PERCEIVED FACIAL SIMILARITY: RETHINKING ITS USE IN CONCEPTUALISING ROMANTIC COUPLES

Rachel Cassar

Submitted in fulfilment of the requirements for the Degree of PhD

School of Psychology

College of Medical, Veterinary and Life Sciences

University of Glasgow

Date of Submission: April 2022

Acknowledgments

My PhD journey would not have been the same were it not for some very important individuals to whom I owe my heartfelt gratitude. First and foremost, my supervisor, Professor Lisa DeBruine. I feel very fortunate that I had Lisa as a role model and mentor as I tried to navigate my way into the world of academia. Lisa ignited my love for coding, introduced me to open science, took me to my first conference and so much more than I could have hoped for. The quality of my work has improved leaps and bounds thanks to her detailed feedback and careful observations of my analyses and writing. My second supervisor, Professor Ben Jones has also been influential throughout my PhD. Ben helped me get my foot in the door with my first publication and guided me in developing my presenting and academic writing skills. He also made academia fun and welcoming. I am especially grateful for the wonderfully inclusive and collaborative environment that Lisa and Ben fostered within the lab.

I wish to thank Dr Victor Shiramizu, whom I was introduced to only halfway through my PhD and has been my biggest cheerleader ever since. When I didn't know where to look, Victor pointed me in the right direction. From our long chats about study designs to helping me interpret results, Victor has been my go-to person and continues to inspire every day. Likewise, Dr Iris Holzleitner has gone above and beyond to get me to the finish line. Whenever I needed a push, Iris knew exactly what to say, and was always happy to share her wisdom on academia and research (and life).

My office mates, Dr Lingshan Zhang, Rebecca Lai and Sarah Alharbi have been important figures throughout, as we supported one another through our PhDs, ate cake and celebrated our individualities. I am equally grateful for our office neighbours and fellow lab members, Dr Kieran O'Shea, Dr Anthony Lee, Dr Vanessa Fasolt and Dr Jaimie Torrance, all of whom have played a valuable role in getting me to this point. Kieran has been especially helpful in guiding me through my first teaching experience.

I would also like to extend my thanks to Professor Catherine Lido, whose lectures were the reason I decided to do a PhD at all, and who included me in some of her own projects. Dr Sharon Kessler also deserves special mention for giving me an opportunity to learn more about evolutionary psychology from non-human primates, for having faith in me and valuing my perspective.

Last but not least, I would not be here without the unconditional support of my parents, sister, Dean, Ewan, Alison and Madlen.

Table of Contents

List of Figures and Tables

Abstract

'Like attracts like', yet robust hypothesis testing for facial similarity in romantic couples remains lacking. Two main explanations govern research on similarity in couples: (i) an adaptive hypothesis which predicts that similarity is a kinship cue for which preferences guide individuals to choose a mate with an optimal amount of relatedness, and (ii) a by-product hypothesis which predicts that similarity is a consequence of a familiarity effect where repeated exposure to familiar objects enhances their appeal in general. Previous studies on homogamy tend to adopt study designs that make assumptions about the proximate mechanisms driving homogamy, resulting in confounded variables of interest and overzealous support for an explanation without ruling out other possibilities. The aim of this thesis was to study how people respond to facial similarity in different contexts and to assess whether explanations for homogamy hold when tested under different study designs.

Study 1 explored how perceptions of similarity relate to couple judgments and sibling judgments on a set of facial images of couples. A binomial mixed effects model was used to assess whether perceived similarity predicts couple and sibling judgments and whether this effect differed between the two contexts. The findings from this study revealed that perceived similarity strongly predicted sibling judgments but did not predict couple judgments, suggesting a distinction in the way similarity is used to inform assessments of consanguine and affine relationships.

Study 2 replicates Study 1 while incorporating random foil pairs to additionally test how accurately people judge couples, as well as whether sibling judgments distinguish between actual couples and foil pairs. The results from this study are in agreement with Study 1 in that similarity was a better predictor of sibling judgments than couple judgments, although the latter effect was weaker than the former. Furthermore, actual couples were more likely than foils to be judged as both couples and siblings, indicating that the visual information used to make couple judgements is likely to have some overlap with the visual information used to make sibling judgements.

Study 3 examined secondary data on attributions of attractiveness and trustworthiness in two-alternative forced choice tasks on self-resembling and partner-resembling transforms paired with non-resembling controls. Analysis with a mixed effects model revealed conflicting results as well as issues with the original study design. A self or partner-resembling bias was found on both attributions of attractiveness and trustworthiness and, contrary to expectations, this bias was stronger in the attractiveness condition. Additionally,

partner-resembling faces were chosen more frequently when they were in the opposite-sex category and self-resembling faces were chosen more often when they were in the same-sex category. These findings could indicate that the self- and/or partner-resembling bias is driven by a familiarity effect, however, this could also be an artefact of the study design and thus further research is required to address this matter before making any such conclusion.

General Introduction

1.1 Selection pressures in the evolution of mate choice

Mating presents as a non-random activity governed by preferences for and the possession of attractive traits. Darwin's (1871) sexual selection theory provides a framework for selection that highlights the relevance of mate choice in the evolution of sexual ornaments. For example, studies have shown that females prefer males with exaggerated sexual ornaments over males that do not have such desirable traits or extravagant display (Andersson, 2019). Two mechanisms characterise sexual selection, that of competition for mates within sexes, typically associated with males, and that of choice of mate of the opposite sex, often ascribed to females. The proposition that female choice shapes male sexual ornaments was initially met with reservation (Andersson & Simmons, 2006; Jennions & Petrie, 1997). Much emphasis of research on sexual selection had been placed on the male role of competing with fellow suitors for successful copulation with females, but this perspective has since changed with an increasing focus on the influence of female (and male) choice.

Mate choice can evolve through direct selection of preferences for traits that generate fitness benefits and costs, but also through indirect selection of preferences for heritable traits due to genetic covariation (Kokko et al., 2002). Direct benefits, such as current health or resources, are more important in the interim and result from non-heritable traits. Indirect benefits, such as increased offspring attractiveness, derive from heritable traits and enhance life expectancy and reproductive potential, which will be enjoyed by future generations. Two key models explain how mate choice can evolve through indirect selection. Fisher's (1930) 'runaway mechanism' drew attention to female choice for its role in reinforcing desired male traits and female preferences to become genetically correlated. In other words, females who prefer certain traits will mate with males possessing these traits and, to the extent that both the preference and the trait are heritable, their offspring will have the genetic predisposition towards both the preference and the trait, linking the two. Thus, the benefits enjoyed from mating with males with preferred traits will increase with frequency of preferences, as offspring will inherit the attractive trait and retain a mating advantage. This rapid propulsion of coevolution is constrained by a trade-off between attractiveness and survival such that male traits are exaggerated to the detriment of their own fitness and consequently sustain additive genetic variance for the trait (Kokko, 2001). This model is said to reach equilibrium if there is either no variance in the trait i.e., random mating pattern, or the benefits of producing attractive offspring, i.e., 'sexy sons', outweigh the costs to fitness (Kokko et al., 2002; Pomiankowski et al., 1991).

Often depicted as a rival of the Fisherian mechanism, the ‘good genes’ model suggests that attractive traits indicate enhanced genetic quality that would further improve offspring fitness. It follows that female preference can be maintained despite high costs to fitness if the heritable benefit enhances offspring survival and subsequent reproductive success. Studies have shown that these two explanations need not be mutually exclusive, but multiple mechanisms can work together rather well in explaining the evolution of traits and preferences through mate choice (Kokko et al., 2002).

Mate preferences are more costly for females to maintain than mating randomly, as choice incurs the expenditure of energy and time required to find an adequate mate, as well as a risk of encountering predators during their search, all of which have a negative impact on fecundity (Pomiankowski et al., 1991). Costly mate preferences cultivate strategies that factor such considerations into their mate discrimination process. For example, discrepant levels of parental investment in humans denote that females typically bear the larger share of responsibility than males in raising offspring (Trivers, 1972), motivating mating strategies where conditions relevant to temporal context are factored into the strength of preferences. Thus, the costliness of preferences generates differential strategies for long-term and short-term relationships that weigh up the advantages and disadvantages of traits pertaining to each approach (Buss & Schmitt, 1993).

1.2 Mate choice and homogamy

The overarching theme of this current body of work is the study of a pattern of mate choice that favours similarity in a partner. Assortative mating describes an approach to mating that responds to similarity in a positive manner when similarity is favoured in a mate (also known as homogamy), and in a negative manner when similarity is avoided in a mate. Such non-random mating behaviour cannot be addressed without first understanding why preferences exist at all. If preferences are so costly, why wouldn’t sexually reproducing organisms just mate with the first individual they come across, much less take the time to choose a similar mate? The answer to this question lies in the value of the benefits that desired traits can provide. Being choosy by way of mate preferences can be maintained as long as benefits continue to outweigh the costs of choice i.e., a net benefit outcome. Much like males face a trade-off between fitness and attractiveness, females face a trade-off between choosiness and preference costs (Jennions & Petrie, 1997). It is conceivable then that mate preferences are adaptive such that individuals respond to desirable traits in ways that optimise fitness and fecundity. Thus, in questioning the source of homogamy, one must question whether similarity in couples is resultant of preferences that are subject to selection pressures functioning to achieve a net benefit from similarity on survival and reproductive ability.

Similarity hosts its own set of indirect costs and benefits that are indeed highly consequential to the mate selection process. Discriminating between prospective mates with similar heritable traits to oneself raises the issue of genetic compatibility, for which evolving preferences would plausibly take into account in determining whether similarity is desirable in a mate (see Section 1.2.1 for further discussion on adaptive explanations). Homogamy can occur on phenotypic traits which correlate with, and to a certain degree reflect, genotypes as well as environmental influences. Studies on phenotypic homogamy constitute a large proportion of the empirical literature, on account of such traits being observable and possible to manipulate for measurable responses to experimental conditions (see Section 1.4.1 for examples of studies on facial and physical traits). Beyond genetic and phenotypic homogamy, individuals can mate assortatively on other dimensions. Social homogamy occurs passively on traits such as social background, status and geographic surroundings (Luo, 2017). Studies on social homogamy are not as widespread and whilst there is some evidence of similarity in social backgrounds, social homogamy could not completely account for assortative mating (Watson et al., 2004; Zietsch et al., 2011). Convergence occurs when partners become more similar over the course of their relationship. Studies have found little support for convergence, finding instead that similarity between partners would have already existed at the initial stage of the relationship (Mascie-Taylor, 1989) although there is some evidence that couples grow more similar in personality over time (Little et al., 2006).

1.2.1 Adaptive explanations for homogamy

Drawing on the theories of inclusive fitness and optimal outbreeding, an adaptive explanation for homogamy purports that individuals actively seek a similar mate due to the benefits obtained from sharing a higher relatedness coefficient, a measure expressed as a fraction of the genes shared between individuals. The premise of inclusive fitness theory is that prosocial behaviour is moderated by the extent of genetic relatedness between two individuals (Hamilton, 1964). The greater the probability of genes shared, the greater the amount of altruism extended (a process known as kin selection). In other words, if a gene's phenotypic effect is to increase altruism specifically towards individuals who have a higher-than-chance probability of also having that gene, then this sets up conditions under which positive selection can occur (i.e., the prevalence of this gene in the population will increase). According to Hamilton's rule, after taking into account the direct fitness costs and benefits from sharing a proportion of alleles with a potential mate, selection for a trait should result if there is a net benefit to social partners. An individual's behaviour is therefore aimed at maximising 'inclusive fitness', rather than just their own reproductive success, by evaluating the indirect genetic effects on fitness of social partners based on genetic similarity.

In the context of mate choice, the motives for maximising inclusive fitness are opposed by the considerations set out by optimal outbreeding theory. Choosing a mate who is too genetically similar, despite potentially increasing inclusive fitness, risks high costs incurred by inbreeding depression. Increased homozygosity due to inbreeding leads to increased frequency of deleterious alleles, resulting in reduced fitness and depleted reproductive success (Charlesworth & Charlesworth, 1987). A study on inbreeding depression in a critically endangered bird, the helmeted honeyeater, showed that highly inbred birds experienced 87%-90% lower predicted lifetime reproductive success than lesser inbred birds (Harrisson et al., 2019). In contrast, choosing a mate who is too dissimilar runs the risk of outbreeding depression. While initially parental divergence may enhance fitness, higher levels of divergence have a negative impact on fitness due to mechanisms such as heterozygote disadvantage. Evidence for outbreeding depression is less widespread than inbreeding depression, yet reduced mating success has been shown in plants (Barrett & Harder, 1996) and other organisms including haplodiploid ambrosia beetles (Peer & Taborsky, 2005).

The rationale behind optimal outbreeding is that a balanced degree of genetic relatedness between partners is pertinent to minimising the fitness costs associated with excessive inbreeding and outbreeding (Bateson, 1983). Inbreeding avoidance mechanisms are thus said to intervene to make appropriate the level of genetic relatedness in mate choice while still maximising inclusive fitness (Westermarck, 1921). The principles and benefits accompanying inclusive fitness and optimal outbreeding provide an incentive for individuals to favour somewhat genetically similar over dissimilar mates (Thiessen & Gregg, 1980). An adaptive account for homogamy thus predicts that perceived similarity acts as a cue of kinship that enables individuals to seek optimally genetically similar mates.

Preferences for similarity could be acquired through sexual imprinting, an adaptive learning mechanism which, its restricted definition affirms, occurs during the critical period of development and cannot be reversed (Lorenz, 1937). Through experience with family members, typically the opposite-sex parent or caregiver, individuals learn the characteristics of their kin and form a mental template of an appropriate sexual partner. When sexual behaviour develops, the mental template serves as the criterion against which individuals will discriminate between potential mates. Sexual imprinting offers an adaptive solution to species recognition, imperative for distinguishing between predators and non-threatening species. Thus, any resulting homogamy is a non-adaptive by-product of this adaptation. Although later studies have contested some of the stringent theoretical assumptions underlying Lorenz's work (Bateson, 1966, 1978), including that preferences can be modified following the critical development period, his rationale for the process of imprinting-like mechanisms remains highly influential and continues to be studied in cross-fostering experiments on animals (Hess, 1959; Sanchez-Andrade & Kendrick, 2009; Slagsvold et al., 2002) as well as in behavioural and preference tests on humans (Bereczkei et al., 2002, 2004; Rantala & Marcinkowska, 2011; Vakirtzis, 2011; Zietsch et al., 2011). The extent of sexual imprinting has also been shown to be moderated by childhood relationships with parents (Tramm & Servedio, 2008; Wiszewska et al., 2007). Marcinkowska and Rantala (2012) adopted a looser definition of sexual imprinting, that experiences following adolescence could alter impressions formed in early childhood. Their study tested the sexual imprinting hypothesis using perceived similarity ratings on faces of spouses and their opposite-sex parents, finding that mothers and daughters-in-law were perceived as significantly more similar than mothers and controls, but this effect was not observed between fathers and sons-in-law. The authors concluded that whilst the effects found could reflect that positive sexual imprinting is at play, they could not rule out the possibility that preferences for similarity are inherited.

For an adaptive mechanism to be considered plausible, however, a means for detecting genetic relatedness in non-kin is required to be in place. Phenotype matching is a widely hypothesised kin detection mechanism that involves matching traits that are perceived as similar. Studies on facial similarity have indicated that phenotype matching may contribute to the detection of kin across various contexts (DeBruine et al., 2008; Maloney & Dal Martello, 2006; Zietsch et al., 2011). Critics of an adaptive account for homogamy have pointed to theoretical difficulties with a phenotype matching mechanism, adding that to accurately match phenotypes in unrelated individuals requires that genetic similarity and phenotypic similarity are highly correlated (Bereczkei et al., 2002). Moreover, the authors highlight that a detection mechanism for genetic similarity has not yet been demonstrated empirically, owing to the difficulties associated with testing such an innate mechanism in humans.

1.2.2 Non-adaptive explanations for homogamy

A case for non-adaptive explanations for similarity in couples has also gained traction among mate choice researchers. A by-product hypothesis for homogamy predicts that similarity is not necessarily preferred for any such fitness-enhancing or beneficial properties as an adaptive explanation would. Instead, similarity between romantic partners ensues as an auxiliary by-product of another mechanism. Zajonc (1968) describes effects of increased affinity towards objects that individuals have repeated or mere exposure to (Bornstein, 1989). It follows then that similarity in couples could be due to simply liking what is familiar, thus choosing a partner who resembles oneself or one's relatives would be due to having been repeatedly exposed to and developed a preference for familiar phenotypes. Under this by-product hypothesis of couple similarity, perceived similarity does not serve as a cue for genetic relatedness but rather is an artefact of the familiarization effect.

Hinsz (1989) tested three possible explanations for facial resemblance in couples: repeated exposure, environmental co-existence, and perceptual bias (that people are better at discriminating within their own ingroup). To test a perceptual bias hypothesis, participants were divided into two groups based on age (older and younger) and rated facial images of engaged and married couples for similarity on a 9-point scale. Results revealed no effects that reached significance other than that similarity differed between actual and random pairs for both engaged and married couples. That relationship length had no effect on similarity i.e., that there was no difference in the similarity ratings between engaged and married couples, suggested that couples do not grow more similar over the course of the relationship, ruling out environmental co-existence. As participant age group did not interact with relationship length, there was no evidence that participants were less effective at discriminating faces of a different age bracket to their own, thus the perceptual bias hypothesis was also ruled out. The author concluded that the results provide strong support for a facial resemblance phenomenon and that repeated exposure is a viable explanation for the occurrence of similarity in couples due to the similarity effect on both engaged and married couples. The problem with this interpretation is that it reduces a repeated exposure explanation to the inverse of an environmental co-existence explanation by basing predictions on whether similarity ratings differ with relationship length.

Stronger evidence for a repeated exposure hypothesis has since emerged with the use of measures of perceived averageness. Perceptions of averageness are associated with prototype formation, mental models created from repeated experience with faces of a certain type, thus improving individuals' ability to discriminate those patterns of faces (Halberstadt & Rhodes, 2000; Langlois et al., 1987). Faces perceived to be highly average are indicative of prototype formation and consequently, repeated exposure with such faces. DeBruine (2004) tested judges' responses to attractiveness and averageness on a set of self-resembling transforms and average composite faces. A single-prototype hypothesis predicts that visual experience with one's own face affects perceptions of averageness and attractiveness of both same-sex and opposite-sex self-resembling faces equally. A two-prototype hypothesis predicts that visual experience with one's own face generates sex-specific mental prototypes such that averageness and attractiveness perceptions of same-sex and opposite-sex self-resembling faces are influenced by self-resemblance differentially. The study revealed that whilst self-resemblance increased attractiveness for same-sex faces more than other-sex faces, average composites were judged as more average than self-resembling transforms for same-sex and other-sex faces equally. The latter results support a single-prototype hypothesis that repeated exposure should enhance attractiveness of same-sex and other-sex faces to the same degree. The differential effect of self-resemblance on attractiveness of same-sex versus opposite-sex faces is incongruent with a single-prototype hypothesis of repeated exposure, however. Thus, while not ruling out a mere exposure hypothesis entirely, the evidence indicates that other mechanisms must be driving the differing perceptions of attractiveness in same-sex and opposite-sex self-resembling faces.

1.3 Review of studies testing theoretical explanations for homogamy

Assortative mating can theoretically occur on any number of phenotypic traits available to perceivers. Studies on similarity in mate choice have typically focused on a single trait on which similarity is measured (Watson et al., 2004). Moreover, the strength of correlations can differ considerably across traits (Horwitz & Keller, 2022). Strong couple correlations are typically observed on demographics such as age (Buss, 1984; Schwartz & Graf, 2009) and to a slightly lesser extent, attitudes (Feng & Baker, 1994; Zietsch et al., 2011) on matters such as politics and religion (Luo, 2017). Moderate correlations have been reported on cognitive ability and intelligence (Mascie-Taylor, 1989; Watson et al., 2004) and weak correlations on physical (Spuhler, 1968) and personality traits (Caspi et al., 1992; Markey & Markey, 2007). Beyond couple correlations, evidence for assortative mating through sexual imprinting, i.e. correlations between spouse and opposite-sex parent, has been demonstrated on hair and eye colour (Little et al., 2003), as well as on body odour (Jacob et al., 2002).

One prominent trait that has continued to feature in the study of perception and mate choice, as well as within this thesis, is the human face. Attraction to faces has been demonstrated from infancy and throughout childhood (Boothroyd et al., 2014) and has been shown to influence social behaviour at sexual maturity (Saxton, 2016). Faces provide salient cues for species (Marcinkowska & Rantala, 2012) and kin recognition (DeBruine et al., 2009; Maloney & Dal Martello, 2006), serving an important function in the avoidance of predators and attribution of altruistic behaviour (Krupp et al., 2012). Facial expressions further reveal important information about an individual's trait or state (Ekman & Rosenberg, 2005), while some studies have found that faces are perceived to reflect health, although whether healthy-looking faces reflect actual health is still unclear (DeBruine et al., 2010; Jones et al., 2001).

1.3.1 Visual perception and facial similarity

Faces serve as visual stimuli that can be compared against one another for assessments of perceived similarity. Non-human animals can make perceived similarity assessments to work out whether they are interacting with conspecifics (Hansen et al., 2008), while humans might use facial similarity more for identity or kin recognition purposes (DeBruine et al., 2008). In a sexual setting, recognising kin is an important skill to avoid inbreeding and its respective adverse effects on offspring fitness. Notwithstanding, consanguineous couples, especially those who are third and fourth cousins, reportedly enjoy greater reproductive success than unrelated couples (Helgason et al., 2008), although earlier reports of such advantages could be accounted for by confounding socio-economic factors (Bittles et al., 2002). Nevertheless, while closely related couples may enjoy benefits such as longer marriages or retaining wealth within the extended family, costs of inbreeding such as early death of offspring, tend to outweigh these benefits. Accordingly, people have been shown to use olfactory cues to detect and avoid potential mates with a high number of shared MHC alleles, supporting optimal outbreeding theory in guiding individuals to avoid choosing a closely related partner (Lundström et al., 2008). Genetic factors therefore play a key role in shaping perceptions of similarity and how people respond to relatedness in different social settings, including in a sexual context. The ensuing question for the study of homogamy is whether perceived facial similarity is associated with conceptualisations of couples in the same way that it is with conceptualisations of relatives.

This thesis is specifically interested in perceived similarity, which differs from measures of actual (physical) similarity, and which is more informative to understanding the proximate mechanisms that result in homogamy. Similarity can be understood in terms of a continuum of what is perceived and what exists in the physical world. A computational-representational framework of theory of mind assumes that features exist prior to similarity assessments and are later processed to form a judgment, whereas a constructivist position views similarity as only being constructed through the very cognitive process of making that similarity judgment (Shanon, 1988). Here we focus on the former perspective, assuming a metric model to describe perceived similarity. Thus, when making an assessment of perceived similarity, a pair of individuals' faces are compared through a cognitive process that takes into account various attributes relevant to making a similarity judgment on faces. The literature on face perception typically discusses faces as being encoded in so-called "face space" i.e., multidimensional scaling (MDS) models. MDS models describe individual faces as lying along a set of dimensions that each represent an attribute upon which humans discriminate faces, such as face shape, age or race (Valentine, 1991). The distance between the two points where the faces under comparison lie on this psychological space represents a measure of perceived similarity in terms of that particular feature. Critics of a metric model of face perception argue that it reduces the judgment of similarity to an over simplified assessment that does not reflect the complex manner with which humans perceive similarity (Lorusso et al., 2015). However, given the multidimensionality of faces. which cannot all be measured objectively, a machine-measured versus person-perceived definition of similarity will arguably always be subjective to a certain extent. Further, inasmuch as faces vary both within and between populations, the set of features, structure and context of the population from which the faces are drawn can largely affect the way an individual perceives similarity.

Assessments of facial similarity play an important role in the proximate mechanisms that might underpin theoretical explanations for homogamy. Of the 28 studies summarised in Appendix A (see also Section 1.4.1): seven studies obtained direct measures of perceptions of facial similarity using a scaling method i.e., ratings, five of which tested facial similarity in mate choice (Hinsz, 1989; Lorusso et al., 2011; Marcinkowska & Rantala, 2012; Thiessen et al., 1997; Wong et al., 2018), whereas the remaining two did not study mate choice per se but provided relevant evidence for facial similarity as a kin recognition tool (DeBruine et al., 2009; Maloney & Dal Martello, 2006); five studies employed a matching task where it was assumed that matches were guided by perceived facial similarity, if not instructed to do so (Bereczkei et al., 2002, 2004; Chambers et al., 1983; Griffiths & Kunz, 1973; Thiessen et al., 1997); five studies obtained measures of perceived attractiveness on self-resembling faces (DeBruine, 2002, 2004, 2005; DeBruine et al., 2011; Kocsor et al., 2011); and one measured similarity on perceptions of facial attractiveness (Little et al., 2006). Perceptions of facial similarity, if resulting from preferences for similarity, would be more relevant in the study of homogamy than objective measures of physical similarity, being assessed using facial metrics i.e., computer algorithms comparing 2D facial images or face shape in 3D images. It is not clear what such measures of physical similarity inform about homogamy and are better understood in relation to human perception, to more closely inspect whether similarity is considered in mate choice decisions. Perceptions of facial similarity depend on the human ability to visually perceive similarity and make assessments about relatedness. Thus, the visual perception literature contributes to our understanding of how perceiving facial similarity can inform mental computations of relatedness and whether the relationship between these two concepts might inform the mate choice process.

As a visual cue that is easily accessible and a relevant source of attraction (Kendrick et al., 1998), faces make ideal candidates for use in experimental research on mate choice through photographs and computational face transforms. Maloney and Dal Martello (2006) proposed a model, the threshold similarity observer (TSO), that describes similarity as a graded kin recognition symbol where observers first make a similarity evaluation which would lead to a positive kinship judgment if it surpasses a threshold value. This model was applied in a kin recognition experiment using faces of children, half of whom were siblings, where participants were assigned to either rate the face pairs for similarity on a scale of 0-10 or judge the face pairs as siblings or not. The study found that judgments of kinship increased when similarity ratings increased, yet similarity was not informed by age and gender cues, suggesting that perceived similarity in faces of children is equivalent to an estimation of genetic relatedness. DeBruine et al. (2009) conducted a replication of this study using two sets of adult face pairs, one all-female set of twins and one set of half same-sex siblings and half opposite-sex siblings. Overall, results from this study corresponded with those observed in Maloney and Dal Martello's study in that similarity is a good predictor of genetic relatedness, however face pairs varying in age and sex appeared to some degree to convey information supplemental to kinship. Thus, it was concluded that age and sex differences are more influential on similarity assessments for adult faces than child faces. Importantly, the study presented evidence suggesting that judgments of kinship and similarity are dependent on context.

The implication that context-specific criteria moderate responses to similarity suggests our understanding of similarity is more complex than first anticipated and questions the visual processing of perceived similarity asserted by a purely adaptive or by-product hypothesis. Lorusso (2011) revealed further complexities in how people compute perceived similarity by examining response time on kinship, similarity and dissimilarity judgments on face pairs varying in age and sex. A subsequent priming experiment was repeated on ten of the face pairs, assessing the conditional probability of responses to the second task given responses to the first task on the same faces. Participants took considerably longer to make kinship judgments than they did similarity or dissimilarity judgments and were quicker to respond to dissimilar face pairs than those appearing similar or related. The priming study revealed a distinction between responses to similarity and dissimilarity, where a positive judgment of dissimilarity differed from a negative judgment of similarity following kinship judgments. If similarity and dissimilarity represented two ends of a continuum, the conditional probability would have been equal. The results of this study suggest that making kinship judgments requires more visual processing than making assessments of similarity, which factors into concepts of kinship but carries different connotations to judgments of dissimilarity. Salient dissimilarities aid in eliminating the likelihood of kinship, whereas salient similarities require further processing to establish whether the similarity observed is due to relatedness. Moreover, the study concludes that given the association between kinship and similarity, as well as the additional processing involved in estimating relatedness, the former should not be reduced to the latter.

Buckingham and colleagues (2006) demonstrated how visual experience influences attributions of attractiveness and trustworthiness on masculine and feminine faces. Attributions of trustworthiness relate to prosocial behaviour relevant in kinship contexts whereas attributions of attractiveness relate to sexual behaviour relevant to mate choice (DeBruine, 2005). Facial stimuli were created by transforming male faces to resemble male and female prototypes by 50% in shape, resulting in ‘masculinised’ and ‘feminised’ versions of the faces. Following an initial task of making attractiveness or trustworthiness judgments on the faces, participants were assigned to an adaptation phase of visual exposure to either the masculinised or feminised faces. A repetition of the judgment tasks revealed that preferences for the category of faces they were exposed to in the adaptation phase had increased, suggesting that recent experience with patterns of faces may drive proximate mechanisms for perceptions of attractiveness and trustworthiness. It is suggested that distinct neural populations code different categories of faces and respond differently to congruent stimuli after visual adaptation, resulting in “face aftereffects”. Sex-contingent face aftereffects had also been demonstrated in a previous study (Little et al., 2005) that found preferences for face transforms increased post-adaptation, and that such effects were observed in the sex category of the faces participants were exposed to during adaptation.

However, DeBruine (2004) found that perceptions of averageness increased for both same-sex and opposite-sex self-resembling transforms, suggesting that visual exposure to opposite-sex kin may have contributed to averageness attributions of opposite-sex self-resembling stimuli. The results from these studies provide evidence for the influence of recent visual experience on preferences and could indicate adaptation effects on faces subject to longer durations of exposure, such as one's own or familiar faces.

1.3.2 Do people find similarity attractive?

It may be intuitive to approach the question of why similarity is prevalent in couples by evaluating whether people find self-resemblance attractive. Kocscor et al. (2011) tested preferences for self-resemblance using computerised facial morphing techniques to create participant-resembling facial stimuli. In line with methods used in previous studies using self-resembling facial stimuli (DeBruine, 2002, 2005), facial composites for each sex were first produced by separately averaging male and female facial images in shape, colour and texture. The male and female composites were then transformed to resemble each individual participant by 60%, resulting in a same-sex and opposite-sex self-resembling face for each participant. Subjects ranked the face transforms and control images on attractiveness in two conditions for each sex category, one with controlled attractiveness and one with varying attractiveness. The tableaus, for each sex category, consisted of: (i) a self-resembling face and two non-resembling faces with matching attractiveness; and (ii) a self-resembling face, a non-resembling face with matching attractiveness and a non-resembling face with higher attractiveness. Male subjects exhibited a self-resemblance bias for opposite-sex faces when attractiveness was controlled for but preferred the most attractive face when attractiveness varied. No such effects were observed when men were exposed to same-sex faces. Female subjects were indifferent towards self-resemblance in same-sex and opposite-sex faces when attractiveness was controlled for but preferred the most attractive face when attractiveness varied. The authors acknowledge a discord in the responses between participant sexes, concluding that the findings present evidence that suggest the evolution of a self-resembling bias in men. Interpretation of the women's results may point to several possible factors that could intervene with choosing self-resembling faces. The lack of a self-resemblance bias could reflect female subjects judging attractiveness in the context of short-term hypothetical relationships, which would elicit different cost and benefit considerations than would be relevant in long-term relationships. Alternatively, the nature of the study design or traits under study could have contributed to the mixed results.

DeBruine et al. (2005) explored how female preferences for self-resemblance change across the menstrual cycle. Four composite faces representing different age-sex categories (young woman, adult woman, young man, adult man) were transformed to resemble participants by 50% in shape. Participants' own transforms were paired with other participants' transforms for all age-sex categories and presented in a two-alternative forced choice task based on attractiveness. Subjects were grouped based on being in either their late follicular (days 6-14 of menstrual cycle) or luteal (days 17-27 of menstrual cycle) phase, due to the phases' estimated levels of progesterone, estrogen and conception risk. The study reported a bias for self-resemblance that was greater for women in the luteal phase than for women in the late follicular phase. This effect was qualified by an interaction with face sex, being more pronounced for women's faces than men's faces. Women's preferences for self-resemblance in female faces reported in this study correlate with progesterone levels across the menstrual cycle, signalling mechanisms for regulating prosocial behaviour as opposed to inbreeding avoidance. Similar results were obtained in another study investigating responses to self-resemblance in women with brothers (DeBruine et al., 2011). The authors reported self-resemblance biases on judgments of trustworthiness in both men and women, while judgments of attractiveness revealed a self-resemblance bias in women but not in men. This effect was qualified by an interaction with having brothers, finding that women without brothers responded with higher attributions of attractiveness in men but not trustworthiness in men. Assessing the role of similarity in mate choice is thus not a straightforward process that can be established by examining preferences for self-resemblance. Preferences for similarity alone are multifaceted and may be affected in different ways by pressures from opposing forces. Examining similarity from other perspectives, beyond preferences, can provide evidence on the processes used when people perceive similarity and how they could impact on mate choice decisions.

1.3.3 Do independent judges think couples are similar?

Various methodological approaches have been used to capture the perceptual mechanisms underlying mate choice from others' perspective. Matching tasks have been used considerably in empirical studies on homogamy, requiring participants to match faces on the basis of similarity or on the basis of likelihood of being a couple. Griffiths and Kunz (1973) sought to address whether partners appear to resemble one another because similarity exists at the initial stage of a relationship or because third parties associate similarity with couples and perceive resemblance as a result. Participants were presented with faces of individuals subcategorised by length of relationship and instructed to match the faces that appear to belong to married couples. Correct matches of couples who had been married for less than a year exceeded chance but this effect did not increase consistently with length of marriage, ruling out convergence. The authors concluded that the study provided evidence for physiognomic homogamy and that similarity was already present at the start of the relationship. At no point during the experiment were participants instructed to look out for similar traits in completing their matching tasks, nor was perceived similarity on the face pairs measured. Whilst it appears that people are generally good at matching couples correctly, the study design does not demonstrate whether or how similarity has been attended to, to arrive at those correct matches. Consequently, the study does not adequately reflect the trajectory of cognitive processing prescribed by the theoretical framework tested and is effectively equating couple judgments to similarity judgments.

In contrast, Bereczkei et al. (2002) instructed subjects to match faces on the basis of similarity without providing any information on the nature of the relationship between the individuals they were matching. To test the sexual imprinting hypothesis, two matching tasks were conducted where potential wives were matched with husbands in the first experiment, then matched with mothers-in-law in the second experiment. The husbands also completed a retrospective attachment test to incorporate a measure for the influence of parent-child relationships. Wives were correctly matched with husbands at a significantly higher frequency than with foils, yet the matching effect between wives and mothers-in law was significantly larger. The results were further supported by a significant negative effect of maternal rejection on similarity and were later replicated in a study on adoptive families (Bereczkei et al., 2004). However, while correct matches in this study reflect some measure of perceived similarity on couples' faces, correct matches alone, with no context on which to match the faces, cannot infer the mechanisms underlying similarity in couples. Marcinkowska and Rantala (2012) investigated the sexual imprinting hypothesis using perceived similarity ratings rather than a matching task and highlighted further issues with the appropriateness of study designs in testing theoretical frameworks on homogamy. Images of individuals' opposite-sex parent were presented alongside images of their partner and three age-matched controls, such that participants rated the similarity perceived between the parent and the partner and controls. Mixed results showed that mothers were indeed rated as more similar to daughters-in-law than controls, yet this effect was not observed on fathers and sons-in-law. Thus, the findings from this study were not fully consistent with those reported in Bereczkei et al. (2002, 2004) and were not interpreted as evidence of a sexual imprinting hypothesis or heritable preferences.

Similarity ratings offer an alternative and possibly more direct, less confounded means of measuring perceived similarity of face pairs compared to matching tasks. Wong et al. (2018) conducted three experiments. The first experiment obtained ratings on subjects' perceptions of similarity, age, attractiveness, and personality on a set of faces of married couples and foil couples. Couples appeared more facially similar and closer in perceived age and personality than non-couples, yet couples and non-couples were perceived as equally attractive. Furthermore, regression analysis showed that perceived age and personality differences were significant predictors of perceived similarity. In a second experiment, the same stimuli were paired with faces better-matched and worse-matched on perceived personality. A different pool of subjects was recruited to provide similarity ratings on the new set of face pairs. A repeated measures ANOVA revealed a face type effect where couples were rated as more similar than face pairs better-matched or worse-matched in perceived personality, and face pairs better-matched on personality were perceived as more similar than those worse-matched in personality. A final experiment obtained couple judgments on the new set of face pairs by ranking the pairs based on likelihood of being a couple. Consistent with the second experiment, better-matched face pairs were significantly more likely to be judged as couples. However, contrary to the first experiment, couples and better-matched pairs were equally likely to be judged as couples. Recoding the data revealed that the differential use of ranks against ratings could not account for the discrepancy in similarity ratings and couple judgments. The findings from this study show that the separate treatment of concepts of similarity and mate choice exposes how context might be an overlooked but important factor in assortative mating behaviour. Whilst the study design disclosed relevant distinctions between how people perceive similarity and couples, it afforded further testing on the relationship between the two concepts to address the assumptions underpinning theoretical explanations for homogamy.

Conflicting results on responses to similarity across context challenge the robustness of evidence supporting similarity that has previously been presented in the literature. In DeBruine's (2005) study, facial similarity was posed to kin and mate choice contexts to investigate whether perceptions of similarity function to regulate behaviour differently across context. Using male and female participant-resembling transforms, participants made two-alternative forced choice decisions for perceptions of trustworthiness, attractiveness in long-term relationships and attractiveness in short-term relationships. The number of times a self-resembling transform was chosen over another participant's transform was compared with the average number of times that image was chosen by other participants. Self-resembling faces were judged as more trustworthy but less attractive for short-term relationships, whereas self-resemblance had no effect on judgments of attractiveness for long-term relationships. The mixed results observed in this study suggest that responses to similarity are not uniformly positive across kinship and mate choice contexts, as a mere exposure hypothesis would predict, but rather are modulated by pressures specific to the context in which they are professed. These findings are in agreement with an earlier study that found self-resemblance increased attractiveness on same-sex faces to a significantly greater extent than other-sex faces (DeBruine, 2004).

1.4 Thesis scope

Explanations for assortative mating behaviour are inherently difficult to demonstrate empirically given the tacit nature of perception and mate preferences. In developing study designs and methodologies, it is crucial to deconstruct conceptualisations of romantic couples to understand how, if at all, similarity factors into the equation. The primary objective of research on assortative mating must be to verify, within reason, that similarity plays a role in the underlying processes depicted by the theoretical frameworks under study. Thus, methodologies that test individuals' ability to discriminate couples from non-couples boldly assume that an association exists between perceived similarity and what makes a romantic couple discriminable. Yet previous studies testing adaptive and non-adaptive (by-product) hypotheses often fail to distinguish between concepts of similarity and concepts of mate compatibility, concluding that their findings support a theoretical framework without clear evidence for the underlying proximate mechanisms.

1.4.1 Overview of methods used to test theoretical explanations for homogamy

A review of the empirical literature on assortative mating shows that various methodological approaches have been applied across studies that report finding evidence for homogamy. This section will discuss the main methodologies adopted by previous empirical studies on perceived similarity and homogamy (see Appendix A for further details).

Couple correlations on self-reported or survey-obtained data on single and multiple traits have dominated much of the evidence on similarity in couples. Questionnaires and surveys are easy to distribute and, especially where secondary data is readily available, can get large sample sizes. The types of traits typically measured with questionnaires from the individual's perspective include traits such as personality (Gyuris et al., 2010), attitudes (Watson et al., 2004), values (Caspi et al., 1992), demographics such as age and ethnicity (Saxton, 2016), socioeconomic variables (Zietsch et al., 2011) and physical traits such as hair and eye colour

(Little et al., 2003). More often than not, these measures are analysed using correlations and demonstrate that coefficients for couples tend to exceed chance, where the strength of association is typically consistent for the trait type across studies (Horwitz & Keller, 2022). Beyond establishing that couples' traits are positively and significantly correlated, questionnaires have been analysed with regressions and other statistical models to test for sexual imprinting, learning of parental traits or influence of childhood relationship.

Studies examining perceptions of facial similarity have frequently recruited independent observers to rate faces on a given scale (DeBruine et al., 2009; Hinsz, 1989; Lorusso et al., 2011; Maloney & Dal Martello, 2006; Marcinkowska & Rantala, 2012; Thiessen et al., 1997; Wong et al., 2018), including two studies conducted as part of this thesis (see Study 1 and Study 2). Ratings i.e., scores from a rating task, provide a sound measure of perceptions of facial similarity, yet study designs that obtain facial similarity ratings still tend to stop short of analysing how the concept of similarity fits in with mate compatibility. Adopting a method that instructs observes to rate faces has been useful for measuring perceptions of other concepts such as attractiveness (Little et al., 2006; Murstein, 1972; Perrett et al., 2002), enabling the study of homogamy from a different perspective.

Matching tasks (discussed in Section 3.3.3) have also been quite popular among homogamy researchers (Bereczkei et al., 2002, 2004; Chambers et al., 1983; Griffiths & Kunz, 1973; Thiessen et al., 1997). This method typically involves presenting individuals with a series of tableaus featuring one facial image and a group of other facial images. Participants are instructed to match the single face with one of the other faces based on some criteria, typically: most likely to be a couple, most likely to be siblings or perceived similarity. As discussed in Section 3.3.3 and in later chapters (Study 1 and Study 2), correct matches on such tasks are fairly uninformative about how or why similarity in couples is a phenomenon and infer little more than that people are capable of matching faces according to the instructions provided.

A final method discussed in this capacity is the two-alternative forced choice design, which has been employed in previous studies (DeBruine, 2004, 2005; DeBruine et al., 2011) as well as in one study in this thesis (Study 3). This methodological approach tasks individuals with indicating their preference for one of two available options, in this case, faces. For homogamy studies, the instructions provided typically aim to measure perceptions on social judgments such as by indicating which face is the most attractive or the most trustworthy, by presenting face pairs that differ on some quality that elicits dichotomised responses to the concept under study. Two-alternative forced choice tasks generate simple and clear measures of perceptions and facilitate effective comparisons of perceptions on two groups or categories of stimuli, one of which typically being the control.

1.4.2 Outline of empirical studies conducted

A growing number of studies have shown that, contrary to earlier findings supporting purely adaptive or by-product hypotheses, responses to facial similarity are context-specific, indicating that similarity serves as a kinship cue that functions to increase affiliative behaviour in kinship contexts and moderate its appeal in sexual contexts. Measures that elicit context-specific responses, such as attractiveness for sexual contexts and trustworthiness for prosocial contexts, have added to the literature by examining effects of similarity on comparatively relevant social judgments. In the face of such contradictory findings, the theoretical explanations discussed above are likely not mutually exclusive, but all contribute towards conceptualisations of and responses to similarity and mate choice.

The objective of this thesis is to explore whether adopting a different methodological and analytical approach to investigating perceptions of similarity in couples demonstrates the ineffectiveness of previous studies in testing theoretical frameworks for homogamy. This objective will be addressed by approaching the subject of homogamy in ways that treat concepts of similarity and mate compatibility as separate variables of interest, posing them to different contexts and analysing the relationship between them.

Study 1 explores how perceived similarity relates to couple judgments compared with how it relates to sibling judgments. Previous studies investigating the use of similarity in conceptualising romantic couples with matching tasks (Bereczkei et al., 2002, 2004) inadvertently confound perceptions of similarity with perceptions of couples. Subsequent evidence indicating that perceived similarity is sensitive to context (DeBruine, 2005) strengthens the need for studies to examine similarity's role in different contexts. This study addresses both of these needs by separately measuring perceptions of similarity, likelihood of being a couple and likelihood of being siblings on a set of face pairs of actual couples. Perceived similarity is measured on a scale of 0-10, whereas likelihood of being couples and likelihood of being siblings are measured with binary 'yes' or 'no' judgments. Using mixed effects modelling, Study 1 analyses whether perceived similarity predicts how likely the face pairs will be judged as couples and siblings, and whether this role differs significantly between the two contexts.

Study 2 is a replication of Study 1 but incorporates random pairs in the stimulus set. Thus, in addition to investigating how similarity influences kinship and couple judgments, it examines individuals' ability to correctly judge couples' faces as couples and assess whether this differs for sibling judgments. This additional analysis uses mixed effects models to assess whether the type of face pair i.e., actual couple or foil, determines the likelihood that the face pair will be judged as a couple and likewise as siblings, and whether this effect differs depending on judgment type. This latter analysis can be arguably likened to assessments of correct matches in matching tasks on couples' faces, while making an important point of demonstrating the ineffectiveness of this measure at informing anything specific about conceptualisations of romantic couples.

The third and final study in this thesis investigates preferences for self-resemblance in sexual versus prosocial contexts. Perceptions of attractiveness and trustworthiness have been measured in previous studies investigating responses to self-similarity in explicitly sexual and prosocial contexts (DeBruine, 2005; DeBruine et al., 2011). The findings from these studies indicate that self-resemblance increases attributions of trustworthiness more than attributions of attractiveness, while reporting a stronger self-resemblance bias in same-sex faces than in opposite-sex faces. Study 3 uses self-resembling and partner-resembling transforms paired with non-resembling transforms in a two-alternative forced choice design where participants make decisions on the face pairs for both attractiveness and trustworthiness. A statistical difference in self-resemblance biases on attractiveness and trustworthiness would strengthen the evidence for a context-specific response to similarity, whereas if an equal self-resemblance bias is observed on perceptions of both attractiveness and trustworthiness, then preferences for similarity could be due to mere exposure.

Study 1

Facial similarity predicts sibling judgments but not couple judgments

2.1 Abstract

Facial similarity has been demonstrated to act as a kinship cue that guides individuals to behave altruistically towards kin (DeBruine, 2004; DeBruine et al., 2009; Maloney & Dal Martello, 2006). Yet in mate choice contexts, the use of facial similarity has sparked considerable debate. Some argue that mate preferences have adapted to produce a fitness-enhancing response to genetically similar individuals, whilst others maintain that a non-adaptive preference for familiarity results in similarity as a by-product. The current study assessed how similarity is used in different contexts by testing whether similarity ratings predict couple judgments and sibling judgments on the same set of face pairs of actual couples. We found that whilst individuals used perceived similarity to make sibling judgments, they did not for couple judgments, shedding light on methodologies using couple matching tasks to study homogamy. The findings reported in this study provide further evidence that responses to facial similarity are context dependent.

2.2 Introduction

A widely discussed phenomenon in the study of mate choice is the prevalence of similarity in couples, otherwise known as 'homogamy' or 'positive assortative mating'. Evidence for homogamy has been reported across various physical (Chambers et al., 1983; Hinsz, 1989; Spuhler, 1968), personality (Little et al., 2006; Mascie-Taylor, 1989) and other non-physical personal traits in couples (Thiessen et al., 1997; Watson et al., 2004), suggesting that partners appear to be more similar than random pairs.

Alongside environmental influences, heritable genetics contribute a great deal towards an individual's observable traits or phenotype and effectively, family resemblance (for review see Kohn, 1991). Thus, similarity among kin is a well-recognised cue for kinship, enabling kin recognition through phenotype matching (Lacy & Sherman, 1983). One hypothesis for homogamy proposes that similarity is used as a kinship cue in mate choice contexts, where individuals choose a similar mate to maximise 'inclusive fitness', the collective reproductive potential of the individual and their genetically similar social partners (Bateson, 1983). This account predicts that selection could favour preferences that generate a response to similarity that balances the fitness benefits from sharing a higher relatedness coefficient with offspring against the costs of extreme inbreeding (e.g. inbreeding depression) and outbreeding (e.g. hybridisation). Under this adaptive hypothesis, similarity is used to guide the individual to find a partner with an optimum amount of genetic relatedness (Hamilton, 1964; Rantala & Marcinkowska, 2011).

Faces provide a clear visual cue for phenotype matching in humans and have been used extensively in experimental research on kin recognition and homogamy alike. Maloney and Dal Martello (2006) examined subjects' ability to identify siblings from a set of child face pairs and found that pairs that were rated as more similar were more likely to be judged as siblings, regardless of age or sex differences. A replication study found that age and/or sex *did* have some influence on similarity ratings in adult face pairs of twins and siblings, indicating that whilst similarity conveys kinship information, it also conveys information that is not relevant for making kinship judgments (DeBruine et al., 2009). This raises potential difficulties with an adaptive explanation for homogamy, where seeking genetic relatedness drives similarity in couples.

An alternative explanation for homogamy suggests that a mere exposure effect could be responsible for similarity observed in couples. Research has shown that repeated exposure to a stimulus enhances attitudes towards it, resulting in a general preference for familiar objects. Similarity is hereunder argued to emerge as an incidental by-product of a non-adaptive preference for familiar stimuli, such as one's own face or those of familiar others. DeBruine (2005) assessed how responses to similarity differ in attributions relevant to kin (trustworthiness) and mate-choice (attractiveness) contexts. Facial resemblance to self was found to increase trustworthiness and decrease attractiveness in short-term relationships, with no effect on attractiveness in long-term relationships. Although an additional study found that attractiveness to self-resembling faces was lower among women who had brothers than women who did not (DeBruine et al., 2011). Consequently, similarity improves attitudes in contexts where kin is favoured, such as when judging whether to trust an individual, and deteriorates attitudes where kin selection is not favourable, such as when judging attractiveness, which is reportedly more sensitive to genetic quality in short-term than long-term relationships. This contradicts the mere exposure explanation where similarity should increase preferences to familiar stimuli in any context.

The accumulating evidence supporting a context-specific explanation for homogamy brings into question existing findings supporting alternative theoretical frameworks. A common experimental approach observed in the literature on homogamy is to task subjects with correctly matching couples from a set of photographs of individuals' faces (Bereczkei et al., 2002; Chambers et al., 1983; Griffiths & Kunz, 1973). These empirical studies purportedly find evidence for homogamy at initial assortment by comparing the number of correct matches against chance, but typically fail to employ adequate controls in the study design. Studies that instruct subjects to match couples based on facial resemblance do not effectively capture whether people do use similarity to identify couples. Studies that instruct subjects to match couples on likelihood of being married do not obtain a direct measure for perceived similarity at all, merely assuming similarity underlies their cognitive processing during the task. Without closer inspection of the relationship between similarity and couple judgments or exploring how similarity is used differently in different contexts, evidence for homogamy is weak and could be confounded with any number of factors that are not controlled for.

The current study examines the role of similarity in different contexts by separately measuring perceived similarity ratings, kin relationship (sibling) judgments and sexual relationship (couple) judgments on pairs of faces from actual couples. If homogamy can be explained by the adaptive hypothesis or mere exposure to familiar faces, perceived similarity should be used to inform both sibling judgments and couple judgments. If responses to similarity are context-specific, perceived similarity will be used when making sibling judgments, but not couple judgments.

2.3 Method

2.3.1 Design

To investigate how perceived similarity is used in different contexts, the study examined whether similarity ratings predict couple judgments and sibling judgments on the same set of face pairs. Participants were randomly allocated to one of two tasks:

1. couple judgments task: judge whether pairs of faces look like couples; or
2. sibling judgments task: judge whether pairs of faces look like siblings.

The dependent variable consisted of the judgment scores for which participants provided 'Yes' or 'No' answers. The independent variables were: *similarity* (obtained from a separate study, Holzleitner et al., 2019); and *judgment type* (couple or sibling).

2.3.2 Participants

The target sample size was determined using a power simulation performed on statistical software package R (R Core Team, 2021). Sample data for the simulation was generated by extracting the random intercepts from a linear mixed effects model on a distribution with the same mean (M = 3.80) and standard deviation (SD = 2.34) as the similarity ratings. A target sample size of 20 participants for each task was determined to detect an effect (estimate = -0.24) where similarity increases the proportion of sibling judgments but not couple judgments.

A total of 169 participants were recruited online via a link that was shared on social media platforms (Facebook, Twitter and Reddit), consisting of 137 women (mean age = 33.95) and 30 men (mean age = 33.11). This total includes 47 participants (44 women and 3 men) who were recruited after the link was posted in the comments of an online article about similarity in couples. 84 participants completed the sibling judgments task and 85 completed the couple judgments task. No restrictions were placed on age, sex, sexual preference or location.

2.3.3 Stimuli

The facial images were collected as part of a previous study and included 139 couples aged between 19 – 72 years old (Holzleitner, O'Shea, et al., 2019). The 3D images were collected with a DI3D system whereby subjects are photographed from 6 different angles simultaneously. The setup for taking the images was standardised, with subjects positioned 90cm from the rig and flash units used for lighting. Make-up, glasses and jewellery were removed, and subjects were provided with a headband to keep any hair away from their face and ears. Therefore, any potential cues beyond the hairline and faces were removed from the images. Photographs were taken with a neutral expression (for further details, DeBruine & Holzleitner, 2022 and Holzleitner, DeBruine, et al., 2019).

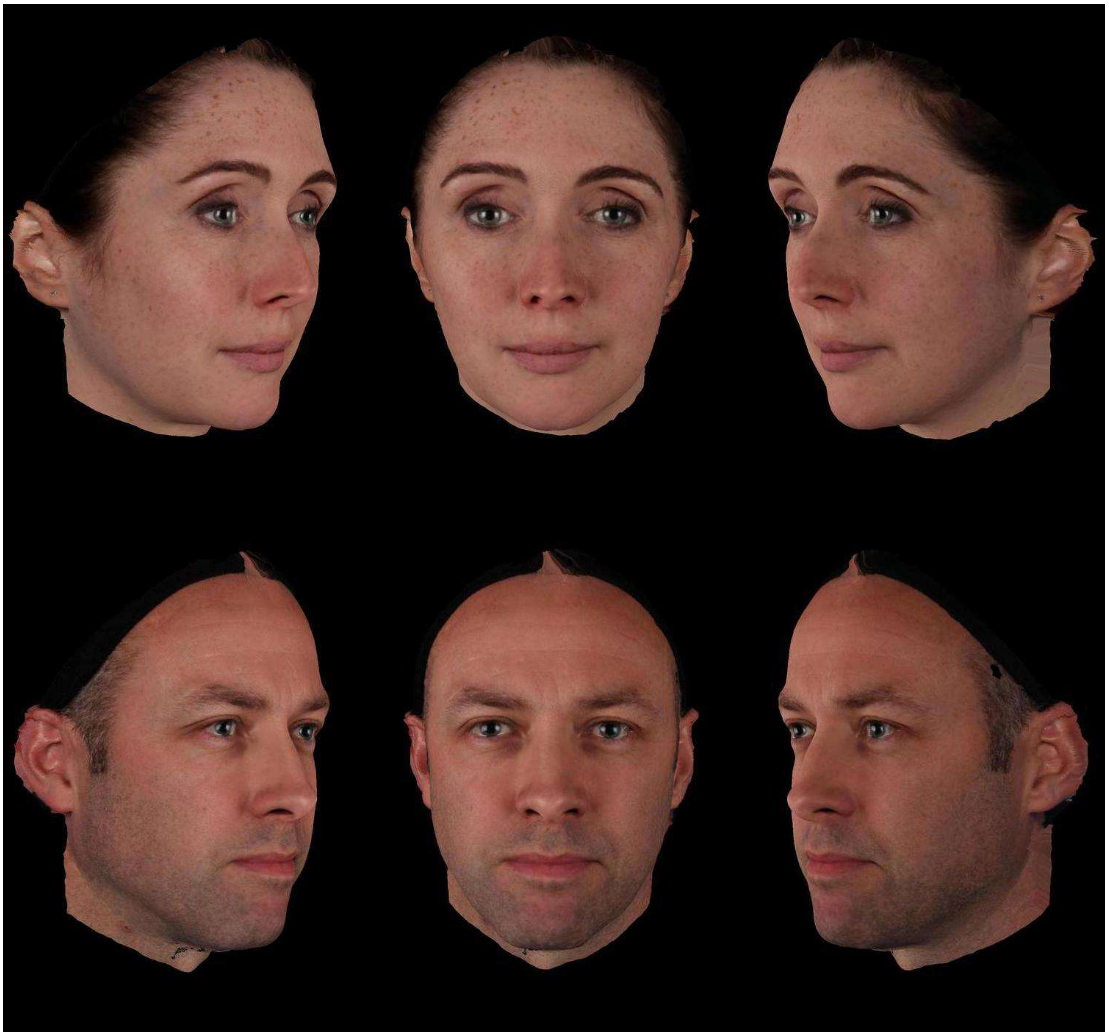


Figure 1: Example of stimuli as presented in the study. 3D facial images showed the face from three angles on a black background. Each face of an opposite-sex pair was positioned on top of one another.

The similarity ratings, ranging from 0 (not similar) to 10 (similar), were obtained from 25 participants, of which 12 were female and 13 were male and aged between 19 – 40 years old (Holzleitner, O'Shea, et al., 2019). In accordance with recommendations for the smallest sample size required to obtain reliable mean stimulus ratings (DeBruine & Jones, 2018), a minimum of 15 raters was planned. Subsequently, inter-rater reliability on the similarity ratings using Cronbach's alpha was determined to be 98%.

2.3.4 Procedure

The experiment was hosted on faceresearch.org, where participants were debriefed on what to expect, without disclosing the objective of the study. Participants were assured that their identities would remain anonymous and were presented with the option to consent to participate in the study with the freedom to leave at any point. Participants were allocated equally to either the couple task or the sibling task. Each image pair was presented on an individual page and in a random order. Participants were asked to judge the face pairs on whether they were likely to be couples (couple task) or siblings (sibling task) using the 'Yes' or 'No' buttons provided. Instructions were repeated for each pair of faces.

2.4 Results

The data was analysed using statistical software package R v3.5.0 (R Core Team, 2021), lmer (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017). All data and code are available at osf.io/uey7s/.

2.4.1 Analysis 1: All data collected

A binomial mixed effects model was used to analyse the judgment scores with the between-subjects and within-item factor of judgment task completed (effect-coded as couple = +0.5 and sibling = -0.5) and the within-subject and between-items factor of perceived similarity scores. Similarity scores were the by-item random intercepts obtained from a linear mixed effects model of similarity judgments from Holzleitner et al (2019) (see Figure 2). Random effects were specified maximally (Barr et al., 2013), with a by-subject random slope of similarity with a random intercept and a by-item random slope of judgment type with a random intercept.

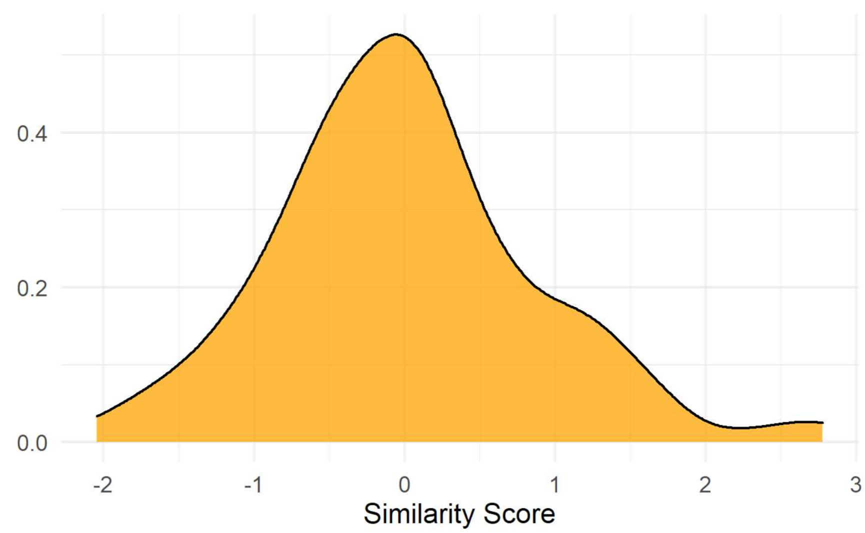


Figure 2: Density plots showing the distribution of similarity scores from 25 participants on 139 pairs of facial images of actual couples. Similarity scores were derived from the random intercepts for face pairs extracted from a linear mixed effects model with raw similarity scores (DV) and random effects for face pairs and participants.

Analysis of all data collected on sibling judgments (M = 0.45, SD = 0.50) and couple judgments (M = 0.54, SD = 0.50) showed a main effect of similarity (estimate = 0.51, se = 0.05, z = 10.3, $p < 0.001$), indicating that the likelihood of both couple and sibling judgments increased as similarity ratings increased. This main effect was qualified by an interaction with judgment type (estimate = -0.71, se = 0.08, z = - 8.82, $p < 0.001$). Separate analyses were conducted for couple and sibling judgments, finding that although similarity predicted the likelihood of being rated as siblings (estimate = 0.85, se = 0.06, z = 14.1, $p < .001$), the effect of similarity in the couple condition was weak (estimate = 0.16, se = 0.07, z = 2.38, p = .018) (see Figure 3).

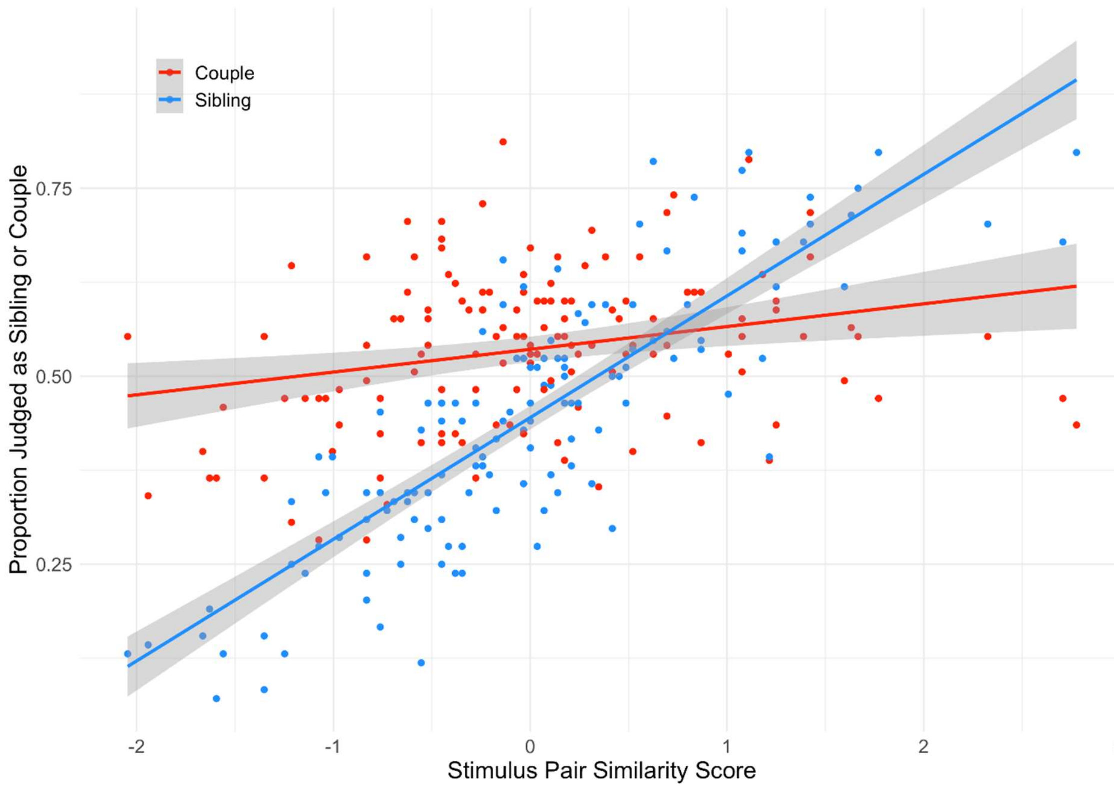


Figure 3: Scatterplot describing the relationship between similarity ratings and couple (estimate = 0.16) and sibling judgments (estimate = 0.85) on 139 face pairs. The y-axis depicts the mean scores for couple and sibling judgments. The x-axis depicts the intercepts for the similarity ratings on the images.

2.4.2 Analysis 2: Effect of couple similarity article on couple judgments

Participants who accessed the study from the link shared on the article on similarity in couples were effectively informed of the study's true interest in whether they attend to similarity to assess couple likelihood. A binomial mixed effects model examining the impact of the article on couple judgments only was specified with a by-subject random slope of similarity with a random intercept and a by-item random effect with a random intercept. This analysis showed a 2-way interaction effect of similarity and couple article status (i.e. before or after participants had seen the article) (estimate = 0.33, se = 0.11, z = 3.03, $p < 0.01$), indicating that participants did not use similarity information to make couple judgments when they did not have any prior knowledge about the subject, but did when they had been potentially exposed to the article on couple similarity (see Figure 4).

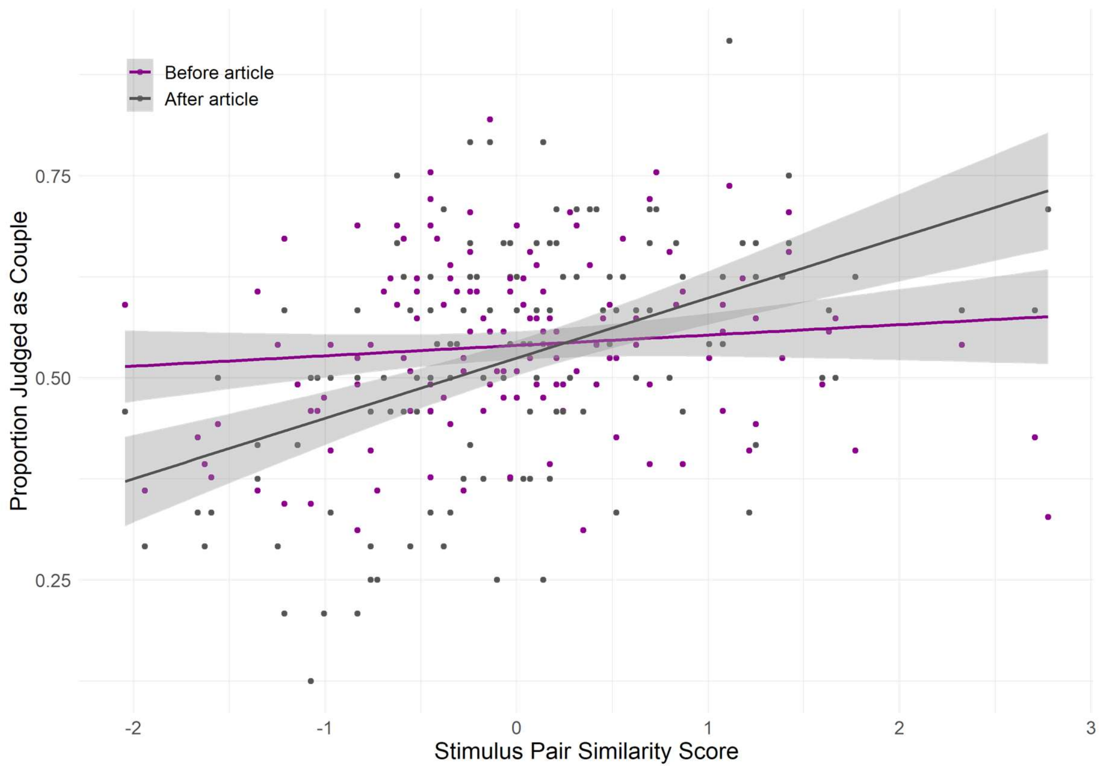


Figure 4: Scatterplot describing the relationship between similarity ratings and couple judgments on 139 face pairs before (estimate = 0.06) and after (estimate = 0.16) being exposed to an article on couple similarity. The y-axis depicts the mean scores for couple judgments. The x-axis depicts the intercepts for the similarity ratings on the images.

2.4.3 Analysis 3: Data collected before couple similarity article

A subset of the data collected, excluding all subjects who participated in the study after viewing the couple similarity article, determined by date of participation, was reanalysed using the same model as in Analysis 1. The results showed a main effect of similarity (estimate = 0.47, se = 0.05, z = 9.28, $p < 0.001$), indicating that the likelihood of both couple (M = 0.54, SD = 0.50) and sibling judgments (M = 0.47, SD = 0.50) increased as similarity ratings increased. This main effect was qualified by an interaction with judgment type (estimate = -0.80, se = 0.08, z = -9.54, $p < 0.001$). Separate analyses were conducted for couple and sibling judgments, finding that similarity predicted the likelihood of being rated as siblings (estimate = 0.86, se = 0.07, z = 13.2, $p < .001$), but not as a couple (estimate = 0.06, se = 0.07, z = 0.97, p = .33) (see Figure 5).

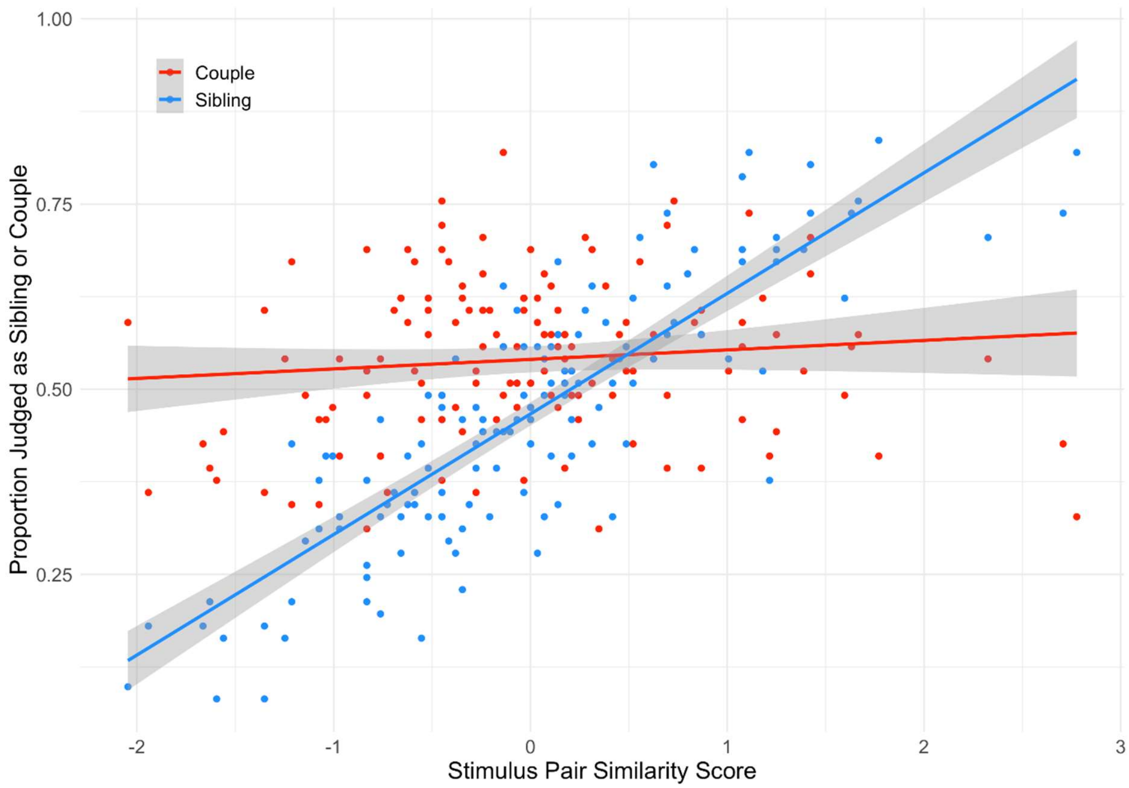


Figure 5: Scatterplot describing the relationship between similarity ratings and couple and sibling judgments on 139 face pairs before being exposed to an article on couple similarity. The y-axis depicts the mean scores for couple and sibling judgments. The x-axis depicts the intercepts for the similarity ratings on the images.

Visual examination of the similarity scores distribution gave no indication that participants were not completing the task as expected and QQ plots for residual error were as expected given the binomial structure of the data. Therefore, there was no reason to inspect the data further for abnormalities.

2.5 Discussion

The findings in this study suggest that whilst people attend to facial similarity to make sibling judgments, they do not for couple judgments. These results provide further evidence that responses to facial resemblance are context dependent, consistent with DeBruine (2002, 2004).

The contrasting effects observed under the sibling and couple condition reflect similarity's systematically different functions under kin and mate choice contexts. Perceived similarity strongly predicted sibling judgments, thus supporting its use as a cue for genetic relatedness. In a mate choice context, similarity did not predict couple judgments, at least not when participants were unaware of the subject under the study.

As explained, 47 of the 169 participants were recruited after having accessed an article about similarity in couples. This event seems to have resulted in participants using similarity to make couple judgments where they otherwise might not have, thereby confounding the couple judgments with their underlying similarity estimations. Analysis of the impact of the article on couple judgments inadvertently demonstrated the relevance of context in attending to facial similarity. When participants had no prior information about the study hypotheses, similarity predicted sibling judgments but not couple judgments. Once subjects were made aware of the nature of the study, similarity predicted both sibling and couple judgments, albeit weakly in the latter context.

The findings reported under this study suggest that similarity ratings and couple judgments are qualitatively different from one another and thus contradicts assumptions that people mainly use similarity to match couples. This corroborates with findings from a study that compared similarity ratings to the likelihood of being a couple on identical face pairs, which found that whilst pairs better matched on perceived personality were rated as more similar than actual couples, couple judgments were equally likely across the two conditions (Wong et al., 2018).

So how might an individual make judgments about the likelihood of a pair being a couple? Whilst the results reported in this study suggest that people do not use similarity to make couple judgments, we cannot comment on why people have been accurate at matching couples in previous studies (Bereczkei et al., 2002, 2004; Chambers et al., 1983; Griffiths & Kunz, 1973). They may be attending to information other than similarity to make couple judgments, such as perceived personality or complementarity of traits (Wong et al., 2018). Given the complexity in defining complementarity in physical and facial traits, this line of reasoning is difficult to test for and consequently, research on complementarity is less widespread (Štěrbová & Valentová, 2012).

One limitation of the study is that all face pairs presented were of actual couples, and thus it did not measure participants' accuracy at identifying couples. To further ascertain whether people are accurate at making couple judgments, and whether perceived similarity moderates their accuracy, further studies would need to incorporate control images of non-couples.

This study has demonstrated a difference in the use of similarity in kinship and mate choice contexts. We highlighted that correctly matching couples' faces does not necessarily reflect the use of similarity to inform couple judgments, casting doubt on the reliability of methodologies that attribute correct matches to an "all-context" response to similarity. By introducing foil couples in a follow-up study (Study 2), we will investigate how accurate people are at identifying couples and whether similarity improves or deteriorates this accuracy.

Study 2

Facial similarity, sibling judgments and couple judgments: A replication study using actual couples and random pairs

3.1 Abstract

Previous studies have shown that couples demonstrate higher-than-chance levels of similarity compared with random couples across various traits (Horwitz & Keller, 2022). Some evidence supporting theoretical explanations for homogamy arises from matching tasks, where participants are instructed to correctly match couples' faces based on perceived similarity or likelihood of being a couple (Bereczkei et al., 2002, 2004; Chambers et al., 1983; Thiessen et al., 1997). In this study, we replicate an earlier study (Study 1) obtaining similarity ratings, couple judgments and sibling judgments on a set of face stimuli of couples and random pairs. Consistent with Study 1, we show that similarity is a weak predictor of couple judgments compared with sibling judgments. Moreover, by introducing random pairs in this study, we are able to examine the accuracy of couple judgments. This additional analysis revealed that actual couples were significantly more likely than foils to be judged as couples (and as siblings). The current study presents evidence that study designs that don't explicitly test similarity in different contexts risk confounding perceived similarity with other ways of matching couples.

3.2 Introduction

'Homogamy' refers to a form of assortative mating where couples are more similar than average. Numerous studies have demonstrated correlations of varying strengths on couples' attributes (Watson et al., 2004), with stronger correlations observed on traits such as age (Buss, 1984) and attitudes (Feng & Baker, 1994; Zietsch et al., 2011); and weaker but still positive correlations reported on personality traits (Gyuris et al., 2010; Štěrbová et al., 2017).

One of the more dominant hypotheses explaining homogamy points to an adaptive mechanism where individuals use perceived similarity as a cue of relatedness when choosing a partner. This perspective claims that individuals choose similar mates to maximise inclusive fitness i.e., the reproductive success of genetic relatives (Hamilton, 1964). Inclusive fitness theory is considered alongside optimal outbreeding theory (Bateson, 1983), which maintains that an equilibrium of genetic relatedness is needed in order to maximise such fitness benefits with the fitness costs associated with extreme inbreeding.

A competing explanation for homogamy argues that similarity in a partner is not the objective behind mate choice decisions, but rather is the unintended by-product of another underlying mechanism that has a knock-on effect of similarity on mate choice. This theoretical perspective proposes that similarity in couples can result from a mere exposure effect, where repeated exposure to a stimulus increases its appeal to the observer (Zajonc, 1968). Thus, exposure to faces of family members during childhood would enhance the appeal of familiar phenotypes, increasing the likelihood that individuals would choose a mate with similar traits to their relatives. This account therefore predicts that any similarity observed between partners is a non-adaptive artefact due to the shared genetic coefficient between the individual and the family members whose traits they were exposed to.

Within the scope of assortative mating, faces provide an important avenue for homogamy research given that they are incredibly salient for identity and impression formation. Several studies test whether couples are more similar than random pairs using some form of measure for perceived similarity, being arguably more appropriate than objective measures of physical similarity at capturing the perceptual processes that shape homogamy (see Section 1.3.1). One study employed facial stimuli to investigate whether couples are perceived as similar at the initial assortment stage or converge, that is, grow to appear more similar over time (Griffiths & Kunz, 1973). Participants were presented with facial photographs of couples who had been married for various durations and were asked to match the pairs they thought looked most like a couple. Couples who were married for less than a year were correctly matched at a rate that exceeded chance, although the frequency of correct matches did not continue to increase with relationship length. The authors concluded that similarity already exists at the start of the relationship, rejecting convergence as a possible explanation for homogamy. Similar conclusions were reached from other studies (Chambers et al., 1983; Hinsz, 1989; Little et al., 2006), although some evidence of convergence in married couples also exists (Zajonc et al., 1987).

Using such photo-matching tasks to assess perceived similarity in a mate choice context, whilst not uncommon, can be methodologically problematic. In some cases, as with Griffiths and Kunz' (1973) study, judges are instructed to match faces on the basis of 'likelihood of being a couple'. Correct matches are consequently assumed to indicate that people are attending to observed similarity to form their judgments. Without directly measuring perceived similarity, however, studies of this kind omit the numerous other possible bases on which people could be making couple judgments. For example, perceivers could be attending to *complementary* features, rather than similar features, such as men who look particularly masculine being perceived as more likely to be paired with women who look particularly feminine.

In other cases of photo-matching tasks, judges are instructed to match faces on the basis of 'similarity', where correct matches are interpreted as evidence of similarity in couples, despite not investigating how the observed similarity relates to judges' concept of a couple in a mate choice context. One study adopted a photo-matching task to examine the sexual imprinting paradigm using facial photographs of men, their wives and their mothers (Bereczkei et al., 2002). Sexual imprinting offers an alternative explanation for homogamy, which predicts that individuals imprint on their caregiver during a critical period of development and form a mental template of their phenotypes to match against potential mates. This allows individuals to detect conspecifics, which can be important when it is possible to encounter individuals of closely related, but different species, with whom reproduction is impossible or offspring are sterile. For example, mules are the offspring of horses and donkeys, and are typically sterile, so horses or donkeys who learn what an appropriate mate looks like from their parent should be more reproductively successful. While humans currently don't have closely related species that they might confuse for appropriate mates, different hominid species lived in the same range through much of hominid evolution. Under this theory, closer-than-average similarity of within-species couples is a non-adaptive by-product of this adaptive process. To test this theory, judges were recruited to match husbands to wives and wives to husbands' mothers on the basis of similarity. The authors reported that the wife's photograph was chosen at a significantly higher rate than controls and moreover, that the wives were judged as more similar to husbands' mothers than to husbands, suggesting that imprinting-like mechanisms drive assortative mating. In a later study, Bereczkei and colleagues (2004) similarly found that adopted women were correctly matched to husbands at a significantly higher rate than controls ($z = -6.12$, $p < 0.001$) and this accuracy increased when matching husbands to wives' adoptive fathers ($z = -7.618$, $p < 0.001$).

Testing theoretical explanations for homogamy requires interpreting people's concepts of both perceived similarity and mate choice, to then understand how the two concepts are associated. Thus, both photo-matching experimental designs (on the bases of similarity or couple likelihood) provide an incomplete picture of the perceptual processes that could be driving homogamy, by exploring only one facet of the relationship between perceived similarity and mate choice decisions. In so doing, they fail to control for other possible factors that could contribute to mate choice decisions; factors that people might even value more highly than similarity in a mate choice context, such as complementarity. Despite reports of correct matches above chance levels, the extent to which these results demonstrate explanations for homogamy is inevitably impacted by the lack of further investigations into the role of context in perceived similarity.

The objective of the current study is to investigate whether people respond to similarity uniformly across kinship and mate choice contexts by assessing whether similarity ratings predict couple judgments and sibling judgments on the same set of facial photographs. This will inform theories of homogamy that emphasise genetic relatedness (e.g., optimal outbreeding theory) versus similarity as a by-product (e.g., sexual imprinting or mere exposure). Further, by incorporating foil pairs, the study aims to examine whether people are any more or less likely to make couple and sibling judgments on actual couples when compared with random pairs. The specific research questions that will be addressed are:

1. Are couple judgments accurate? This would be evidenced by pairs of faces being more likely to be judged as couples if they belong to actual couples than foils.
2. Are couple judgments different from sibling judgments? This would be evidenced by a difference in the effect of pair type (i.e., actual vs. foil pair) on the likelihood of making sibling judgments versus couple judgments.
3. Does perceived similarity drive sibling/couple judgments? This would be evidenced by pairs who are judged as more facially similar being more likely to be judged as siblings or couples.

4. Does perceived similarity drive couple judgments less than sibling judgments? This would be evidenced by a difference in the effect of perceived similarity on the likelihood of making sibling judgments and couple judgments.

3.3 Method

3.3.1 Design

To investigate whether perceived similarity predicts couple judgments and sibling judgments on actual couple and foil face pairs, participants were randomly allocated to one of the following tasks:

1. similarity ratings: rate the pairs of faces for similarity on a scale from 0-10;
2. couple judgments: judge whether pairs of faces look like couples; or
3. sibling judgments: judge whether pairs of faces look like siblings.

The dependent variable consisted of the judgment scores, for which participants provided 'Yes' or 'No' answers. The independent variables were: *similarity* (see Section 3.4.1), *judgment type* (couple or sibling), and *pair type* (actual couple or foil).

A randomised block design, where each participant viewed only one of 3 subsets of 90 face pairs, was used to keep the task duration to a minimum and avoid participant fatigue. Thus, participants were randomly assigned to complete one of the tasks on one of the three subsets of face pairs.

3.3.2 Participants

A power analysis was conducted by simulating the expected data using the statistical software package R (R Core Team, 2021). With half the effect sizes reported in the pilot study and a sample size of 20 participants per group of face pairs (i.e., 60 participants per judgment task), it was determined that the study would be powered at least 89% to detect the highest-order predicted interaction (code and data available at https://osf.io/rqvxp/).

A total of 182 participants, aged between 18 – 73 years old and from at least 23 different countries (mostly from the UK, Malta and the US), were recruited online via a link that was shared on social media platforms (Facebook, Twitter and Reddit). 125 participants identified as female, 55 as male, 1 as non-binary and 1 chose not to provide their gender. Participants aged under 18 years old were excluded from the study. In accordance with the power analysis of a sample size of at least 20 participants per group per task, 61 participants completed the similarity rating task, 61 participants completed the couple judgment task, and 60 participants completed the sibling judgment task. Because the study needed to be manually disabled after running analyses to determine the number of eligible participants, the actual numbers slightly exceeded the planned numbers.

3.3.3 Stimuli

The facial images were collected as part of a previous study and included 270 opposite-sex pairs (540 individuals) aged between 19 – 77 years old, half of which were actual couples and half were foil pairs (Holzleitner, O'Shea, et al., 2019). All stimuli were of white individuals, so foils were only matched on age (+/- 3 years) and sex. The 3D images were collected with a DI3D system and delineated using MorphAnalyser 2.4. Participants were instructed to remove any make-up, glasses or jewellery, and were provided with a headband to keep hair away from their face and ears. The setup for taking the images was standardised with flash lighting and seats positioned 90cm away from the rig. 3D photographs, using cameras at 6 different angles, were taken with a neutral expression and later cropped around the face and ears. Thus, hair cues beyond the hairline were omitted from the final face images which were placed on a black background (for further details, DeBruine & Holzleitner, 2022 and Holzleitner, DeBruine, et al., 2019).

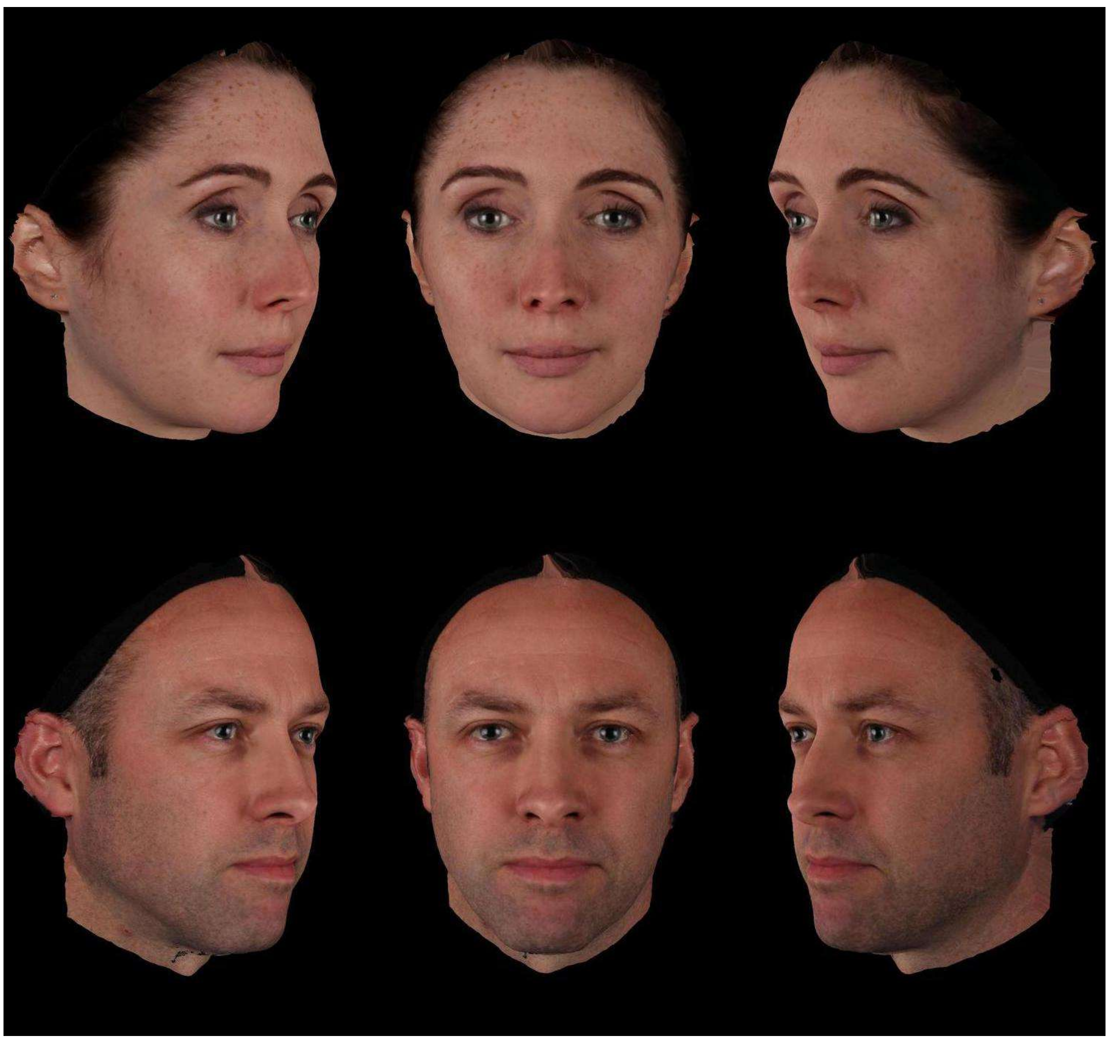


Figure 6: Example of stimuli as presented in the study. 3D facial images showed the face from three angles on a black background. Each face of an opposite-sex pair was positioned on top of one another.

3.3.4 Procedure

The experiment was hosted on Experimentum (exp.psy.gla.ac.uk) where participants were given instructions about the task that they were assigned to, without disclosing the objective of the study. Participants were assured that their identities would remain anonymous and were given the option to consent to participate in the study, with the freedom to leave at any point. Each image pair was presented on an individual page and in a random order and brief instructions were repeated on every new page.

Participants allocated to the similarity task were asked to rate the face pairs for similarity on a scale of 0 (not similar) – 10 (similar). Participants allocated to the couple or sibling tasks were asked to judge the face pairs on whether they were likely to be couples (couple task) or siblings (sibling task) using the 'Yes' or 'No' buttons provided.

3.4 Results

The data was analysed using statistical software R v3.5.0 (R Core Team, 2021) and tidyverse (Wickham et al., 2019), lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017) and scienceverse (Lakens & DeBruine, 2020) packages. All data and code are available at https://osf.io/rqvxp/. Generalised linear mixed effects models were used to conduct the analyses with crossed random effects for face pairs and participants specified maximally (Barr et al., 2013):

3.4.1 Similarity Scores

Similarity scores for the analyses were calculated by extracting the random intercepts for face pairs from a linear mixed effects model predicting the individual raw similarity scores with random effects for face pairs and participants (Figure 7).

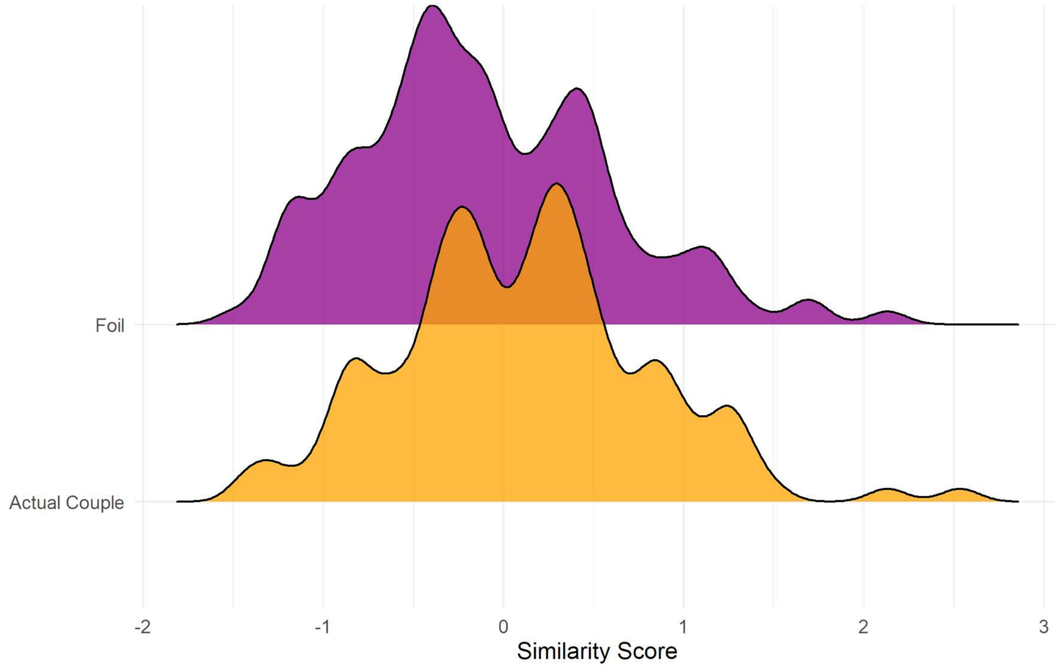


Figure 7: *Density plots showing distributions of similarity scores from 182 participants on 135 pairs of facial images of actual couples and 135 pairs of facial images of foils. Similarity scores were derived from the random intercepts for face pairs extracted from a linear mixed effects model with raw similarity scores (DV) and random effects for face pairs and participants.*

3.4.2 Analysis 1: Pair type effect on couple judgments

Hypothesis 1: We predict that pairs of faces are more likely to be judged as couples if they belong to actual couples than to foils (significant main effect of pair type on couple judgments).

A binary mixed effects model was used to analyse the effect of pair type on couple judgments. This model included a random effect for participants with a random slope of pair type (couple/foil) and uncorrelated random intercept and a random effect for face pairs with a random intercept. The analysis found a main effect of pair type (estimate = 0.30, se = 0.10, $z = 2.90$, $p < 0.001$), indicating that actual couples were more likely to be judged as couples than foil pairs in the couple condition.

3.4.3 Analysis 2: Difference in pair type effect on couple and sibling judgments

Hypothesis 2: We predict that the effect of pair type on the likelihood of making couple judgments will be larger than the effect of pair type on making sibling judgments (significant interaction between judgment and pair type).

A binary mixed effects model with fixed factors of judgment type, similarity and pair type predicting judgments from both tasks was used to analyse the difference in the effect of pair type on couple versus sibling judgments (see Table 1 for descriptive statistics). Random effects were specified for participants and face pairs, the former having a random slope of the interaction of similarity and pair type with an uncorrelated intercept, the latter with a random slope for judgment type and an uncorrelated intercept. This particular analysis focussed on the two-way interaction of judgment task and pair type. A significant interaction of pair type and judgment task type was not found (estimate = 0.09, se = 0.10, z = 0.89, p = 0.37), indicating that actual couples were more likely than foil pairs to be judged as both couples and siblings, and that the sizes of these effects were not significantly different (see Figure 8).

Table 1: Descriptive statistics for couple and sibling judgments

Pair Type	N	Mean	SD
Couple Judgments			
Actual couple	61	0.58	0.49
Foil	61	0.52	0.50
Sibling Judgments			
Actual couple	60	0.43	0.50
Foil	60	0.37	0.48

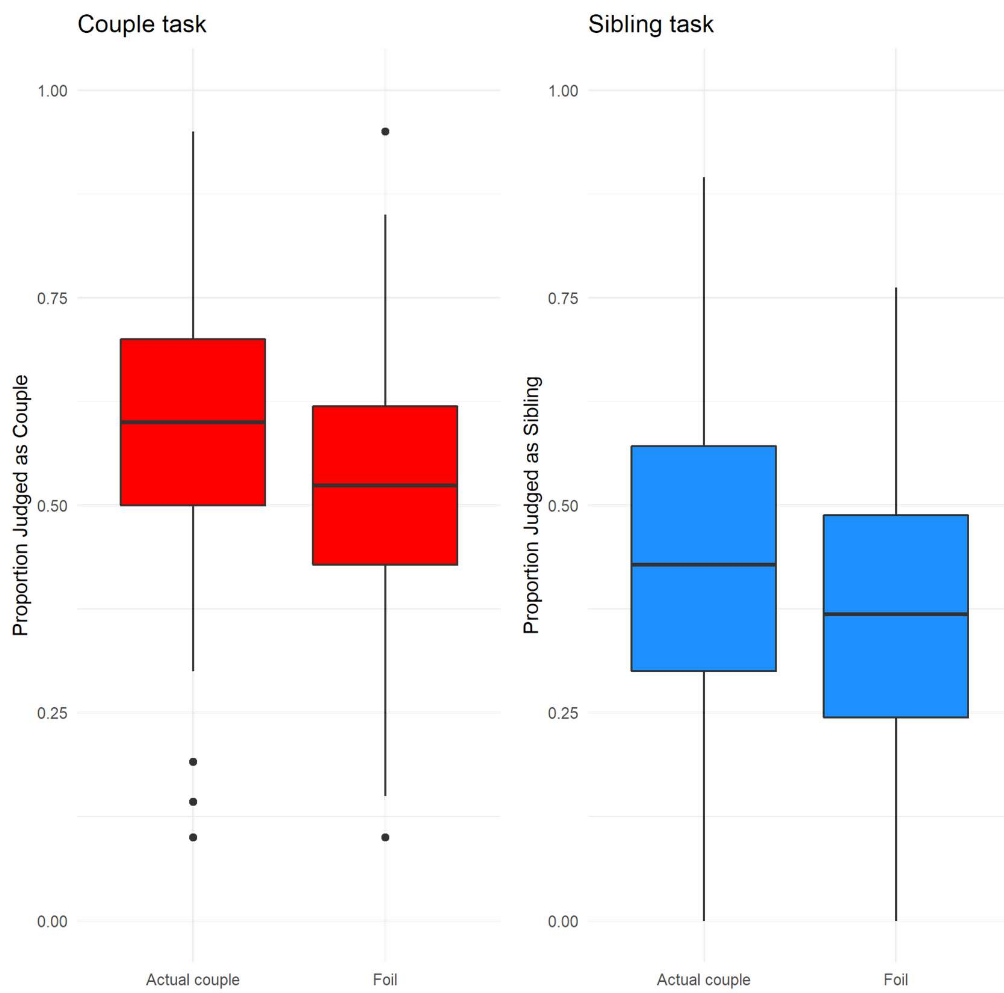


Figure 8: Boxplots depicting the median,25th and 75th percentile of the proportion of actual couples and foil pairs judged as couples and the proportion of actual couples judged as siblings.

3.4.4 Analysis 3: Similarity effect on sibling/couple judgments

Hypothesis 3: We predict that pairs who are judged as more facially similar are more likely to be judged as siblings (significant main effect of perceived similarity on sibling judgments) and as couples (significant main effect of perceived similarity on couple judgments).

A binary mixed effects model of sibling judgments predicted by similarity was used to analyse the effect of similarity on sibling judgments (M = 0.40, SD = 0.49). This model included a random effect of participants with a random slope for similarity and uncorrelated intercept and a random effect for face pairs with a random intercept. The analysis found a main effect of similarity (estimate = 0.93, se = 0.07, z = 12.66, $p < 0.001$), indicating that the likelihood of sibling judgments increased as similarity ratings increased.

Although not pre-registered, we also include here for comparison a model of *couple judgments ~ similarity*, used to analyse the effect of similarity on couple judgments (M =

0.55, SD = 0.50). The analysis found a main effect of similarity (estimate = 0.46, se = 0.08, $z = 5.71$, $p < 0.001$), indicating that the likelihood of couple judgments increased as similarity ratings increased.

3.4.5 Analysis 4: Difference in similarity effect on couple and sibling judgments

Hypothesis 4: We predict that the effect of perceived similarity on the likelihood of making sibling judgments will be larger than the effect of similarity [1]on making couple judgments (significant interaction between judgment and perceived similarity).

The same model that was used in Analysis 2 was used to analyse the difference in the effect of similarity on couple and sibling judgments, focussing on the two-way interaction of judgment type and similarity. The analysis found a significant interaction between similarity and judgment task type (estimate = -0.51, se = 0.10, $z = -5.31$, $p < 0.001$), indicating that the effect of similarity on the likelihood of affirmative judgments was larger for sibling than couple judgments (see Figure 9).

[1] The preregistration document has an obvious and nonsensical typo here, stating "...larger than the effect of **pair type**...".

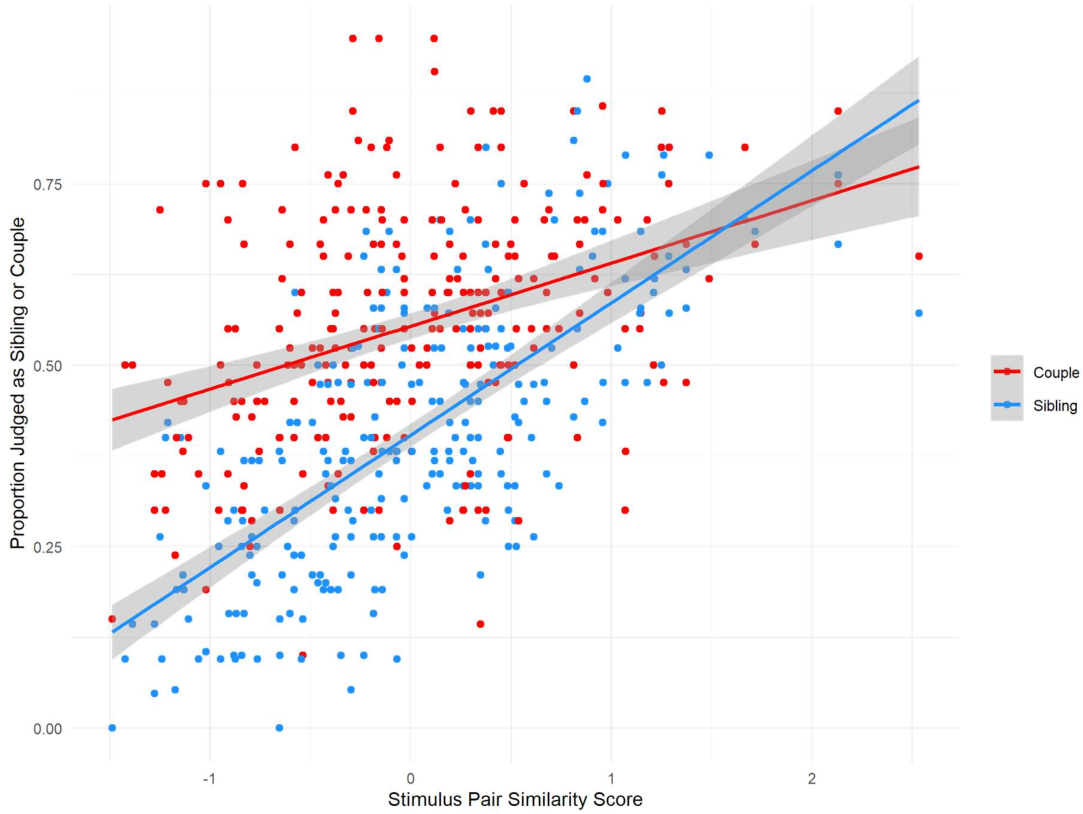


Figure 9: Scatterplot describing the relationship between similarity ratings and couple and sibling judgments on 135 facial images of actual couples and 135 images of foil pairs. The y-axis depicts the proportions judged as couples and siblings. The x-axis depicts the intercepts for the similarity ratings on the face pairs.

The first completion of each trial was kept for each participant, thus sessions with only 90 trials were kept in the dataset. Where participants may have completed more than one task per session, only the first entry was kept. Visual examination of the similarity scores distribution gave no indication that participants were not completing the task as expected and QQ plots for residual error were as expected given the binomial structure of the data. Therefore, there was no reason to inspect the data further for abnormalities.

3.5 Discussion

Using a combination of similarity ratings, couple judgments and sibling judgments on a set of facial photographs, the current study found that similarity in face pairs increased sibling judgments to a greater extent than couple judgments. We also found that actual couples were more likely than foils to be judged as both couples and siblings.

In answer to the specific questions:

1. Are couple judgments accurate? Yes. Pairs of faces were more likely to be judged as couples if they belong to actual couples than foils.
2. Are couple judgments different from sibling judgments? No, there was no significant difference in the effect of pair type (i.e., actual vs. foil pair) on the likelihood of making sibling judgments versus couple judgments.
3. Does perceived similarity drive sibling/couple judgments? Yes. Pairs who were judged as more facially similar were more likely to be judged as siblings and as couples, although the effect size for couple judgments was about half the size of the effect size for sibling judgments.
4. Does perceived similarity drive couple judgments less than sibling judgments? Yes. The effect of perceived similarity on the likelihood of making sibling judgments was significantly larger than for couple judgments.

Phenotypic traits reflect genotypes, about 50% of which are shared among siblings. It is conceivable, therefore, that the similarity effect observed on sibling judgments indicates that similarity acts as a cue of kinship. Previous studies on facial resemblance in siblings corroborate this finding using facial images of children (Maloney & Dal Martello, 2006) and adults (DeBruine et al., 2009). That similarity is a better predictor of sibling judgments than couple judgments points to a stark difference in how similarity shapes people's concepts of siblings and couples. The specialised use of perceived similarity in different contexts mimics findings from Study 1 using only couples' faces, where we found that similarity predicted sibling judgments (estimate = 0.86, se = 0.07, $z = 13.2$, $p < .001$) but not couple judgments (estimate = 0.06, se = 0.07, $z = 0.97$, $p = .33$). DeBruine (2005) also examined responses to facial resemblance in different contexts, finding that self-resemblance increased prosocial judgments of trustworthiness, decreased perceptions of attractiveness in short-term relationships and had no effect on attractiveness judgments in long-term relationships. The findings reported in the current study are in agreement with DeBruine (2005) and suggest that responses to perceived similarity diverge across kin and mate choice contexts as a means of calibrating how we interact with genetic relatives in different relationship settings.

The current study reported a weaker but still significant effect of perceived similarity on couple judgments, despite a stronger effect observed on sibling judgments. This result could partly be explained by optimal outbreeding theory, as underlying mechanisms operate to control the amount of relatedness that would be appropriate in mate choice. Alternatively, the weak association between similarity and couple judgments could be a consequence of elements that were not controlled for. Whilst matching on ethnicity has been demonstrated previously (Thiessen et al., 1997), all facial stimuli used in this study belonged to white individuals, although subtle ethnic differences within the broader "white" category were not controlled for.

By introducing random foil pairs, the current study sought to assess whether people can detect differences between actual couples and random pairs, as photo-matching tasks have indicated previously (Bereczkei et al., 2002, 2004; Chambers et al., 1983; Griffiths & Kunz, 1973; Thiessen et al., 1997). Indeed, subjects judged actual couples as couples at a rate that exceeded chance, although this effect was observed with sibling judgments as well, which would seem to suggest that humans use visual information to identify actual couples that is not unique to a mate choice context but is also informative to making kin-related judgments. This finding complicates earlier studies' interpretations of correct matches in couple-matching tasks as evidence of similarity in couples, strengthening the case that there has been a sustained misconception in the literature on the meaning of people's ability to reassort couples' faces. Whilst it does seem to be the case that people have some ability to distinguish between actual couples and random pairs, it does not necessarily mean that couples are actually similar. Future research might benefit from the inclusion of actual sibling pairs in similar study designs.

In conclusion, our findings support a mechanism that guides responses to similarity in a manner that is specific to the context it is presented in. Similarity appears to act as a kinship cue to which people respond positively in prosocial contexts and moderately or negatively in sexual contexts.

Study 3

Self-resembling and partner-resembling biases in perceptions of attractiveness and trustworthiness

4.1 Abstract

Studies have demonstrated self-resemblance biases in prosocial contexts, where cues of relatedness encourage altruistic behaviour, but not in sexual contexts, where mechanisms intervene to minimise opportunities for inbreeding (DeBruine, 2005; DeBruine et al., 2011). The current study tests participants' attributions of attractiveness and trustworthiness in two-alternative forced choice tasks using face pairs of self-resembling or partner-resembling transforms and non-resembling controls. Contrary to previous studies, self-resembling and partner-resembling biases were found for attributions of both attractiveness and trustworthiness, with a significantly stronger effect on attributions of attractiveness. The study also shows main effects for target face and face sex, which were qualified by an interaction such that stronger biases were noted in opposite-sex faces that were partner-resembling and in same-sex faces that were self-resembling. We conclude that while self-resembling and partner-resembling biases could be due to mere exposure, issues with the study design may have contributed to these results and thus further testing is required.

4.2 Introduction

Self-resemblance in couples has fascinated the masses, from consumers of online "listicles" of celebrity couples (Ro, 2015), to academic circles studying twins and their partners (Zietsch et al., 2011). Several studies have reported evidence for similarity in couples, also known as 'homogamy', on traits including hair and eye colour (Little et al., 2003; Saxton, 2016), body odour (Jacob et al., 2002), age (Feng & Baker, 1994) and personality (Caspi et al., 1992; Watson et al., 2004). The question that scientists today face is what proximate mechanisms explain this widespread affinity for similarity.

Broadly, two groups of explanations exist for similarity in couples, although evidence for both is mixed. First, homogamy could be attributable to an adaptive mechanism, where people use similarity as a cue of kinship to choose a partner with similar traits in pursuance of maximising inclusive fitness (Hamilton, 1964). However, high costs of consanguinity create opposing pressures on selection for similarity, giving rise to inbreeding avoidance behaviours that facilitate mate choice for an optimum degree of genetic relatedness (Bateson, 1983). The second main category of explanation for homogamy, non-adaptive hypotheses, predicts that similarity arises as a by-product of a general preference for traits that individuals have been repeatedly exposed to (Zajonc, 1968). Hence, a by-product explanation does not ascribe a purposeful role to similarity in mate choice, arguing that resemblances within couples emerge as a consequence of a separate mechanism.

The task of investigating why similarity is so prevalent in non-kin relationships can be approached at by examining how people assess self-resemblance in different social settings. Computational techniques have made it possible to test responses to perceived similarity through the use of self-resembling morphs (DeBruine, 2005; Kocsor et al., 2011). DeBruine's (2002) study adopted a computer trust game where participants played against a series of computer-programmed players depicted by avatars of self-resembling and non-resembling morphs. The game required participants to make one of two decisions, depending on which player role they were assigned: (1) a decision by Player 1 on whether or not to trust Player 2 with sharing a monetary reward; and (2) a decision by Player 2 on whether or not to reciprocate Player 1's trust with an unselfish move. Participants played 6 rounds as Player 1 and 6 rounds as Player 2 and the sequence of roles and decisions by computer-programmed players were kept constant. The study found that individuals trusted self-resembling morphs significantly more than non-resembling morphs in the Player 1 role but their decisions as Player 2 were indiscriminately unselfish for both types of morphs. The lack of an effect on Player 2 decisions could have been attributable to the monetary reward structure, although further research would be necessary to explain this result. Trustworthiness is a prosocial attribution that self-resemblance is expected to increase and thus the results from this study support the hypothesis that self-similarity enhances prosocial behaviour, at least in preliminary trust decisions, but could not completely rule out a familiarity hypothesis.

To examine how self-resemblance influences sexual behaviour and whether familiarity facilitates preferences, DeBruine (2004) tested responses to self-resemblance on perceptions of attractiveness and averageness. Shape transforming methods (DeBruine et al., 2005), 2005) were used to manipulate same-sex and opposite-sex composite faces to make them resemble participants' faces. Participants made two-alternative forced choice decisions on same-sex and opposite-sex face pairs of all possible combinations for self-resembling, other participant-resembling and average composite faces. One group of participants was instructed to choose the face they found most attractive, whereas a second group of participants was instructed to choose faces on the basis of averageness, translated as the most 'typical' or 'ordinary'. This latter experiment was designed to test how typical self-resemblance is perceived compared to computer-transformed average composites, given that attributions of averageness reflect prototypical formations arising from repeated exposure to patterns of faces (Langlois et al., 1987). The results showed that facial resemblance increased attractiveness in same-sex face composites more than other-sex face composites. Further, average composites were chosen as more average than self-resembling transforms for both same-sex and other-sex faces, while self-resembling transforms were chosen as more average than other participant-resembling transforms equally for both same-sex and other-sex faces. Perceptions of averageness reported in this study indicate that recent experience with faces affects perceptions of both same- and other-sex faces, suggesting that while familiarity does influence perceptions of averageness in self-resembling faces, it could not explain the differential levels of attractiveness observed in same-sex and other-sex self-resembling faces. The increase in non-sexual preferences for self-resemblance has been interpreted as being in line with predictions of inclusive fitness theory, while decreased sexual preferences for self-resemblance has been interpreted as being in line with optimal outbreeding theory.

In a later study, DeBruine (2005) showed that responses to facial resemblance indeed differ across prosocial and sexual contexts. Participants viewed a set of other-sex participant-resembling face transforms and made two-alternative forced choice decisions based on perceived trustworthiness, attractiveness in long-term relationships and attractiveness in short-term relationships. The chosen measures should prompt considerations pertinent to their specific contexts, allowing for robust testing of responses to self-resemblance against predictions from adaptive and familiarity theoretical frameworks. Self-resembling biases were calculated as the number of times the self-resembling transform was chosen compared with the average number of times other participants chose the same image. Analysis revealed that self-resemblance enhanced perceptions of trustworthiness. On the other hand, self-resemblance decreased perceptions of attractiveness for short-term relationships and did not influence perceptions of attractiveness for long-term relationships. Such context-specific responses to self-resemblance were further supported by a study examining whether having opposite-sex siblings affects women's responses to self-resembling faces on attractiveness and trustworthiness (DeBruine et al., 2011). In a similar design to DeBruine (2005) but incorporating non-resembling foils, participants made two-alternative forced choices between self-resembling transforms and foils, and between other participant-resembling transforms and foils, for each sex category. Scores for other participant-resembling pairs were deducted from the scores for self-resembling pairs to calculate the extent to which self-resemblance was found to be more attractive or trustworthy by each participant. The authors reported self-resemblance biases for same-sex trustworthiness, opposite-sex trustworthiness, and same-sex attractiveness, but not opposite-sex attractiveness. In accordance with inbreeding avoidance predictions, women with brothers showed weaker self-resemblance preference for male attractiveness than women who did not have brothers. The presence of brothers has no effect on self-resemblance preferences for male trustworthiness. The results from these two studies suggest that mere exposure alone cannot explain attitudes to self-similarity, which would predict preferences for self-resemblance to increase in general and equally across contexts.

On the other hand, an earlier study testing different theoretical explanations for homogamy concluded it could only support a repeated exposure hypothesis (Hinsz, 1989). Participants, divided into 'younger' and 'older' categories based on their age, rated the similarity of face pairs of engaged couples, married couples and foils on a 9-point scale. If similarity in couples could be explained by a repeated exposure hypothesis, younger and older participants' responses to similarity would be positive and consistent for engaged and married couples' faces. If environmental co-existence was responsible for perceived similarity in couples, married couples would be perceived as more similar than engaged couples. Finally, if similarity in couples was due to a perceptual bias, younger raters would attribute higher similarity to married couples than engaged couples and vice versa with older raters. This latter prediction stems from an impaired ability to discriminate faces belonging to a different social group to one's own (Tajfel, 1981 as cited in Hinsz, 1989). A repeated measures

ANOVA revealed no difference in responses from younger and older participants and no difference in perceived similarity between engaged and married couples, yet both engaged and married couples were perceived as more similar than random pairs. Hence, the author concluded that the findings support a repeated exposure hypothesis for homogamy. Considerations were not made, however, in regards to appropriating facial similarity to a sexual context when interpreting the results. Participants were free to make their judgments in whatever manner felt appropriate and were not informed that any of the face pairs belonged to couples. Consequently, the faces could have just as easily been subconsciously processed as opposite-sex siblings as they were couples, bringing into question the robustness of this study's design in demonstrating theoretical explanations for similarity in a sexual context.

Research on effects of visual processing enhance our understanding of how exposure shapes perceptions of attractiveness in faces. Recent experience with faces has been shown to influence attributions of attractiveness and trustworthiness on prototype faces (Buckingham et al., 2006). Participants were presented with masculinized and feminized transforms of adult male faces and assigned to either rate the faces for attractiveness or rate the faces for trustworthiness on a 4-point scale. Following a period of adaptation, where participants viewed either the masculinised or feminised faces twice as a means of becoming familiar with a pattern of faces, participants repeated the pre-adaptation task on the same faces. The results indeed showed that recent experience with the type of faces viewed in the adaptation phase increased the strength of preference for those types of faces in the post-adaptation task. The study demonstrates that experience enhances attitudes towards stimuli encountered and does so across contexts. Thus, while it appears that repeated exposure plays a role in shaping attitudes towards faces, some other mechanisms must also be functioning to explain differential responses to facial similarity in kinship and mate choice contexts.

Results from previous studies on the effects of self-resemblance on social judgments are mixed. Evidence for an adaptive mechanism strongly suggest that responses to similarity are attuned to the context in which they are situated. Yet effects of familiarity could not be completely ruled out either and studies indicate that repeated exposure influences preferences for similarity to some degree. In the current study, we investigate responses to attractiveness and trustworthiness in self-resembling and partner-resembling facial transforms. If self-resemblance and/or partner-resemblance are perceived as more trustworthy but less attractive, the findings from this study will replicate previous studies supporting similarity as a cue of kinship that modulates responses in prosocial and sexual contexts. If self-resemblance and/or partner-resemblance enhances attitudes of trustworthiness and attractiveness equally, the findings from this study will support a mere exposure explanation for preferences for similarity.

4.3 Method

4.3.1 Design

The current study investigated judgments of attractiveness and trustworthiness on a set of facial images using a cross-classified design. Participants viewed 20 faces (10 self-resembling and 10 partner-resembling) paired with 10 non-resembling controls in a repeated measures experiment with two-level conditions for judgment type (attractiveness and trustworthiness) and face sex (same-sex and opposite-sex), forming 4 tests in total.

As a result, each participant made a total of 80 two-alternative forced choice judgments. The dependent variable was binary judgment scores indicated by which face was chosen, self-/partner-resembling or control. The independent variables were categorical within-subjects factors of: *judgment type* (attractiveness or trustworthiness), *target face* (self-resembling or partner-resembling), and *face sex* (same-sex or opposite-sex).

4.3.2 Participants

Data was collected by the lab as part of another, larger study (Wincenciak et al., 2015) and consisted of 126 participants who formed 63 couples recruited from the University of Glasgow subject pool. One same-sex couple was excluded from the dataset, leaving 62 couples (62 men and 62 women) aged between 18 and 36 years old.

4.3.3 Stimuli

2D facial images of all participants were collected. Participants were instructed to stand on a platform, standardising height to eye-level with the camera, along with a X-Rite colour checker chart. Head-to-camera distance was kept constant across participants and photographic lighting was installed on either side of the camera. The images were then aligned on the pupils and delineated using WebMorph.org (DeBruine, 2018). Composite young white male and female faces were created by averaging 20 images from an openly available dataset (DeBruine & Jones, 2007) of young adults from Ontario, Canada. 10 individual male and female faces from the same image set were chosen as unfamiliar faces. The composite male and female faces were transformed in WebMorph (DeBruine, 2018), using methods used in earlier studies (DeBruine, 2005; DeBruine et al., 2011), to resemble each of the 10 male and 10 female unfamiliar faces by 50% in shape, creating foils for each sex category. Likewise, the male and female composites were transformed to resemble each participant and their partner, creating a self-resembling and partner-resembling face for each sex category. These methods are described in detail in a previous study (DeBruine et al., 2005). Each of the foils was then paired with the self-resembling and partner-resembling face of the congruent sex category (see example in Figure 10).

(a)

(b)

Figure 10: Example of stimuli as presented in the study, transformed using faces from an open dataset (DeBruine & Jones, 2007). (a) Same-sex category: The image on the left depicts a self/partner-resembling face and the image on the right

depicts a non-resembling foil; (b) Opposite-sex category: The image on the left depicts the same non-resembling foil and the image on the right depicts the same self/partner-resembling face. Face stimuli were created by transforming an average composite to resemble the participant by 50% in shape (self/partner-resembling) and an unfamiliar face by 50% in shape (foil). Each self/partner-resembling faces was paired with a foil and placed on a white background. For this demonstration, the identity depicted is not an actual participant but was taken from the same open dataset as the foil.

4.3.4 Procedure

The data was collected in the lab via faceresearch.org. Participants were recruited on a voluntary basis and were free to leave the study at any point. The four tests were presented to participants in a random order and, depending on the condition, participants were instructed to choose which face in a pair looked more trustworthy or physically attractive. Participants were given 4 blocks of 20 trials, one block for each combination of face sex and judgment, in a random order for each participant. In each block, they viewed 10 trials choosing the more trustworthy or attractive face from a pair of faces consisting of a self-resembling face and one of the 10 sex-matched foil-resembling faces, and a further 10 trials comparing the partner-resembling face with the 10 foils. Trials within each block were presented in a random order for each participant. Participants clicked on the face they found more attractive or trustworthy to make the judgment for each face pair.

4.4 Results

Data analysis was conducted on R v3.5.0 (R Core Team, 2021) statistical software, using tidyverse (Wickham et al., 2019) and lmerTest (Kuznetsova et al., 2017) packages. Judgment scores were analysed using a binary mixed effects model with a random slope for the three-way interaction of target face (self/partner), sex type (same/opposite) and judgment type (attractiveness/trustworthiness) with uncorrelated intercepts varying within factors participant pairs and stimulus pairs . When specified maximally, the model failed to converge, therefore the model was specified with a random slope for just the three-way interaction among random effects. All data and code are available at https://osf.io/e8caf/.

Analysis of attractiveness and trustworthiness judgments (see Table 2 for descriptive statistics) of all faces revealed a significant positive intercept (estimate = 1.287, se = 0.15, z = 8.49, $p < 0.001$), such that self- or partner-resembling faces were consistently judged as more attractive and trustworthy than the foil faces. We'll refer to this effect as a self and/or partner bias.

There was a main effect of judgment type (estimate = 0.20, se = 0.05, z = 3.98, $p < 0.001$), such that participants were more likely to show this positive bias for attractiveness judgments than trustworthiness judgments. There was a main effect for target face (estimate = -0.13, se = 0.05, z = -2.71, $p < 0.01$), such that participants were more likely to show a bias for partner-resembling than self-resembling faces. A main effect of face sex (estimate = -0.15, se = 0.05,

$z = -3.07$, $p < 0.01$) indicated significantly stronger biases for opposite-sex faces than same-sex faces.

However, the latter two main effects were qualified by an interaction between target face and face sex (estimate = 0.33, se = 0.10, $z = 3.36$, $p < 0.001$), indicating significantly stronger biases for opposite-sex categories when faces are partner-resembling and same-sex categories when faces are self-resembling (see Figure 11). No interaction effects were found between judgment type and target face ($p = 0.82$) or between judgment type and face sex ($p = 0.49$) (see Table 3).

Table 2: Descriptive statistics for attractiveness and trustworthiness judgments. Target Face	Sex Type	N	Mean	SD
Attractiveness Judgments				
Partner-resembling	Opposite-sex	124	0.80	0.40
Partner-resembling	Same-sex	124	0.76	0.43
Self-resembling	Opposite-sex	124	0.76	0.43
Self-resembling	Same-sex	124	0.76	0.42
Trustworthiness Judgments				
Partner-resembling	Opposite-sex	124	0.78	0.41
Partner-resembling	Same-sex	124	0.72	0.45
Self-resembling	Opposite-sex	124	0.73	0.45
Self-resembling	Same-sex	124	0.73	0.45

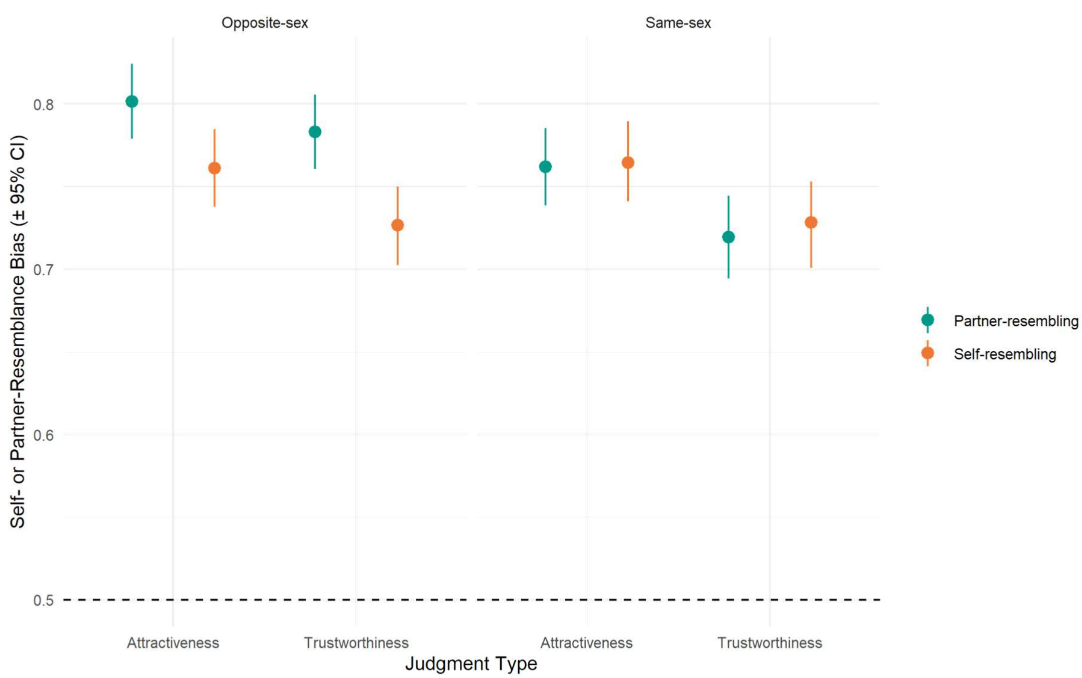


Figure 11: Proportion of self- or partner-resembling faces chosen as more attractive or trustworthy for opposite-sex and same-sex self-resembling and partner-resembling faces. 0.5 is chance -- choosing self/partner-resembling and foil-resembling faces with equal probability.

Table 3: Summary statistics for non-significant effects.

Effect	Estimate	Std Error	Statistic	P
Judgment Type *Target Face	0.022	0.098	0.225	.821
Judgment Type *Sex Type	0.067	0.098	0.685	.493
Judgment Type * Target Face * SexType	-0.125	0.218	-0.573	.566

Quality checks involved ensuring that both partners of a couple were present in the dataset and making sure that partners were of opposite sexes. Given the difficulty in checking for incidences where participants may have repeatedly chosen an image on the same side, completed tasks too slowly or too quickly, exclusions on such basis of participant performance were not conducted so as to avoid the possibility of "p-hacking". QQ plots for residual error were as expected given the binomial structure of the data. Therefore, there was no reason to inspect the data further for abnormalities.

4.5 Discussion

In the current study, we reported positive self or partner biases in judgments of attractiveness and trustworthiness, with a stronger tendency to judge these faces as attractive than trustworthy. The reported attractiveness and trustworthiness biases were larger for self- than partner-resembling faces in same-sex categories and larger for partner- than self-resembling faces in opposite-sex categories.

Our findings indicate that facial resemblance influences social decisions, but the directionality of effects observed remain puzzling in comparison with previous studies. Whereas DeBruine (2005) suggests that facial resemblance acts as a cue of relatedness which elicits specialised responses in prosocial and sexual contexts, our findings do not reflect this context-specific function of similarity. The present study did not replicate the stronger self-resemblance bias in trustworthiness than attractiveness reported in previous studies (DeBruine, 2005; DeBruine et al., 2011). Indeed, the self-resemblance bias for trustworthiness was significantly smaller than for attractiveness.

The strength of the self or partner bias observed in perceptions of both attractiveness and trustworthiness could be an artefact of the study design, however. That ten different sex-matched foil images were paired with the ten self-resembling and ten partner-resembling transforms may have created a visual exposure effect within the study design itself in that self- and partner-resembling faces were viewed 10 times more frequently than each non-resembling identity. To directly test for this incidental effect of the study's design, we would need to replicate the study using the same face stimuli but with a different group of participants who are not the identities reflected in the transforms, and compare their preferences with those of the original self/partner participants.

Alternatively, this conflicting result might be compatible with a mere exposure explanation insomuch as the self and/or partner bias was stronger for attractiveness attributions than trustworthiness attributions. Due to the inclusion of partner-resembling faces, the self and/or partner bias in attractiveness and trustworthiness might reflect familiarity effects overriding pressures from inclusive fitness and inbreeding avoidance. Furthermore, it is plausible that familiarity with a partner's face could be driving preferences for partner-resembling faces. An individual's experience with their partner's face could surpass experience with their own face, given the fewer opportunities to encounter one's own face compared with a partner's, particularly if cohabiting. The stronger bias for partner-resembling faces than self-resembling faces for opposite-sex faces may otherwise be echoing further limitations in the design of the current study. By assessing responses to both self-resembling and partner-resembling stimuli our study fails to account for the fact that participants would already have established a likeness for their own partner's facial traits, and this may be reflected in the responses.

Nevertheless, a mere exposure hypothesis could not explain the interaction effect between target face and face sex given the incompatibility of a sex difference with a mere exposure prediction. DeBruine (2004) showed that same-sex and opposite-sex self-resembling faces were perceived as equally average, indicating that prototype formation of typical faces is sex-neutral. Thus, if the self and/or partner bias were entirely due to a mere exposure effect, same-sex and opposite-sex faces would have been perceived as equally attractive and trustworthy, which was not observed in this study. Alternatively, the findings from DeBruine (2004) might indicate that opposite-sex self-resembling faces are perceived as more average not because of visual experience with one's own face, but because of visual experience with opposite-sex relatives, who are likely to resemble oneself.

To conclude, the findings reported in this study were difficult to interpret in light of the issues with the study design. The absence of context-specificity in self and/or partner biases may be indicative of a familiarity effect, although mere exposure could not explain the stronger bias for opposite-sex partner-resembling faces. That participants viewed self-resembling and partner-resembling faces more often than foils may have led to a visual exposure effect intrinsic to the study's setup. Moreover, the inclusion of partner-resembling stimuli made it difficult to isolate effects of self-similarity from established preferences for partner-resembling traits. A follow-up study, with a new set of participants who are not affiliated with the individuals reflected in the facial stimuli would shed more light on whether the study design confounded the effects of self-resemblance with a visual adaptation effect.

General Discussion

The objective of this PhD thesis was to examine how similarity fits into mate choice considering previous studies' use of confounded measures of similarity in demonstrating support for theoretical explanations for homogamy. Reports of perceived or self-reported trait correlations in couples give a reasonable indication of the extent of similarity shared within couples on traits measured (Horwitz & Keller, 2022). Yet studies testing theoretical explanations for homogamy do not effectively isolate measures of similarity from other perceptions that could affect couple judgments in their methodological approaches, nor do they adequately address the causal assumptions underpinning those theoretical frameworks. These shortcomings were addressed in this thesis with three studies reviewed below.

5.1 Summary of empirical studies conducted

5.1.1 Review of Study 1 'Facial similarity predicts sibling judgments but not couple judgments'

The first study discussed in this body of work inspected the relationship between three abstract concepts that are often confounded in experimental research on homogamy: *perceived facial similarity*, *likelihood of being siblings* and *likelihood of being a couple*. Participants made either couple judgments or sibling judgments on a set of face pairs of 139 couples, whereas similarity ratings had been obtained as part of another study prior. Binomial mixed effects models were used to assess how well similarity predicted the likelihood of the couples being judged as siblings or as a couple, thus comparing the way people might use perceived similarity to inform their concepts of consanguine (related by blood) versus affine (related by marriage/partnership) relationships.

The findings from this study initially revealed that similarity predicted both sibling and couple judgments but had a significantly stronger effect on sibling judgments. During the online recruitment stage, an invitation to participate in the experiment was shared as a comment on a social media post about a news article on similarity in couples. Given that the article's subject matter was very close to that of the study, an analysis was conducted to determine the impact of being exposed to the article, that of informing subjects to use similarity when making couple judgments. This latter analysis was conducted on subjects who were assigned to make couple judgments on and after the date that the article was shared on social media. Interestingly, the results indicated that this subset of participants did indeed use similarity to make couple judgments to a significantly greater extent than did earlier participants who weren't exposed to the article. A reanalysis of the original model, after restricting data to that collected before the article was posted online, showed that perceived similarity no longer predicted couple judgments, providing evidence of a context-specific response to similarity.

The findings from this study add to the body of knowledge on homogamy by highlighting a stark difference in the role of perceived similarity in guiding behaviour in prosocial and sexual contexts. The strong effect of similarity on sibling judgments suggests that people use phenotype matching to assess genetic relatedness through perceived similarity, a distinct cue of kinship. That similarity was a strong predictor of sibling judgments but not couple judgments, for the final subset of participants at least, indicates that this mechanism for detecting relatedness is not exercised in a sexual context. Hence, while perceived similarity may have an evolutionary purpose in detecting kin to appropriate prosocial behaviour towards genetic relatives over non-relatives, the evidence presented in this study does not support an adaptive hypothesis for detecting an optimum amount of relatedness in potential mates.

Incidentally, the study demonstrated the importance of isolating conceptualisations of similarity from couple judgments, casting doubt on the use of matching tasks to support theoretical explanations for similarity in couples. The lack of non-couple stimuli limited the study's ability to assess whether people are in fact accurate at making couple judgments as indicated by studies using matching tasks.

5.1.2 Review of Study 2 'Facial similarity, sibling judgments and couple judgments: A replication study using actual couples and random pairs'

The second study discussed in this dissertation replicated the first study using stimuli of both actual couples and random matched pairs. As similarity ratings needed to be obtained on the new stimuli, a third condition was added to the experiment such that participants were assigned to either rate the face pairs for similarity, judge whether the face pairs looked like couples, or judge whether the face pairs looked like siblings. The three measures collected were analysed with binomial mixed effects models to assess effects of similarity and pair type (i.e., whether the face pairs are actual couples or foils) on judgments. Consistent with the first study, perceived similarity was a stronger predictor of sibling judgments than couple judgments, although the effect on couple judgments was still significant. In comparison with foil pairs, actual couples were significantly more likely to be judged as couples, but this effect was observed with sibling judgments as well.

This study provides further support that similarity generates distinct responses in contexts where genetic relatedness is beneficial versus contexts where it is not. That similarity was a weak but still significant predictor of couple judgments is in line with optimal outbreeding theory functioning to moderate the appeal of similarity in a sexual context. Specifically, despite not replicating the non-significant effect of perceived similarity on couple judgments from Study 1, the difference in the similarity effect across judgment types reported in Study 2 remains significant and hence indicative of a context-specific response to similarity, while the weak effect of perceived similarity on couple judgments could be reflecting evolutionary forces regulating the extent of relatedness considered appropriate for couples. By incorporating foil pairs in the design, the study demonstrated that the accuracy of couple judgments is not necessarily indicative that people use similarity to make these judgments.

5.1.3 Review of Study 3 'Self-resembling and partner-resembling biases in perceptions of attractiveness and trustworthiness'

In the final study, we investigated preferences for self-resemblance and partner-resemblance from two standpoints: *attractiveness* and *trustworthiness*. Self-resembling and partner-resembling transforms from 62 couples were created for same-sex and opposite-sex categories and paired with non-resembling foils matched on sex and age. The participants made two-alternative forced choice decisions on all face pairs for both attractiveness and trustworthiness. The data was analysed with a mixed effects model examining effects of judgment type, face sex and target face (i.e., self-resembling or partner-resembling).

The results from this study were contrary to expectations. Evidence for a mere exposure hypothesis would have seen an equal self or partner bias in attractiveness and trustworthiness judgments, indicating that repeated exposure to familiar faces enhances attitudes towards similar faces across all contexts. A self and/or partner bias was indeed found for both attractiveness and trustworthiness, but this effect was stronger for attractiveness judgments, which does not support the theory that perceived similarity is a cue to kinship which people respond to positively in prosocial contexts and negatively or moderately in sexual contexts. Furthermore, partner-resembling faces showed a stronger bias than self-resembling faces, while opposite-sex faces showed a stronger bias than same-sex faces. However, these latter two main effects were qualified by an interaction effect that revealed stronger biases for partner-resemblance when faces were of the opposite-sex category and for self-resembling when faces were of the same-sex category. Attractiveness and trustworthiness judgments did not differ significantly between target face or face sex categories. The findings for this study, therefore, contradict those from Study 1 and Study 2 by lacking support for an evolutionary mechanism for perceived similarity that generates a more favourable response towards relatives than potential mates. Whilst a mere exposure effect might explain the self and/or partner bias in both conditions, it could not account for the interaction effect of target face and face sex.

Issues with the study design may have been the cause behind the unexpected interaction effect of target face and face sex which did not differ between judgment types. The way the study was set up made it so that subjects viewed self-resembling and partner-resembling faces more often than foils, potentially resulting in a visual adaptation effect that generated the self and/or partner bias across both conditions. Further, by including both self-resembling and partner-resembling transforms, the study failed to account for established preferences for partner-resembling traits, given that participants would already find their partner attractive and trustworthy.

5.2 Position within the wider literature

As expected, the findings from Study 1 and Study 2 were compatible with Maloney and Dal Martello (2006) and DeBruine et al. (2009) in that similarity strongly predicted sibling judgments. This effect indicates that a considerable amount of information available from perceived similarity is reflective of kinship and thus assessments of perceived similarity serve as a valid means for kin recognition. Posing the same analysis to couple judgments comparably evaluates similarity's role in conceptualising the relationship between two individuals but in a sexual context. Consistent with DeBruine (2002, 2004, 2005), we reported a significantly stronger effect of similarity on sibling judgments than couple judgments in both Study 1 and Study 2, although the effect on couple judgments was not significant in Study 1 and weak but significant in Study 2. Nevertheless, this finding is consistent with optimal outbreeding theory that inbreeding avoidance behaviours function to discourage mating with close kin by mitigating the appeal of similarity in sexual relationships. Such effects would not be observed if similarity was merely a by-product of familiarity, but the studies do not exclude the possibility that familiarity contributes towards similarity to some extent.

Whilst the studies presented here did not use a matching task, Study 2 examined the equivalent of correct matches by assessing how likely participants were to judge actual couples as couples compared with matched foil couples. As observed in Griffiths and Kunz (1973), participants were significantly more likely to judge actual couples as couples than they were foils. Although not tested in Griffiths and Kunz (1973), or any other paper as far as we are aware, actual couples were also more likely than foils to be judged as siblings, highlighting important considerations on interpreting results from correct matches on couples' faces. However, perceptions of similarity predicted sibling judgments to a significantly greater extent than they predicted couple judgments. Though not in direct conflict with previous studies that have used matching tasks, we show that evidence of correct matches does not necessarily mean that people are using similarity in the same way to make kin and couple judgments.

Our findings from Study 3 did not replicate the context-specific effects observed in DeBruine (2005) and DeBruine et al. (2011). Confounds brought about by the study’s design may have contributed towards the contradictory effects. By incorporating partner-resembling stimuli, the study inadvertently permitted pre-established preferences for partner-resemblance to override biases for self-resemblance. Moreover, the way foils were matched with target faces meant that participants viewed self-resembling and partner-resembling transforms more often than non-resembling transforms, creating a visual adaptation effect within the study which may have contributed to the consistent effects on attractiveness and trustworthiness. Therefore, whereas DeBruine (2005) and DeBruine et al. (2011) found that self-resemblance enhanced attributions of a prosocial kind and moderated attributions of a sexual kind, this bias was equally positive across contexts.

While the facial stimuli used in the studies conducted eliminated cues from features such as hair or makeup, other cues (that are difficult to control for), such as cues from eyebrow grooming or facial piercings, may have escaped into the final images. It is certainly possible that such cues, that we did not intend to measure for perceived similarity, were picked up on by participants anyway and in turn affected their judgments.

This particular thesis studies static facial morphological cues, however of course humans may use any number of other cues available to them to make judgments outside of a controlled experiment, such as vocal or olfactory cues. Even with a clear experimental design and all controls in place, it is still possible that the judgments made in the experiments conducted do not reflect judgments humans would make in the real world. Lorusso and colleagues (2015) argue that a rating measure of similarity does not reflect the complexity with which humans make judgments of similarity. Likewise, a two-alternative forced choice design may not capture judgments of couples or siblings in the way such concepts are studied in this thesis. Given the lack of empirical evidence available to rule this out, we can only acknowledge that such a prospect is indeed possible despite designing experiments to capture the human judgments under study as intended.

5.3 Contributions to research and future scope

Study 1 contributed to research by taking a different methodological approach to assessing similarity’s role in mate choice, compared with what has been adopted in the literature. This was achieved by modelling similarity as a linear predictor of couple judgments and sibling judgments on couples' faces, thereby comparing the direction and extent to which similarity is associated with conceptualisations of romantic couples and siblings. Given the inclusive fitness benefits of genetic relatedness in an adaptive hypothesis for assortative mating, a comparison of similarity's association with kinship versus its association with couples provides a deeper insight into whether and to what degree similarity is sought out in a mate. This assessment is of interest to the study of homogamy as it set out to separately measure abstract concepts of similarity, likelihood of being a couple and likelihood of being siblings, without making any causal assumptions about the relationships between these concepts that subscribe to any particular theoretical framework. The evidence provided is thus informative to understanding similarity's role in social interactions and how it interacts with other concepts, including mate choice, without confounding variables of interest.

By incorporating random pairs in a replication of Study 1, Study 2 showcased that correct matches on matching tasks are not evidence of similarity inasmuch as they are evidence that people are capable of matching couples correctly. The finding that the effect of pair type on couple judgments is no different than it is on sibling judgments demonstrates that while people can distinguish between couple pairs and random pairs, the basis on which they do so is informative to judging kinship and couples.

A final contribution to research that emerged from this thesis presented an investigation into and direct comparison of preferences for self-similarity and partner-similarity. Despite conflicting results and design flaws, the findings from Study 3 could reveal interesting considerations about how people assess self- and partner-resemblance, that preferences for self-resemblance are important in social interactions, but might not be as important as partner-resemblance or general attractiveness (Kocsor et al., 2011).

As noted in several other studies previously, the facial stimuli used in the studies reported on in this thesis were limited to subjects from WEIRD societies. Thus, future research would benefit from a more diverse stimulus set, allowing for the capturing of cultural influences and other potentially relevant mechanisms to assortative mating, such as matching on ethnicity. Future replications of Study 2 could consider including actual sibling pairs in the stimulus set, in addition to couple pairs and foils. The design for this proposed future study would involve the same similarity task conducted on facial stimuli of three types: sibling pairs, couple pairs and foil pairs. The sibling judgments task would be conducted on sibling pairs against foil pairs, while the couple judgments task would be conducted on couple pairs against foil pairs. Analysis would involve applying the same mixed effects models to address the same research questions. It is likely that this study design would generate a stronger effect of similarity on sibling judgments and thus a greater difference in the similarity effect across judgment types. Additionally, the inclusion of sibling pairs for the sibling judgment task will allow for a direct comparison of the accuracy with which humans make judgments (i.e., pair type effect) across contexts.

The design flaws in Study 3 incidentally create an opportunity for future research to address the need for independent judges to perform the task, as well as to rectify the unequal viewing frequency of resembling and non-resembling faces. Thus, a potential future study to address these shortcomings could involve recruiting independent judges who are not the identities of the individuals behind the self and partner resembling faces used as stimuli. In so doing the study could add further clarity on whether mere exposure could explain the effects observed.

The studies in this thesis support the hypothesis that perceived similarity acts as a cue of kinship for which preferences have adapted to direct individuals to behave favourably towards similar others where higher degrees of relatedness would be beneficial and to assert caution with similar others where higher degrees of relatedness could be deleterious. The findings from Study 3 also present some evidence in support of a familiarity hypothesis although given the issues with the study design, this interpretation cannot be fully supported without further research addressing the problems with the way the study was originally set up.

References

Andersson, M. (2019). *Sexual Selection*. Princeton University Press. https://doi.org/10.1515/9780691207278

Andersson, M., & Simmons, L. W. (2006). Sexual selection and mate choice. *Trends in Ecology & Evolution*, *21*(6), 296–302. https://doi.org/10.1016/j.tree.2006.03.015

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

Barrett, S. C., & Harder, L. D. (1996). Ecology and evolution of plant mating. *Trends in Ecology & Evolution*, *11*(2), 73–79. https://doi.org/10.1016/0169-5347(96)81046-9

Bates, D., Maechler, 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. P. G. (1966). The Characteristics and Context of Imprinting. *Biological Reviews*, *41*(2), 177–217. https://doi.org/10.1111/j.1469-185X.1966.tb01489.x

Bateson, P. P. G. (1978). Sexual imprinting and optimal outbreeding. *Nature*, *273*(5664), 659–660. https://doi.org/10.1038/273659a0

Bateson, P. P. G. (1983). Optimal outbreeding. In *Mate Choice* (pp. 257–278). Cambridge University Press.

Bereczkei, T., Gyuris, P., Koves, P., & Bernath, L. (2002). Homogamy, genetic similarity, and imprinting; parental influence on mate choice preferences. *Personality and Individual Differences*, *33*(5), 677–690. https://doi.org/10.1016/S0191-8869(01)00182-9

Bereczkei, T., Gyuris, P., & Weisfeld, G. E. (2004). Sexual imprinting in human mate choice. *Proceedings. Biological Sciences*, *271*(1544), 1129–1134. PubMed. https://doi.org/10.1098/rspb.2003.2672

Bittles, A. H., Grant, J. C., Sullivan, S. G., & Hussain, R. (2002). Does inbreeding lead to decreased human fertility? *Annals of Human Biology*, *29*(2), 111–130. https://doi.org/10.1080/03014460110075657

Boothroyd, L. G., Meins, E., Vukovic, J., & Burt, D. M. (2014). Developmental changes in children's facial preferences. *Evolution and Human Behavior*, *35*(5), 376–383. https://doi.org/10.1016/j.evolhumbehav.2014.05.002

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, *106*(2), 265–289. https://doi.org/10.1037/0033-2909.106.2.265

Buckingham, G., DeBruine, L. M., Little, A. C., Welling, L. L. M., Conway, C. A., Tiddeman, B. P., & Jones, B. C. (2006). Visual adaptation to masculine and feminine faces influences generalized preferences and perceptions of trustworthiness. *Evolution and Human Behavior*, *27*(5), 381–389. https://doi.org/10.1016/j.evolhumbehav.2006.03.001

Buss, D. M. (1984). Marital assortment for personality dispositions: Assessment with three different data sources. *Behavior Genetics*, *14*(2), 111–123. https://doi.org/10.1007/BF01076408

Buss, D. M., & Schmitt, D. P. (1993). Sexual Strategies Theory: An evolutionary perspective on human mating. *Psychological Review*, *100*(2), 204–232. https://doi.org/10.1037/0033-295X.100.2.204

Caspi, A., Herbener, E. S., & Ozer, D. J. (1992). Shared experiences and the similarity of personalities: A longitudinal study of married couples. *Journal of Personality and Social Psychology*, *62*(2), 281–291. https://doi.org/10.1037//0022-3514.62.2.281

Chambers, V. J., Christiansen, J. R., & Kunz, P. R. (1983). Physiognomic homogamy: A test of physical similarity as a factor in the mate selection process. *Biodemography and Social Biology*, *30*(2), 151–157. https://doi.org/10.1080/19485565.1983.9988529

Charlesworth, D., & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*, *18*(1), 237–268. https://doi.org/10.1146/annurev.es.18.110187.001321

Darwin, Charles. (1871). *The descent of man, and selection in relation to sex: Vol. v.1 (1871)*. London :John Murray,. https://www.biodiversitylibrary.org/item/192632

DeBruine, L. M. (2002). Facial resemblance enhances trust. *Proceedings. Biological Sciences*, *269*(1498), 1307–1312. https://doi.org/10.1098/rspb.2002.2034

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

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

DeBruine, L. M. (2018). *debruine/webmorph: Beta release 2* (v0.0.0.9001) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.1162670

DeBruine, L. M., & Holzleitner, I. J. (2022). *KINSHIP: How do humans recognize kin?* osf.io/5ymzk

DeBruine, L. M., & Jones, B. C. (2007). *Young Adult White Faces with Manipulated Versions* (Version 1). figshare. https://doi.org/10.6084/m9.figshare.4220517.v1

DeBruine, L. M., & Jones, B. C. (2018). *Determining the number of raters for reliable mean ratings*. https://doi.org/10.17605/OSF.IO/X7FUS

DeBruine, L. M., Jones, B. C., Crawford, J. R., Welling, L. L. M., & Little, A. C. (2010). The health of a nation predicts their mate preferences: Cross-cultural variation in women's preferences for masculinized male faces. *Proceedings. Biological Sciences*, *277*(1692), 2405–2410. https://doi.org/10.1098/rspb.2009.2184

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. https://doi.org/10.1007/s10508-007-9266-0

DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2005). Women's attractiveness judgments of self-resembling faces change across the menstrual cycle. *Hormones and Behavior*, *47*(4), 379–383. https://doi.org/10.1016/j.yhbeh.2004.11.006

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

Ekman, P., & Rosenberg, E. L. (Eds.). (2005). *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS), 2nd* *ed.* (pp. xxi, 639). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195179644.001.0001

Feng, D., & Baker, L. (1994). Spouse similarity in attitudes, personality, and psychological well-being. *Behavior Genetics*, *24*(4), 357–364. https://doi.org/10.1007/BF01067537

Fisher, R. A. (1930). *The genetical theory of natural selection.* (pp. xiv, 272). Clarendon Press. https://doi.org/10.5962/bhl.title.27468

Griffiths, R. W., & Kunz, P. R. (1973). Assortative mating: A study of physiognomic homogamy. *Biodemography and Social Biology*, *20*(4), 448–453. https://doi.org/10.1080/19485565.1973.9988075

Gyuris, P., Járai, R., & Bereczkei, T. (2010). The effect of childhood experiences on mate choice in personality traits: Homogamy and sexual imprinting. *Personality and Individual Differences*, *49*(5), 467–472. https://doi.org/10.1016/j.paid.2010.04.021

Halberstadt, J., & Rhodes, G. (2000). The attractiveness of nonface averages: Implications for an evolutionary explanation of the attractiveness of average faces. *Psychological Science*, *11*(4), 285–289. https://doi.org/10.1111/1467-9280.00257

Hamilton, W. D. (1964). The genetical evolution of social behavior, parts I and II. *Journal of Theoretical Biology*, *7*(1), 1–16. https://doi.org/10.1016/0022-5193(64)90038-4

Hansen, B. T., Johannessen, L. E., & Slagsvold, T. (2008). Imprinted species recognition lasts for life in free-living great tits and blue tits. *Animal Behaviour*, *75*(3), 921–927. https://doi.org/10.1016/j.anbehav.2007.07.023

Harrisson, K. A., Magrath, M. J. L., Yen, J. D. L., Pavlova, A., Murray, N., Quin, B., Menkhorst, P., Miller, K. A., Cartwright, K., & Sunnucks, P. (2019). Lifetime Fitness Costs of Inbreeding and Being Inbred in a Critically Endangered Bird. *Current Biology*, *29*(16), 2711-2717.e4. https://doi.org/10.1016/j.cub.2019.06.064

Helgason, A., Pálsson, S., Gudbjartsson, D. F., Kristjánsson, T., & Stefánsson, K. (2008). An association between the kinship and fertility of human couples. *Science (New York, N.Y.)*, *319*(5864), 813–816. https://doi.org/10.1126/science.1150232

Hess, E. H. (1959). Imprinting, an effect of early experience, imprinting determines later social behavior in animals. *Science (New York, N.Y.)*, *130*(3368), 133–141. https://doi.org/10.1126/science.130.3368.133

Hinsz, V. B. (1989). Facial Resemblance in Engaged and Married Couples. *Journal of Social and Personal Relationships*, *6*(2), 223–229. https://doi.org/10.1177/026540758900600205

Holzleitner, I. J., DeBruine, L. M., Lee, A. J., Fasolt, V., O'Shea, K. J., & Shiramizu, V. K. M. (2019). *Image Collection Procedure*. osf.io/gs5wm

Holzleitner, I. J., O'Shea, K. J., Fasolt, V., Lee, A. J., DeBruine, L. M., & Jones, B. C. (2019). *Distribution of facial resemblance in romantic couples suggests both positive and negative assortative processes influence human mate choice*. https://doi.org/10.31234/osf.io/pw5c2

Horwitz, T. B., & Keller, M. C. (2022). A comprehensive meta-analysis of human assortative mating in 22 complex traits. *BioRxiv*. https://doi.org/10.1101/2022.03.19.484997

Jacob, S., McClintock, M. K., Zelano, B., & Ober, C. (2002). Paternally inherited HLA alleles are associated with women's choice of male odor. *Nature Genetics*, *30*(2), 175–179. https://doi.org/10.1038/ng830

Jennions, M. D., & Petrie, M. (1997). Variation in mate choice and mating preferences: A review of causes and consequences. *Biological Reviews of the Cambridge Philosophical Society*, *72*(2), 283–327. https://doi.org/10.1017/s0006323196005014

Jones, B. C., Little, A. C., Penton-Voak, I. S., Tiddeman, B. P., Burt, D. M., & Perrett, D. I. (2001). Facial symmetry and judgements of apparent health: Support for a "good genes" explanation of the attractiveness–symmetry relationship. *Evolution and Human Behavior*, *22*(6), 417–429. https://doi.org/10.1016/S1090-5138(01)00083-6

Kendrick, K. M., Hinton, M. R., Atkins, K., Haupt, M. A., & Skinner, J. D. (1998). Mothers determine sexual preferences. *Nature*, *395*(6699), 229–230. https://doi.org/10.1038/26129

Kocsor, F., Rezneki, R., Juhász, S., & Bereczkei, T. (2011). Preference for facial self-resemblance and attractiveness in human mate choice. *Archives of Sexual Behavior*, *40*(6), 1263–1270. https://doi.org/10.1007/s10508-010-9723-z

Kohn, L. A. P. (1991). The Role of Genetics in Craniofacial Morphology and Growth. *Annual Review of Anthropology*, *20*, 261–278. JSTOR.

Kokko, H. (2001). Fisherian and "good genes" benefits of mate choice: How (not) to distinguish between them. *Ecology Letters*, *4*(4), 322–326. https://doi.org/10.1046/j.1461-0248.2001.00224.x

Kokko, H., Brooks, R., McNamara, J. M., & Houston, A. I. (2002). The sexual selection continuum. *Proceedings. Biological Sciences*, *269*(1498), 1331–1340. https://doi.org/10.1098/rspb.2002.2020

Krupp, D. B., DeBruine, L. M., Jones, B. C., & Lalumière, M. L. (2012). Kin recognition: Evidence that humans can perceive both positive and negative relatedness. *Journal of Evolutionary Biology*, *25*(8), 1472–1478. https://doi.org/10.1111/j.1420-9101.2012.02553.x

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lacy, R. C., & Sherman, P. W. (1983). Kin Recognition by Phenotype Matching. *The American Naturalist*, *121*(4), 489–512. JSTOR.

Lakens, D., & DeBruine, L. M. (2020). *Improving Transparency, Falsifiability, and Rigour by Making Hypothesis Tests Machine Readable*. PsyArXiv. https://doi.org/10.1177/2515245920970949

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), 363–369. https://doi.org/10.1037/0012-1649.23.3.363

Little, A. C., Burt, D. M., & Perrett, D. I. (2006). Assortative mating for perceived facial personality traits. *Personality and Individual Differences*, *40*(5), 973–984. https://doi.org/10.1016/j.paid.2005.09.016

Little, A. C., DeBruine, L. M., & Jones, B. C. (2005). Sex-contingent face after-effects suggest distinct neural populations code male and female faces. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1578), 2283–2287. https://doi.org/10.1098/rspb.2005.3220

Little, A. C., Penton-Voak, I. S., Burt, D. M., & Perrett, D. I. (2003). Investigating an imprinting-like phenomenon in humans: Partners and opposite-sex parents have similar hair and eye colour. *Evolution and Human Behavior*, *24*(1), 43–51. https://doi.org/10.1016/S1090-5138(02)00119-8

Lorenz, K. Z. (1937). The Companion in the Bird's World. *The Auk*, *54*(3), 245–273. https://doi.org/10.2307/4078077

Lorusso, L., Brelstaff, G., & Grosso, E. (2011). A new model for the visual judgement of kinship. *2011 IEEE 10th IVMSP Workshop: Perception and Visual Signal Analysis*, 75–80. https://doi.org/10.1109/IVMSPW.2011.5970358

Lorusso, L., Pulina, L., & Grosso, E. (2015). The Measure of Perceived Similarity Between Faces: Old Issues for a New Method. *Review of Philosophy and Psychology*, *6*(2), 317–339. https://doi.org/10.1007/s13164-014-0229-9

Lundström, J. N., Wysocki, C. J., Olsson, M. J., Preti, G., & Yamazaki, K. (2008). Biological basis of the third-cousin crush. *Science (New York, N.Y.)*, *320*(5880), 1160–1161. https://doi.org/10.1126/science.320.5880.1160

Luo, S. (2017). Assortative mating and couple similarity: Patterns, mechanisms, and consequences. *Social and Personality Psychology Compass*, *11*(8), e12337. https://doi.org/10.1111/spc3.12337

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

Marcinkowska, U. M., & Rantala, M. J. (2012). Sexual Imprinting on Facial Traits of Opposite-Sex Parents in Humans. *Evolutionary Psychology*, *10*(3), 147470491201000320. https://doi.org/10.1177/147470491201000318

Markey, P. M., & Markey, C. N. (2007). Romantic ideals, romantic obtainment, and relationship experiences: The complementarity of interpersonal traits among romantic partners. *Journal of Social and Personal Relationships*, *24*(4), 517–533. https://doi.org/10.1177/0265407507079241

Mascie-Taylor, C. G. N. (1989). Spouse similarity for IQ and personality and convergence. *Behavior Genetics*, *19*(2), 223–227. https://doi.org/10.1007/BF01065906

Murstein, B. I. (1972). Physical attractiveness and martial choice. *Journal of Personality and Social Psychology*, *22*(1), 8–12. https://doi.org/10.1037/h0032394

Peer, K., & Taborsky, M. (2005). Outbreeding depression, but no inbreeding depression in haplodiploid ambrosia beetles with regular sibling mating. *Evolution*, *59*(2), 317–323. https://doi.org/10.1111/j.0014-3820.2005.tb00992.x

Perrett, D. I., Penton-Voak, I. S., Little, A. C., Tiddeman, B. P., Burt, D. M., Schmidt, N., Oxley, R., Kinloch, N., & Barrett, L. (2002). Facial attractiveness judgements reflect learning of parental age characteristics. *Proceedings. Biological Sciences*, *269*(1494), 873–880. https://doi.org/10.1098/rspb.2002.1971

Pomiankowski, A., Iwasa, Y., & Nee, S. (1991). The evolution of costly mate preferences I. Fisher and biased mutation. *Evolution; International Journal of Organic Evolution*, *45*(6), 1422–1430. https://doi.org/10.1111/j.1558-5646.1991.tb02645.x

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rantala, M. J., & Marcinkowska, U. M. (2011). The role of sexual imprinting and the Westermarck effect in mate choice in humans. *Behavioral Ecology and Sociobiology*, *65*(5), 859–873. https://doi.org/10.1007/s00265-011-1145-y

Ro, C. (2015). 9 Celebrity Couples Who Look Like Each Other. *BuzzFeed*. https://www.buzzfeed.com/crystalro/celebrity-couples-who-share-the-same-face

Sanchez-Andrade, G., & Kendrick, K. M. (2009). The main olfactory system and social learning in mammals. *Behavioural Brain Research*, *200*(2), 323–335. https://doi.org/10.1016/j.bbr.2008.12.021

Saxton, T. K. (2016). Experiences during specific developmental stages influence face preferences. *Evolution and Human Behavior*, *37*(1), 21–28. https://doi.org/10.1016/j.evolhumbehav.2015.06.001

Schwartz, C., & Graf, N. (2009). Assortative matching among same-sex and different-sex couples in the United States, 1990-2000. *Demographic Research*, *21*(28), 843–878.

Shanon, B. (1988). On the similarity of features. *New Ideas in Psychology*, *6*, 307–321. https://doi.org/10.1016/0732-118X(88)90043-8

Slagsvold, T., Hansen, B. T., Johannessen, L. E., & Lifjeld, J. T. (2002). Mate choice and imprinting in birds studied by cross-fostering in the wild. *Proceedings. Biological Sciences*, *269*(1499), 1449–1455. PubMed. https://doi.org/10.1098/rspb.2002.2045

Spuhler, J. N. (1968). Assortative mating with respect to physical characteristics. *Eugenics Quarterly*, *15*(2), 128–140. https://doi.org/10.1080/19485565.1968.9987763

Štěrbová, Z., Bártová, K., Martinec Novakova, L., Varella, M., Havlicek, J., & Varella Valentova, J. (2017). Assortative mating in personality among heterosexual and male homosexual couples from Brazil and the Czech Republic. *Personality and Individual Differences*, *112*, 90–96. https://doi.org/10.1016/j.paid.2017.02.036

Štěrbová, Z., & Valentová, J. (2012). Influence of homogamy, complementarity, and sexual imprinting on mate choice. *Anthropologie – International Journal of the Science of Man*, 47–59.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. Cambridge University Press. http://www.loc.gov/catdir/enhancements/fy1004/80041200-t.html

Thiessen, D., & Gregg, B. (1980). Human assortative mating and genetic equilibrium: An evolutionary perspective. *Ethology and Sociobiology*, *1*(2), 111–140. https://doi.org/10.1016/0162-3095(80)90003-5

Thiessen, D., Young, R. K., & Delgado, M. (1997). Social pressures for assortative mating. *Personality and Individual Differences*, *22*(2), 157–164. https://doi.org/10.1016/S0191-8869(96)00181-X

Tramm, N. A., & Servedio, M. R. (2008). Evolution of mate-choice imprinting: Competing strategies. *Evolution; International Journal of Organic Evolution*, *62*(8), 1991–2003. https://doi.org/10.1111/j.1558-5646.2008.00419.x

Trivers, R. (1972). Parental Investment and Sexual Selection. In *Sexual Selection and the Descent of Man* (p. 378).

Vakirtzis, A. (2011). Mate Choice Copying and Nonindependent Mate Choice: A Critical Review. *Annales Zoologici Fennici*, *48*(2), 91–107. https://doi.org/10.5735/086.048.0202

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *43*(2), 161–204. https://doi.org/10.1080/14640749108400966

Watson, D., Klohnen, E. C., Casillas, A., Simms, E. N., Haig, J., & Berry, D. S. (2004). Match makers and deal breakers: Analyses of assortative mating in newlywed couples. *Journal of Personality*, *72*(5), 1029–1068. https://doi.org/10.1111/j.0022-3506.2004.00289.x

Westermarck, E. (1921). *The history of human marriage, 3 vols., 5th ed.* Macmillan.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wincenciak, J., Fincher, C. L., Fisher, C. I., Hahn, A. C., Jones, B. C., & DeBruine, L. M. (2015). Mate choice, mate preference, and biological markets: The relationship between partner choice and health preference is modulated by women's own attractiveness. *Evolution and Human Behavior*, *36*(4), 274–278. https://doi.org/10.1016/j.evolhumbehav.2014.12.004

Wiszewska, A., Pawlowski, B., & Boothroyd, L. G. (2007). Father-daughter relationship as a moderator of sexual imprinting: A facialmetric study. *Evolution and Human Behavior.*, *28*(4), 248–252.

Wong, Y. K., Wong, W. W., Lui, K. F. H., & Wong, A. C.-N. (2018). Revisiting facial resemblance in couples. *PLoS ONE*, *13*(1). https://doi.org/10.1371/journal.pone.0191456

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

Zajonc, R. B., Adelmann, P. K., Murphy, S. T., & Niedenthal, P. M. (1987). Convergence in the physical appearance of spouses. *Motivation and Emotion*, *11*(4), 335–346. https://doi.org/10.1007/BF00992848

Zietsch, B. P., Verweij, K. J. H., Heath, A. C., & Martin, N. G. (2011). Variation in human mate choice: Simultaneously investigating heritability, parental influence, sexual imprinting, and assortative mating. *The American Naturalist*, *177*(5), 605–616. https://doi.org/10.1086/659629

Appendices

Appendix A

Summary of methodologies from a selection of previous studies testing homogamy.

Citation	Testing	Method	N	Materials	Measures	Analysis	Conclusion
(Caspi et al., 1992)	Shared environmental influences	**Self-reported data**	165 married couples from KLS dataset	Self-reported questionnaires	Study of values Views on ideal marriage	Factor analyses	Shared environmental experiences contribute towards similarity
(Gyuris et al., 2010)	Influence of childhood relationship on similarity	**Self-reported data**	294 individuals from 49 couples and opposite-sex parents	Self-reported questionnaires	Big Five Inventory s-EMBU	Correlations Intraclass correlations	Parental influence on similarity in couples' personality traits
(Little et al., 2003)	Similarity in hair and eye colour	**Self-reported data**	697 individuals	Self-reported traits	Eye colour Hair colour	Spearman's Rho	Supports sexual imprinting hypothesis
(Mascie-Taylor, 1989)	Initial choice or convergence	**Self-reported data**	Unreported - two samples from communities near Cambridge and Oxford	Self-reported questionnaires	WAIS EPI	Correlations Hierarchical multiple regression	Similarity exists at initial choice
(Saxton, 2016)	Influence of childhood relationship on similarity	**Self-reported data**	145 individuals	Self-reported questionnaires	Demographics Hair and eye colour Emotional support	Correlations Regression	Childhood experiences influence mate preferences

Citation	Testing	Method	N	Materials	Measures	Analysis	Conclusion
(Watson et al., 2004)	Similarity across multiple domains	**Self-reported data**	263-276 married couples	Self-reported traits	Demographics Big Five Inventory PANAS Disinhibition scale Ego resiliency scale Emotional expression Adult attachment Religious and political attitudes Values Intelligence Relationship variables	Correlations Confirmatory factors analyses Partial correlations Multiple regressions Hierarchical regression	Similarity exists at initial choice
(Zietsch et al., 2011)	Genetic and family environmental influences	**Self-reported data**	22,861 twins, partners, parents, children and siblings	Self-reported questionnaires	Education Yearly income Religiosity Social attitudes Personality Height and age BMI Length of relationship	Correlations	Local reproductive conditions and mating markets likely account for variation in mate choice

Citation	Testing	Method	N	Materials	Measures	Analysis	Conclusion
(Hinsz, 1989)	Repeated exposure, environmental co-existence, perceptual bias explanations	**Ratings**	48 individuals	Faces of human adults	Similarity	ANOVA	Evidence for repeated exposure but not environmental co-existence or perceptual bias
(Little et al., 2006)	Similarity and convergence	**Ratings**	22 individuals (study 1) 19 individuals (study 2)	Faces of human adults Questionnaires	Attractiveness Masculinity Averageness Big five inventory Perceived age	Correlations	Married couples do not grow more similar physically but do grow more similar in personality
(Marcinkowska & Rantala, 2012)	Sexual imprinting hypothesis	**Ratings** **Self-reported data**	120 individuals	Faces of human adults Self-reported questionnaires	Similarity s-EMBU	Parametric tests Non-parametric correlations	Childhood experiences influence men's mate preferences but cannot conclude support for sexual imprinting
(Murstein, 1972)	Similarity on attractiveness	**Ratings** **Self-reported data**	99 couples + 8 judges (study 1) 98 couples (study 2)	Faces of human adults Questionnaires	Attractiveness	Frequencies of discrepancies	Mating markets likely account for variation in attractiveness

Citation	Testing	Method	N	Materials	Measures	Analysis	Conclusion
(Perrett et al., 2002)	Learning parental traits	**Ratings** **Self-reported data**	83 individuals	Facial composites Self-reported questionnaires	Attractiveness Desired traits	Correlations ANCOVA ANOVA	Parental influence on facial attractiveness judgments
(Thiessen et al., 1997)	Similarity on self-reported and facial traits	**Ratings** **Self-reported data** **Matching task**	59 couples (study 1) 50 judges (study 2)	Faces of human adults Personal questions	Similarity Self and partner traits Correct matches	Correlations	Reciprocal social interactions contribute towards similarity in couples
(Wiszewska et al., 2007)	Sexual imprinting hypothesis	**Ratings** **Self-reported data** **Facialmetrics**	49 women and their fathers	Faces of human adults Self-reported questionnaires	Attractiveness Father relationship Father absenteeism Facialmetrics	Principal component analysis	Childhood experiences influence mate preferences
(DeBruine et al., 2009)	Kin recognition	**Ratings** **Judgments**	118 individuals	Faces of human adults	Similarity Kinship	Correlations Likelihood analyses Signal detection analyses ANOVA	Similarity conveys some kinship information but sex and age cue non-kin information in adult faces

Citation	Testing	Method	N	Materials	Measures	Analysis	Conclusion
(Lorusso et al., 2011)	Kinship, similarity and dissimilarity	**Ratings Judgments**	19 individuals (exp 1) 130 individuals (exp 2)	Faces of human adults	Kinship Similarity Dissimilarity	Median and robust standard deviation Two-sample Wilcoxon test	Similarity and dissimilarity are not opposite ends of the same concept and visual processing of faces depends on task and stimuli
(Maloney & Dal Martello, 2006)	Kin recognition	**Ratings Judgments**	64 individuals	Faces of human children	Similarity Kinship	Correlations Linear regression	Similarity conveys much kinship information in child faces and sex and age are ignored
(Wong et al., 2018)	Perceived similarity, attractiveness, personality and age	**Ratings Rankings**	51 individuals (exp 1) 60 individuals (exp 2) 60 individuals (exp 3)	Faces of human adults	Perceived similarity Perceived attractiveness Perceived personality Perceived age Likelihood of being couple	Independent-samples t-test Bootstrapping analyses	Facial similarity incorporates perceptions of personality and age
(Bereczkei et al., 2002)	Sexual imprinting hypothesis	**Matching task**	52 individuals	Faces of human adults Self-reported questionnaires	Correct matches s-EMBU	Wilcoxon signed-rank test Regression	Supports sexual imprinting hypothesis

Citation	Testing	Method	N	Materials	Measures	Analysis	Conclusion
(Bereczkei et al., 2004)	Sexual imprinting hypothesis	**Matching task**	242 individuals	Faces of human adults Self-reported questionnaires	Correct matches s-EMBU	Wilcoxon signed-rank test Regression	Supports sexual imprinting hypothesis
(Chambers et al., 1983)	Physiognomic homogamy	**Matching task**	28 individuals	Faces of human adults	Correct matches	Chi-square	Similarity exists at initial choice
(Griffiths & Kunz, 1973)	Initial choice or convergence	**Matching task**	295 individuals	Faces of human adults	Correct matches	Chi-square	Similarity exists at initial choice
(DeBruine et al., 2011)	Similarity in kinship and mate choice contexts	**2AFC**	146 women	Faces of self-resembling and non-resembling morphs	Trustworthiness Attractiveness	ANOVA	Responses to facial resemblance differ in kin and mate choice contexts
(DeBruine, 2004)	Similarity in kinship and mate choice contexts	**2AFC**	108 individuals (exp 1) 78 individuals (exp 2)	Faces of self-resembling and non-resembling morphs	Attractiveness Averageness	One-sample t-test ANOVA	Responses to facial resemblance differ in kin and mate choice contexts
(DeBruine, 2005)	Similarity in kinship and mate choice contexts	**2AFC**	144 individuals	Faces of self-resembling and non-resembling morphs	Trustworthiness Attractiveness long-term Attractiveness short-term	ANOVA	Responses to facial resemblance differ in kin and mate choice contexts

Citation	Testing	Method	N	Materials	Measures	Analysis	Conclusion
(DeBruine, 2002)	Effect of similarity on trust	**Trust game**	24 individuals	Faces of self-resembling and non-resembling morphs	P1 Trust decision P2 Selfishness decision	ANOVA	Self-resemblance increases prosocial attributions of trust
(Kocsor et al., 2011)	Similarity and attractiveness	**Rankings**	44 individuals	Faces of self-resembling and non-resembling morphs	Attractiveness	Friedman tests Pairwise comparisons Between-sex Mann-Whitney U	Preferences for attractiveness outweigh self-resemblance but both contribute towards mate choice
(Spuhler, 1968)	Similarity on physical traits	**Physical measurements**	1-27 population samples	Physical measurements	Physical measurements	Correlations	Weak correlation on physical traits between couples

Appendix B

Power analysis for Study 2, ` Facial similarity, sibling judgments and couple judgments: A replication study using actual couples and random pairs'.

1. Background

1.1 Sample size

This is a power simulation for a study investigating couple and sibling judgments on a set of 270 face pairs, consisting of 135 actual couples and 135 foil pairs.
Stimuli will be subdivided into 3 groups of 90 face pairs (45 actual couples and 45 foil pairs).
Each group of face pairs will be judged by n=20 participants for each judgment task (couple/sibling judgment task), i.e.:
* couple judgment task, target n = 60
* sibling judgment task, target n = 60.

1.2 Variables simulated

IV: Similarity rating scores
Similarity scores will be simulated by extracting the random intercepts from a dataset of 25 previously collected similarity scores.
The sample similarity scores were made on 139 face pairs of actual couples from the same dataset of stimuli that will be used in this study (i.e. same collection procedure).
The similarity ratings, ranging from 0 (not similar) to 10 (similar), consisted of 25 participants, of which 12 were female and 13 were male and aged between 19 – 40 years old (Holzleitner, O'Shea, et al., 2019).

DV: Couple judgment scores
60 fake participants' couple judgment scores will be simulated, powering for half the effect size found in the exploratory study https://osf.io/uey7s/.

DV: Sibling judgment scores
60 fake participants' sibling judgment scores will be simulated, powering for half the effect size found in the exploratory study https://osf.io/uey7s/.

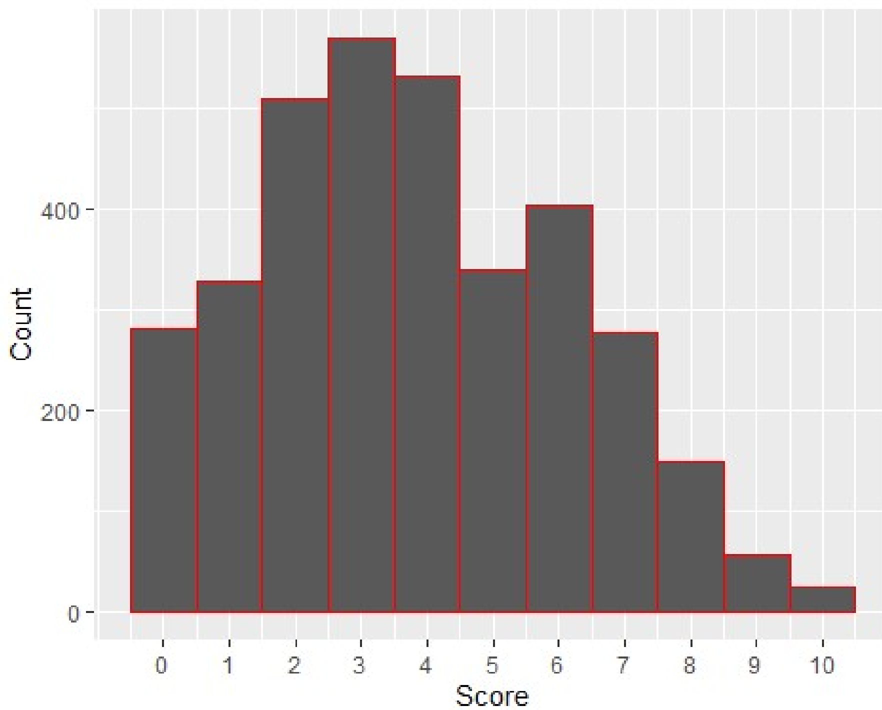
2. Load data

Similarity scores (n = 25) obtained from previously collected data.

```
data <- read_csv("data/Default Query for exp_3640.csv") %>%
  gather("pair_id", "score", t1:t139)
```

2.1 Visualise distribution of similarity scores

```
data %>%
  ggplot() +
  geom_histogram(aes(score), binwidth = 1, color = "red") +
  scale_x_continuous(breaks = 0:10) +
  xlab("Score") +
  ylab("Count")
```



3. Generate similarity scores

3.1 Extract random intercepts

Similarity scores are analysed in a linear mixed effects model (model).

Random intercepts by stimulus pair (pair_i) are stored to be used as similarity scores in the simulation.

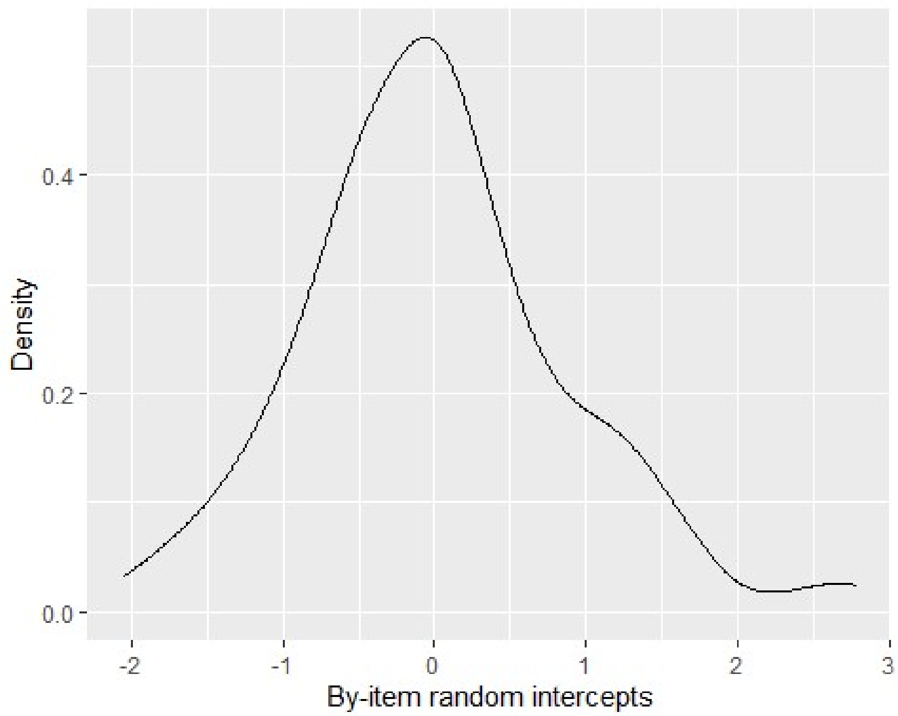
Random intercepts by user (user_i) are stored to be included as random effect by user in the simulation.

```
model <- lmer(score ~ 1 + (1 | user_id) + (1 | pair_id),
              data = data)

pairs <- ranef(model)$pair_id %>%
  rownames_to_column(var = "pair_id") %>%
  rename(pair_i = `(Intercept)`)

users <- ranef(model)$user_id %>%
  rownames_to_column(var = "user_id") %>%
```

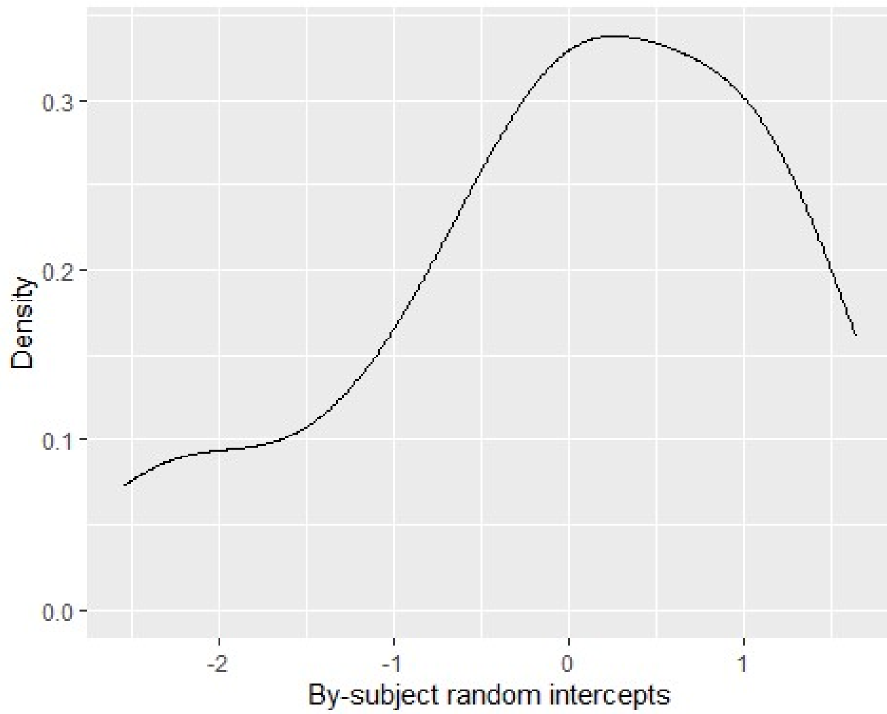
```
  rename(user_i = `(Intercept)`)

grand_i <- fixef(model) %>% pluck()
```

3.2 Visualise distribution of random intercepts by user and stimulus pair for similarity ratings

```
ggplot(pairs) + geom_density(aes(pair_i))+
  xlab("By-item random intercepts") +
  ylab("Density")
```



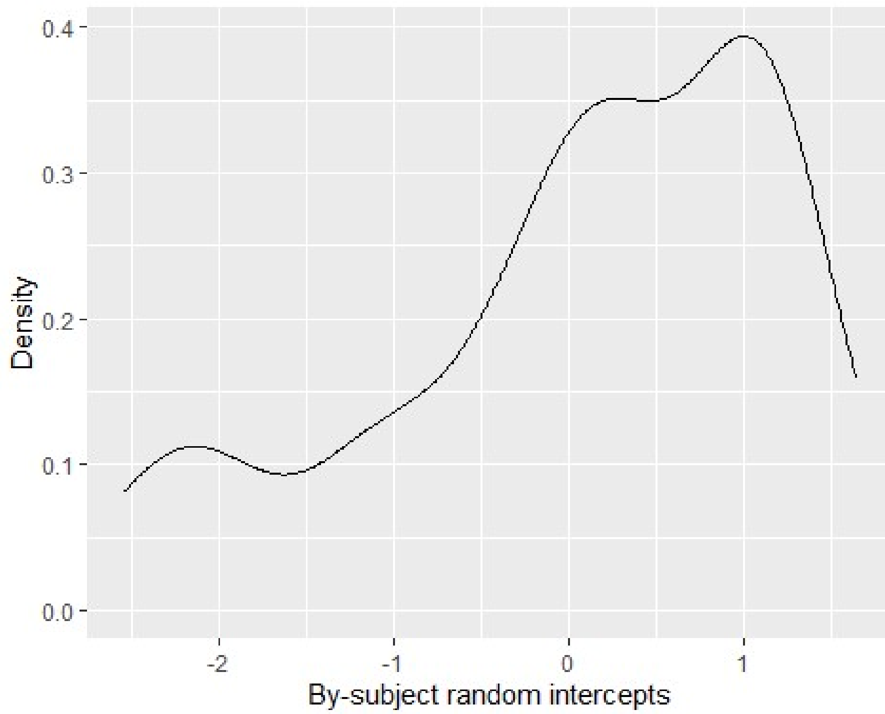

```
ggplot(users) + geom_density(aes(user_i)) +
  xlab("By-subject random intercepts") +
  ylab("Density")
```



3.3 Generate sample raters

Power analysis will simulate 20 raters for each of the three groups of stimuli for both tasks (couple and sibling judgment tasks). Participants recruited for the similarity rating task is not accounted for below as it is justified that a minimum of 15 raters is required to obtain reliable mean stimulus ratings (IV). Therefore, the number of raters that we need to simulate for the judgment tasks is 20 raters X 3 groups of stimulus pairs X 2 tasks = 120 raters.

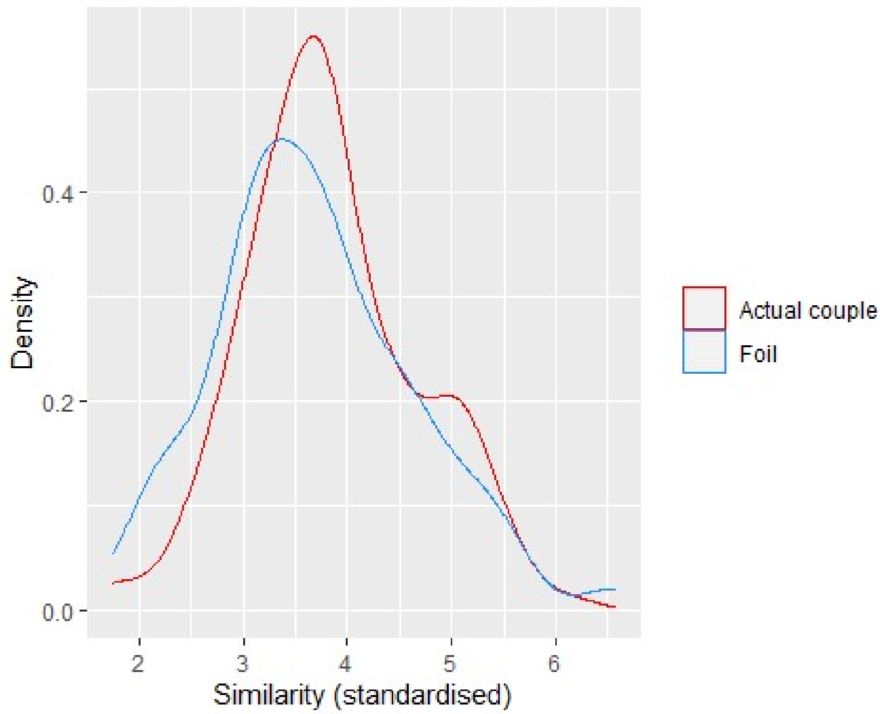
```
rater_n <- 20 #raters per group
group_n <- 3 #number of groups of stimulus pairs

raters <- expand.grid(
  rater_id = 1:rater_n,
  group_id = 1:group_n, #assign raters to one of three groups
  rating = c("sib", "couple")
) %>%
  mutate(
    rater_id = 1:nrow(.),
    user_i = sample(users$user_i, nrow(.), replace = TRUE) #sample rando
m effect of user from similarity scores random intercepts distribution
  )

ggplot(raters) + geom_density(aes(user_i)) +
  xlab("By-subject random intercepts") +
  ylab("Density")
```



3.4 Generate sample pairs

Simulate 3 groups of stimulus pairs to contain 45 real couple pairs and 45 foil pairs each.

```
pair_n <- 45

stim <- expand.grid(
  pair_id = 1:pair_n,
  pair_type = c("actual couple", "foil"),
  group_id = 1:group_n
) %>%
  mutate(
    pair_id = 1:nrow(.),
    sim = grand_i + sample(pairs$pair_i, nrow(.), replace = TRUE)
  )

stim %>%
  mutate(pair_type = recode(pair_type, "actual couple" = "Actual couple"
, "foil" = "Foil")) %>%
  ggplot() + geom_density(aes(sim, color = pair_type)) +
  xlab("Similarity (standardised)") +
  ylab("Density") +
  scale_color_manual(values = c("red","dodgerblue")) +
  theme(legend.title = element_blank())
```



4. Simulation

4.1 Cross raters and pairs

In the actual study, raters will be randomly allocated to one of the 3 groups of stimulus sets.

The simulated data reflects this setup by allocating a group ID (group_id) to raters and to stimulus pairs and matching them via their assigned group ID.

The result is a simulated dataset of 120 raters on 270 stimulus pairs, i.e. 10800 observations.

```
sim_power <- function(n) {
  pair_n <- 45 #number of pairs for each pair type (couple/foil)

  # powering for ~ half the effect sizes from the exploratory study (htt
ps://osf.io/uey7s/)
  int   <-  0    # (Intercept)
  pt    <-  0.1  # pair_type.e
  rt    <-  0.15 # rating.e
  s     <-  0.3  # sim
  rtpt  <-  0.3  # rating.e:pair_type.e
  rts   <- -0.2  # rating.e:sim (similarity increases sib judgments, not
couple judgments)
  spt   <-  0    # sim:pair_type.e
  rtspt <-  0    # rating.e:sim:pair_type.e

  stim <- expand.grid(
    pair_id = 1:pair_n,
    pair_type = c("couple", "foil"),
    group_id = 1:group_n
```

```
  ) %>%
    mutate(
      pair_id = 1:nrow(.),
      sim = grand_i + sample(pairs$pair_i, nrow(.), replace = TRUE),
      sim.s = (sim - mean(sim)) / (sd(sim)),
      pair_i = rnorm(nrow(.), 0, 1)
    )

  fake.data <- raters %>%
    left_join(stim, by = "group_id") %>%
    mutate(
      pair_type.e = recode(pair_type, "couple" = 0.5, "foil" = -0.5),
      rating.e = recode(rating, "sib" = -0.5, "couple" = 0.5),
      # calculate continuous (underlying) score
      x = int + user_i + pair_i +
        sim.s*s + pair_type.e*pt + rating.e*rt +
        rating.e*pair_type.e*rtpt + rating.e*sim.s*rts + sim.s*pair_type
.e*spt +
        rating.e*sim.s*pair_type.e*rtspt,
      prob = exp(x)/(1+exp(x)), # logit transform
      score = rbinom(nrow(.), 1, prob = prob)
    )

  # # sense check of fake data simulation
  # fake.data %>%
  # group_by(pair_type, rating) %>%
  # summarise(
  # mean = mean(score),
  # r = cor(score, sim)
  # )
  # ggplot(fake.data, aes(x, color = rating)) +
  # geom_density() +
  # facet_grid(pair_type~.)

  fake.model <- glmer(
    score ~ rating.e * sim.s * pair_type.e +
      (1 + sim.s * pair_type.e || rater_id) +
      (1 + rating.e || pair_id),
    family = binomial,
    data = fake.data
  )

  broom.mixed::tidy(fake.model, "fixed")
}
```

5. Run simulation

The simulation loops 100 times in the next code chunk and saved as a csv file.

Run this chunk 10 times, updating the file name each time to reflect trial number ‘t_.csv’.

This is done to avoid Rstudio crashing from processing 1000 trials at one go.

```
power <- purrr::map_df(1:100, sim_power)

write_csv(power, "power_2019-08-23_t1.csv") #update trial number before
each run
```

6. Establish power

The 10 trial runs are combined into one dataset and average power is calculate for each effect.

```
power <- list.files(pattern = "power_2019-08-.*\\.csv") %>%
  lapply(read_csv, col_types = cols()) %>%
  do.call("rbind", .)

power %>%
  group_by(term) %>%
  summarise(n = n(),
            effect_size = mean(estimate),
            power = mean(p.value < .05))

## # A tibble: 8 x 4
##   term                            n effect_size power
##   <chr>                       <int>       <dbl> <dbl>
## 1 (Intercept)                  1000    -0.0686  0.025
## 2 pair_type.e                  1000     0.108   0.122
## 3 rating.e                     1000     0.251   0.082
## 4 rating.e:pair_type.e         1000     0.300   0.889
## 5 rating.e:sim.s               1000    -0.200   0.988
## 6 rating.e:sim.s:pair_type.e  1000    -0.00266 0.05
## 7 sim.s                        1000     0.303   0.994
## 8 sim.s:pair_type.e            1000     0.00482 0.043
```