Facial resemblance enhances trust

Lisa M. DeBruine

Department of Psychology, McMaster University, Hamilton, Ontario, Canada L8S 4K1 (debruilm@mcmaster.ca)

Organisms are expected to be sensitive to cues of genetic relatedness when making decisions about social behaviour. Relatedness can be assessed in several ways, one of which is phenotype matching: the assessment of similarity between others’ traits and either one’s own traits or those of known relatives. One candidate cue of relatedness in humans is facial resemblance. Here, I report the effects of an experimental manipulation of facial resemblance in a two-person sequential trust game. Subjects were shown faces of ostensibly playing partners manipulated to resemble either themselves or an unknown person. Resemblance to the subject’s own face raised the incidence of trusting a partner, but had no effect on the incidence of selfish betrayals of the partner’s trust. Control subjects playing with identical pictures failed to show such an effect. In a second experiment, resemblance of the playing partner to a familiar (famous) person had no effect on either trusting or betrayals of trust.

Keywords: kin recognition; phenotype matching; trust; facial resemblance; humans

1. INTRODUCTION

Humans undoubtedly learn who their relatives are by association and verbal communication, but these methods are not equally reliable for all classes of kin. Identification of one’s mother and maternal siblings can be achieved by association, but distinguishing maternal half-siblings from full siblings, identifying one’s father, or even recognizing offspring (for males), may require additional mechanisms such as phenotype matching (Porter 1987; Pfennig & Sherman 1995; Hauber & Sherman 2001). Phenotype matching refers to an implicit evaluation of relatedness on the basis of some trait-based assessment of phenotypic similarity. This assessment of similarity may be either with reference to one’s own phenotype (self-referent phenotype matching) or with reference to the phenotypes of individuals conveying other cues of kinship, such as the patterns of social association characteristic of close family members. Evidence of phenotype matching has been found in several species, including ground squirrels (Holmes & Sherman 1982), baboons (Alberts 1999), golden hamsters (Mateo & Johnston 2000) and rhesus monkeys (Meikle & Vessey 1981), although seldom on the basis of visual cues. People, however, certainly attend to familial resemblances in appearance, especially on the paternal side (Daly & Wilson 1982; Regalski & Gaulin 1993), making facial resemblance a likely candidate for phenotype matching.

Digital morphing techniques provide an opportunity to manipulate facial resemblance experimentally and assess its effects on behaviour. Digital morphing allows photographs of two faces to be melded into a realistic virtual face by delineating corresponding points located at several standard landmarks (such as the corners of the eyes and the mouth) on each of the source faces and generating a new face (the ‘morph’) from a weighted average of landmark-specific parameters of the sources. Morphing may entail the averaging of both shape and colour information from the source faces, or just shape alone. In shape-only morphing (also called warping), the morph retains the coloration of one specified source face. Otherwise, the colour values of each pixel from matching locations on the two faces are also averaged to produce the morph. Morphs generally have smoother-looking skin and are more symmetrical than source faces, but appear realistic. In the present study, all experimental (self-similar) and control faces were morphs, so these effects of morphing were not confounded with cues of resemblance. Of course, facial averaging by digital morphing may not produce resemblances that are identical to those resulting from actual relatedness, but this consideration should only work against the hypothesis that experimental manipulation of resemblance in morphs will affect social responses as predicted. There has been little research on whether and how family members resemble each other. Unrelated people can detect resemblance between parents and infants or children, but their accuracy is far from perfect (Nesse et al. 1990; Brédart & French 1999; McLain et al. 2000; Bressan & Dal Martello 2002). Understandably, even less has been done on the ability to assess resemblance of one’s relatives to one’s self, although Porter et al. (1984) found that, after only 15 min to 20 h (average 4.7 h) of post-natal contact, 22 of 24 mothers could correctly identify their infants from sets of four photographs of infants matched for sex, age, race and amount of hair, and many of the mothers indicated that a family resemblance helped them to identify their child.

If humans use phenotype matching to guide their decisions and actions, they might be expected to exhibit higher levels of prosocial behaviours, such as trust and equitable sharing towards those who resemble them. Economists have developed laboratory analogues of trust and selfish behaviour in the form of sequential bargaining games with monetary incentives. I report on the results of a computer-interactive ‘two-person’ trust game (figure 1), where player 1 (P1) chooses either to dictate an equitable division of a small sum or to trust player 2 (P2) to divide a larger sum equitably, even though the latter has a more selfish option. Standard game-theory models of selfinterested choice suggest that P2 should always act selfishly in an anonymous one-shot game of this sort, and therefore...
P1 should never trust P2. However, experimental studies have shown that P1 often does trust P2 and that P2 often rewards that trust with an unselfish division (Berg et al. 1995; McCabe et al. 1996; Hoffman et al. 1998; Scharrer et al. 2001). The tendency to trust can be manipulated by framing effects, such as labelling the other player with terms implying cooperation or competition: in an experiment by Burnham et al. (2000), the label ‘partner’ produced twice as much trusting behaviour as the label ‘opponent’. In light of this evidence, it is plausible that cues of relatedness will also increase prosocial behaviour in the trust game.

If the ultimate evolutionary reason why pay-offs to oneself have utility is because they enhance expected fitness, then pay-offs to one’s kin should also have utility because of inclusive fitness effects, and a cue of kinship with one’s playing partner should imbue the partner’s pay-offs with utility proportionate to the degree of relatedness, \( r \), that is implied by the cue (Hamilton 1964). In the pay-off structure of the current game, the expected effect of such an evolved psychology would be both to raise the incentive to trust from (Canadian) \$1 to \$(1 + r) and to lower the cost of betrayal from \$1 to \$(1 - r). According to this model of ‘nepotistic’ social motives, people playing P1 should be more willing to trust a P2 who presents a cue of kinship, even if P1’s subjective probability of unselfish reciprocation by P2 is unaffected by the cue (figure 2). Moreover, by similar reasoning, a cue of kinship should reduce P2’s incentive to betray a trusting P1, from \$1 to \$(1 - r).

Figure 1. The trust game tree. Subjects were P1 for half of the games and P2 for the other half. The order of games was chosen randomly and counterbalanced for order of self and non-self morphs. The monetary value of \( X \) ranged from (Canadian) \$2 to \$5 and was balanced between self and non-self morph trials. Subjects were presented with dollar values, not variables, and the arrows were not labelled.

Figure 2. Game-theoretical analysis. The pay-off to P1 who plays ‘not trusting’ is \( X(1 + r) \) and the pay-off to P1 who plays ‘trusting’ is \( u[(X + 1)(1 + r) + (1 - u)] \times [(X - 1) + r(X + 2)] \). The shaded area of the graph represents all combinations of \( r \) and \( u \) where ‘trusting’ results in a greater average pay-off to P1, i.e. when \( r > (1 - 2u)/(2 - u) \). The pay-off to P2 who plays ‘selfish’ is \( (X + 2) + r(X - 1) \) and the pay-off to P2 who plays ‘unselfish’ is \( (1 + r)(X + 1) \). Playing ‘unselfish’ results in a greater pay-off to P2 than does playing ‘selfish’ only when \( r > 0.5 \). (\( X \), the pay-off if P1 terminates the game at the first move; \( r \), relatedness between P1 and P2; \( u \), the probability of P2 playing ‘unselfish’.)

2. METHODS

(a) Preparation of facial stimuli

To create a cue of kinship, facial resemblance was manipulated using digital morphing techniques to combine same-sex faces of persons unknown to the subject (called ‘unknown’ faces) with either the subject’s own face (‘self’) or another unfamiliar same-sex face (‘non-self’).

The unknown and non-self faces were obtained by recruiting volunteers from another Ontario university. Photographs (neutral expression and centred) of 24 male and 24 female Canadian students between the ages of 19 and 25 (mean age 21 years) were standardized for interpupillary distance and cropped to a standard size (320 \times 400 pixels). Clothing and facial jewellery were removed digitally (using Adobe Photoshop Deluxe 2.0). The same procedure was used to prepare the images of the 24 research participants (‘subjects’), who ranged in age from 18 to 24 (mean age 21 years). Photographs of the subjects were taken prior to the experiment, on the pretext that they were needed in order to show the subjects’ faces to their playing partners at other universities.

Colour morphing of ‘white’ with ‘non-white’ persons creates conspicuously artificial images, but students at McMaster University who volunteer for research participation come from a diversity of ethnic backgrounds and it was not practical or fair to limit participation to Caucasians. Morphing bearded men also engenders an artificial appearance. For these reasons, two types of morphs were prepared: shape–colour morphs and shape-only morphs (figure 3). The shape–colour
Facial resemblance enhances trust

L. M. DeBruine

**Figure 3.** Sample morphs. Shape–colour morphs (a) were made by combining 40% of the shape and colour information from the self or non-self face with 60% of an unknown face to make the shape–colour morph. Note that the morph retains all of the unknown person’s hair colour and style. Shape-only morphs (b) were made by combining 50% of the shape information from both faces and 100% of the colour information from the unfamiliar face. Note that the faces are standardized for interpupillary distance, which alters head size as a function of this distance.

morphs were based on pictures of Caucasians and non-bearded men \( n = 11 \), while the shape-only morphs were based on images of non-white persons \( n = 9 \) and men with facial hair \( n = 4 \). Furthermore, combining colour information from different hairstyles led to unnatural-looking hair for the shape–colour morphs, so colour information in the hair area was always limited to that of the unknown face. Shape–colour morphs consisted of 40% of the self or non-self face and 60% of the unknown face, while shape-only morphs used 50% of each. These values were set at the highest level at which subjects in a pilot study did not spontaneously detect resemblance to self. Each unknown face served as a base for self morphs and non-self morphs an equal number of times, distributed equally across trials.

(b) **The trust game**

The particular trust game used in this experiment is a two-node extensive form game (figure 1), meaning that there are only two decisions that could be made and that the decisions are made in sequence, not simultaneously. The first node represents a decision by P1 to trust or not trust P2 to split a sum of money. The second node represents a decision by the second player to respond to P1’s trust selfishly or unselfishly.

In the experiment, 24 subjects played 16 rounds of this bargaining game in sessions with one to four players at individual computer stations. Subjects played with what were ostensibly 16 different playing partners on-line at other universities, but in fact played against programmed choices associated with displayed facial morphs. The subjects’ choices were scored as trusting or not trusting when playing P1 (six rounds), and unselfish or selfish when playing P2 (six rounds). In the remaining four rounds, subjects were assigned the P2 role, but P1 was programmed to make a non-trusting move and terminate the game; these rounds generated no choice data, but were included to provide realism. Orthogonal to the above distinctions was the resemblance cue, with half of the rounds played against self morphs and half against non-self morphs. Thus, each subject could make 0, 1, 2 or 3 prosocial decisions in each of four conditions: trusting self morphs and non-self morphs as P1, and responding unselfishly to trusting self morphs and non-self morphs as P2. To minimize differences in subjects’ responses based on the sequence of roles or the other players’ responses, both were kept constant. The pay-off structure has been shown to affect responses (Bolton & Ockenfels 2000), so pay-off types were balanced equally between self and non-self morphs.

At the end of the session, subjects were paid the appropriate amount from one randomly chosen game, as promised at the beginning of the session. In debriefing interviews after each session, no subject revealed any suspicion that the images had been manipulated nor reported a perception of resemblance to self (although one subject remarked that one of his playing partners looked very much like his brother). Participants evidently
believed that they were playing with real partners via the Internet, often commenting with apparent feeling about how a particular partner had played.

(c) Control conditions

Three control conditions were included in this study.

(i) To control for any peculiarities of the stimuli that may have caused subjects to trust self morphs more than non-self morphs, another 24 ‘race’-matched subjects played exactly the same virtual partners in the same sequence of roles as the players in the experimental condition. Besides providing a matched control for any idiosyncratic influences, this also controlled for effects of race, as any effects caused solely by the race of the experimental subject should also be seen in the same-race control subject.

(ii) To assess the possibility that reactions towards self morphs might be mediated by perceptions of differential attractiveness, all stimulus faces were rated on a 10-point Likert scale (anchor adjectives were 1, ‘very unattractive’ and 10, ‘very attractive’). An independent set of 10 observers at the same pool of research volunteers.

(iii) Finally, to control for the effect of familiarity of self morphs, I repeated the procedure in this experiment using shape-only morphs of famous and non-famous faces in place of self and non-self faces. Eight images of famous and non-famous faces were taken from the Internet and chosen to be approximately equal in attractiveness and picture quality. These faces served in place of the self and non-self faces, respectively, from the main experiment. Each subject played with eight morphs made from one of the same-sex famous faces and eight morphs made from one of the same-sex non-famous faces. After the games were played, subjects rated their familiarity with the original (unmorphed) famous and non-famous faces on a five-point Likert scale. Only subjects who rated the famous faces as familiar (ratings of 4 and 5) and the non-famous faces as unfamiliar (ratings of 1 and 2) were included in the data analyses, although no other criteria for inclusion changed the conclusion.

(d) Statistical analyses

For initial analyses of behaviour towards self morphs versus non-self morphs, the 11 subjects who saw shape–colour morphs and the 13 who saw shape-only morphs were treated as two independent experiments, affording a replication of the basic study with two slightly different methodologies. As the effect on trusting was statistically significant in both cases and the pattern of results was identical (see §3), the data are pooled for further analyses.

To assess whether the effect of self morphs was modulated by sex, race or morph type, a difference score (the number of prosocial responses to self morphs minus the number to non-self morphs) was computed for each subject’s play at P1, and these difference scores were submitted to a three-way analysis of variance (ANOVA).

3. RESULTS

The 11 subjects who saw shape–colour morphs as stimuli were more trusting of self morphs than of non-self morphs when playing P1 (t_{10} = 2.63, p = 0.013), but as P2, they were equally unselfish towards both types of morph (t_{10} = 0.00, n.s.). The 13 subjects who saw shape-only morphs as stimuli behaved similarly, trusting self morphs more than non-self morphs as P1 (t_{12} = 2.13, p = 0.028), but failing to discriminate between self and non-self morphs as P2 (t_{12} = 0.37, n.s.) (figure 4).

Combining the data for all 24 subjects and using a difference score for number of prosocial plays to self-morphs minus non-self morphs, there were no significant effects of sex (F_{1,23} = 0.13, p = 0.73), ‘race’ (white versus non-white; F_{1,23} = 0.83, p = 0.38) or morph type (shape–colour morphs versus shape-only morphs; F_{1,23} = 0.13, p = 0.73), nor were there any statistically significant interactions (all p > 0.50).

There were no statistically significant differences in the behaviour of the matched-control subjects who played the same sets of faces in the same sequence as the experimental subjects. In other words, these control players did not play differentially towards faces that had been self morphs versus non-self morphs for the original subjects, neither in their trusting responses as P1 (t_{23} = −1.50, n.s.) nor in their selfish choices as P2 (t_{23} = −0.35, n.s.) (figure 5).

The attractiveness ratings by the independent set of 10 judges also showed no difference; the faces that had served as self morphs for the original subjects were actually rated slightly less attractive (M = 4.11 ± 1.17 s.d. on a 10-point scale) than the non-self morphs (M = 4.27 ± 1.13 s.d.), but the difference was not significant. Moreover, there was no difference in attractiveness ratings between morphed images that did or did not elicit trust from subjects playing P1 (M_{trusting} = 4.11 ± 1.06, M_{not trusting} = 4.11 ± 1.22, t_{141} = 0.00, n.s.), nor between those that elicited selfish and unselfish moves by subjects playing P2 (M_{selfish} = 4.30 ± 1.18, M_{unselfish} = 4.26 ± 1.20, t_{142} = −0.18, n.s.). If we confine attention to the non-self morphs, there was still no significant difference and those faces that elicited trust were actually rated slightly less attractive than
Facial resemblance enhances trust. L. M. DeBruine

1.0 0.8 0.6 0.4 0.2 0.0 0.2 0.4 0.6 0.8 1.0
average difference

-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1.0

experimental subjects
control subjects

Figure 5. Resemblance effects. Average difference (± s.e.) between number of prosocial responses to self and non-self morphs for experimental and control subjects (both morphing methods combined) (grey bars, trusting; open bars, unselfish). Differences were scored as self minus non-self, so that positive differences indicate more prosocial responses towards self morphs and negative differences indicate more prosocial responses towards non-self morphs. An additional set of control subjects saw the same face stimuli as the experimental subjects. These control players showed no statistically significant differences in either trusting or unselfish behaviour, while experimental subjects were significantly more trusting of self than non-self morphs ($p < 0.005$).

those that did not ($M_{trust} = 4.14 ± 1.04$, $M_{notrust} = 4.28 ± 1.25$, $t = -0.53$, n.s.).

Finally, subjects in control condition (iii) played against morphs based on famous or non-famous faces, without significant effects. In the P1 role, subjects trusted ‘famous’ opponents 1.40 times on average and non-famous opponents 1.25 times ($t_{19} = 0.51$, n.s.). In the P2 role, subjects responded unselfishly to a trusting move by a famous P1 1.70 times and by a non-famous P1 1.80 times ($t_{19} = -0.42$, n.s.).

4. DISCUSSION

Experimental subjects, who believed that they were playing against pictured opponents while unaware that information from their own faces had been incorporated into the ‘morphed’ faces of some of those supposed opponents, trusted opponents who resembled themselves significantly more than they trusted other opponents, but did not reward trusting moves by their opponents differentially. These results were replicated in two independent groups of subjects, using two distinct facial morphing procedures.

These results show that facial resemblance can modulate social behaviour. An enhanced positive inclination towards individuals that resemble self was anticipated on the basis of kin selection theory (Hamilton 1964), but this theory is mute with regard to the mechanisms by which such an inclination is elicited. One possibility is that self morphs are perceived as more attractive (Penton-Voak et al. 1999) by the subjects and that more attractive people are trusted more. However, analysis of the subjects’ game choices in relation to the independent ratings of the attractiveness of the stimulus faces does not support this possibility, as no effect of attractiveness on either trusting or selfishness was seen, even for non-self morphs.

The absence of an effect of familiarity using famous faces indicates that the relevant cue(s) modulating discriminative trusting moves may be more than simply familiarity. However, this is only weak evidence against familiarity as a mediator of the effect of resemblance on trusting behaviour, as the familiarity that one can gain through television and movies may be insufficient or of a different quality from that gained through personal experience.

The fact that self morphs were treated preferentially by subjects playing P1 but not by the same subjects playing P2 may be explicable in terms of the different pay-off structures in the two-person bargaining game. Following the logic of inclusive fitness theory (Hamilton 1964), P1 is expected to be indifferent between trusting and not trusting if the coefficient of relatedness between the players is zero ($r = 0$) and P2’s probability of a selfish betrayal of trust is equal to the probability of reciprocating trust (that is, if $p = 0.5$ for both moves), because the expected pay-off is the same in each case. Any increase in the probability of P2 reciprocating trust or in the utility that P1 derives from P2’s pay-offs as a result of kinship cues would then raise P1’s expected pay-off from trusting above that from not trusting. However, P2’s expected utility from the results of an unselfish decision will exceed that from a selfish decision only if $r > 0.5$ (the equivalent of full siblings). Clearly, subjects did not behave in full accordance with such a model, because it cannot explain why they sometimes rewarded the trust of non-self morphs with unselfish responses. However, in conjunction with the additional considerations needed to explain why non-nepotistic generosity ever occurs, this asymmetry may help to explain why resemblance affected only the subjects’ initial gestures of trust and not the reciprocation of trust by their opponents.

If humans are indeed using phenotype matching to make decisions about social behaviour, several interesting questions are raised. Foremost is whether humans use the self or others as a referent. Our human ancestors did not have mirrors until relatively recently and reflections in water would provide quite degraded information. Using a genetically encoded template of what one’s face looks like seems highly implausible. Hence, the most likely mechanism for phenotype matching is expected to depend on familial referents. The weakness in relying on facial resemblance of an unknown person to that of a relative identified as such by association or family history, such as mother and maternal siblings, is that this mechanism would not provide reliable information about paternal relatives nor allow one to differentiate full siblings from half siblings.

It would be interesting to replicate the present study using ‘genetic’ siblings and adopted unrelated siblings. If people use self-referential phenotype matching, self morphs made from adopted and non-adopted subjects should elicit similar effects on trusting behaviour. If subjects are shown morphs made from their genetic or adopted siblings, the effect on trust should be completely absent or only present for morphs made from genetic siblings, inasmuch as they resemble the subject. However, if humans instead use other kin as referents, the data should
Facial resemblance enhances trust

Self morphs should elicit little effect on trust for adopted subjects, but morphs made from both genetic and adopted siblings should elicit at least as large an effect as self-morphs do for non-adopted subjects. Of course, both types of phenotype matching could occur at the same time, as has been shown to be the case in Belding’s ground squirrels (Holmes & Sherman 1982). Alternatively, only other-referent phenotype matching may occur, but the advent of mirrors has caused humans to experience their own phenotypes as they would experience a sibling’s. Either alternative would result in an effect on trust of morphs made from both self and siblings and would be observed in both adopted and non-adopted subjects. To distinguish this pattern from one resulting from simple familiarity, morphs made from non-family members, such as close friends, should be assessed for their ability to influence trust. To conclude, the results discussed in this paper indicate that facial resemblance is a candidate cue for human visual phenotype matching; this is a first step in assessing whether and how we recognize kin by facial resemblance.

I thank M. Wilson, M. Daly, D. Earn, P. Ramos, P. Cooper, T. Addison, A. Clark, P. Barclay, M. Burke and the Animal Behaviour journal club for their comments and support. Financial support for this research was provided by grants from the Social Science and Humanities Research Council of Canada to M. Wilson and from the National Sciences and Engineering Research Council of Canada to M. Daly.

REFERENCES


