). Vocal alliances in Chacma baboons (*Papio hamadryas ursinus*). *Behavioral Ecology and Sociobiology, 61*, 6. doi:10.1007/s00265-006-0319-5

Zhu, G., Evans, D. M., Duffy, D. L., Montgomery, G. W., Medland, S. E., Gillespie, N. A., … al. (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