Kin recognition and the perceived facial similarity of children

Laurence T. Maloney
Department of Psychology and Center for Neural Science, New York University, USA

Maria F. Dal Martello
Department of General Psychology, University of Padova, Italy

We examine the connection between a hypothetical kin recognition signal available in visual perception and the perceived facial similarity of children. One group of observers rated the facial similarity of pairs of children portrayed in photographs. Half of the pairs were siblings but the observers were not told this. A second group classified the pairs as siblings or nonsiblings. An optimal Bayesian classifier, given the similarity ratings of the first group, was as accurate in judging siblings as the second group. Mean rated similarity was also an accurate linear predictor ($R^2 = .96$) of the log-odds that the rated pair portrayed were, in fact, siblings. Surprisingly, mean rated similarity did not vary with the age difference or gender difference of the pairs, both of which were counterbalanced across the stimuli. We conclude that the perceived facial similarity of children is little more than a graded kin recognition signal and that this kin recognition signal is effectively an estimate of the probability that two children are close genetic relatives.

Keywords: kin recognition, face perception, similarity, resemblance

Introduction

The term similarity leads a double life. In one sense, it describes aspects of perceptual and cognitive experience: We perceive that objects are more or less similar. Yet the term similarity has a second, more abstract, employment. In this second sense, the term “similarity” serves as an explanatory construct. In the past half century “abstract similarity” has played a central role in ethology and animal learning (Guttman & Kalish, 1956; Sutherland & Mackintosh, 1971) as well as the study of human cognition (Shepard, 1987; Tversky, 1977). Analogues of similarity play an important role in artificial intelligence (Falkenheimer, Forbus, & Gentner, 1989; Riesbeck & Schank, 1989), in statistical classification algorithms (Duda & Hart, 1973), in formal models of decision making (Gilboa & Schmeidler, 2001) and in models of kin recognition (Chapais & Berman, 2004; Fletcher & Michener, 1987; Hepper, 1991). Similarity in this second sense is the conceptual glue that binds experience together: “From causes which appear similar we expect similar effects” (Hume, 1748/1993).

In the kin recognition literature, in particular, researchers assign special importance to hypothetical sensory kin recognition signals that allow humans (Porter & Cernoch, 1991; Porter, Cernoch, & Balogh, 1984) and other animals (Sherman, Reeve, & Pfenning, 1997) to recognize their close genetic relatives and, in animals, to assess the degree of genetic relatedness between themselves and members of their own species (Masters & Forester, 1995; Mateo, 2002). These signals also allow, at least in primates, the assessment of the degree of genetic relatedness between other members of their own species (Cheney & Seyfarth, 1989; Cheney & Seyfarth, 1999). Cheney and Seyfarth (1999, p. 67) find, for example, that monkeys “recognize the close relationships that exist among individuals other than themselves” and tend to attack the relatives of monkeys who have attacked them or their relatives in the past. A comparable ability in humans is presupposed by “social mirror” theories of paternal resemblance which assume that human males make use of the judgments of others in assessing their own resemblance to a possible offspring (Burch & Gallup, 2000; Daly & Wilson, 1982; Regalski & Gaulin, 1993).

In this article we examine the connection between a hypothetical kin recognition signal available in human visual perception and the perceived facial similarity of children. In examining the portrait details taken from Gainsborough’s The Marsham Children (Figure 1), for example, it is hard not to register similarities and dissimilarities between the children portrayed. If challenged to pick the most similar pair, viewers may disagree. There is no “right” answer as there would be if we asked the viewer to pick the youngest child. We do not know what objective measurement or rule captures perceived facial similarity.

There is initially every reason to expect that perceived facial similarity will depend on many factors, including age and gender as well as features that signal genetic relatedness. It would not be surprising to discover that perceived facial similarity depends in part on facial characteristics that signal genetic relatedness. However, we find that the perceived facial similarity of children is little more than a
graded kin recognition signal, unaffected by age or gender difference. We also find that the kin recognition signal captured by perceived facial similarity is equivalent to a visual estimate of the probability that the children are close genetic relatives.

**Similarity**

Similarity is typically modeled as an aggregate measure based on matching of distinct features (Tversky, 1977; Quine, 1969) or dimensional differences (Shepard, 1964). Following Tversky, we will use the term “visual feature” to refer to any measurement that can be performed on the visual input. The visual input to similarity judgments is a pool of visual features (Figure 2A), and the observer, asked to make a similarity judgment, selects a subset of features and combines them by some rule to arrive at a similarity measure. Tversky (1977), for example, proposed that a weighted linear combination of common and distinct “features of similarity” could account for similarity judgments. This sort of linear cue combination model is common in the vision literature (for a review, see Landy, Maloney, Johnston, & Young, 1995).

In making other visual judgments, the observer presumably makes use of subsets of the same common feature pool. Several authors have investigated the facial features that signal gender (Brown & Perrett, 1993; Bruce et al., 1993; O’Toole et al., 1998; Wild et al., 2000). Others have examined the features that signal age or are age invariant (Enlow & Hans, 1996; George & Hole, 1998; Pittenger, Shaw, & Mark, 1979). Researchers have also attempted to determine the features that convey information about health (Jones et al., 2001) or degree of genetic relatedness (Kohn, 1991; Salter, 1996).

Of course, an observer may use different subsets of facial features and different rules of combination in rating similarity and in recognizing kin (as illustrated in 

![Figure 1](http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932838/)

Figure 1. The Marsham Children. The children of Charles Marsham, First Earl of Romney. Details of The Marsham Children by Thomas Gainsborough, 1787 (Staatliche Museen, Berlin, Photo credit: Bildarchiv Preussischer Kulturbesitz/Art Resource, NY).

![Figure 2](http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932838/)

Figure 2. Two models of the flow of visual information in similarity tasks and signal detection tasks. (A) Both similarity and signal detection judgments are based on a subset of available visual features, the feature pool. Features are combined to form measures of similarity, of genetic relatedness, etc. The different measures need not make use of the same rules of combination or even the same visual features, but it is possible that certain visual features will be used in more than one kind of task. The observer asked to provide a similarity rating on a scale of 0 to 10 might respond "7." The same observer asked to judge whether a pair of children are siblings or not might respond “siblings” or “not siblings” based on a computation using a completely different rule. (B) TSO. The computation of a similarity measure based on a subset of available visual features is the basis for both ratings of similarity and signal detection judgments. The similarity measure is compared to a threshold in deciding whether to respond “siblings” or “not siblings.”
It is not implausible, for example, that the most useful features for kin recognition would be roughly invariant with age and of little use in judging age difference and unaffected by gender difference. Yet, if facial similarity is an aggregate measure based on many different facial features some of which may be of use in particular signal detection tasks, then it is plausible that facial similarity measures would carry some, but not all, objective information of use in judging degree of kinship.

It is possible that similarity contains only a fraction of the information available to the observer that is relevant to kin recognition. Conversely, it is possible that similarity contains all of the task-relevant information available to the observer for kin recognition. In Figure 2B, we illustrate a model that leads to the latter outcome. Suppose that the observer carries out a kin recognition task, by first forming a measure of similarity. If the value of the similarity measure exceeds a threshold value, the observer responds “yes” (“related”), otherwise “no” (“not related”). For this thresholded similarity observer (TSO) the similarity measure is effectively a graded kin recognition signal.

Something like this model is typically presupposed in the experimental literature concerning human kin recognition. Experimenters often do not ask observers to judge directly whether two individuals are closely related (the relevant signal detection task). Instead, they ask them to judge the facial similarity of pairs of individuals portrayed in photographs (Bréard & French, 1999; Bressan & Dal Martello, 2002; Christenfeld & Hill, 1995). In their analysis or discussion, they assumed that individuals rated as more similar would be more likely to be related. Our research is the first test of this assumption for pairs of children’s faces (for an analogous experiment using adult–child pairs, see Bressan & Grassi, 2004).

Notation

In the experiment below, observers are presented with pairs of photographs of children and asked to rate the similarity s of each pair on a scale from 0 (not at all similar) to 10 (very similar). We will use the term “related” to mean that two children have the same biological parents. Half of the pairs of children are related (R) and half are not (R̅). The order of presentation is randomized, and therefore the prior probability P[R] that the children in each pair are related is 1/2. We record the frequency of use of each of the ratings, separately for related and for unrelated pairs. Their expected values are proportional to the true underlying conditional probabilities, P[s|R], the likelihood that the observer says “s” when confronted with a related pair of children, and P[s|R̅], the probability that the observer says “s” when confronted with an unrelated pair of children. If the distributions P[s|R] and P[s|R̅] were identical, then we could conclude that the observer’s similarity ratings contain no information about kinship. We can compute the posterior odds

\[
P[R|s] / P[ R̅|s] \text{ that the children are related by Bayes’ theorem in odds form (Mood, Graybill, & Boes, 1974),}
\]

\[
P[R|s] = \frac{P[s|R] \times P[R]}{P[R|s]} = \frac{P[s|R]}{P[R|s]},
\]

where \( P[R]/P[ R̅] \) is the prior odds that the children are related. For the circumstances of our experiment, the posterior odds that a pair of children rated s is equal to the ratio of likelihoods,

\[
P[R|s] = \frac{P[s|R]}{P[s|R̅]},
\]

because \( P[R] = P[R̅] \).

Optimal Bayesian classification

In Figure 2B, we schematized the visual processing of an observer whose performance in a signal detection task is based entirely on his or her similarity rating. It is probably not obvious how to convert a similarity rating into a judgment of kinship. The optimal method for doing so (Duda & Hart, 1973; Green & Swets, 1966/1974; Mood et al., 1974) turns out to be intuitively plausible, as we explain next. We describe it for signal detection tasks involving kinship. Given the similarity rating s, compute the log posterior odds,

\[
D(s) = \log \left( \frac{P[R|s]}{P[R̅|s]} \right),
\]

and compare it to a threshold \( \Delta \). If the log posterior odds is above the threshold, say “yes” in the signal detection task, otherwise say “no.” The choice of \( \Delta \) depends on prior odds and payoffs associated with different outcomes (Green & Swets, 1966/1974).

Suppose, for example, that we choose \( \Delta = 0 \). Then \( D(s) > 1 \) if and only if the posterior odds \( P[R|s]/P[R̅|s] \) are greater than 1:1. The decision rule for this choice of \( \Delta \) can be paraphrased as, “Say yes if the rating is more likely to have come from a related pair than from an unrelated pair, otherwise say no.” If \( D(s) < 0 \), then the posterior odds \( P[R|s]/P[R̅|s] \) are less than 1:1. Again, we emphasize that this decision rule is not simply plausible or intuitively appealing but also optimal in the sense that you will make the fewest possible classification errors on average (Duda & Hart, 1973).

Materials and methods

Participants

Sixty-four people, recruited in public places in Padova, Italy, were alternately assigned to one of two conditions
(‘‘similarity’’ or ‘‘kinship’’) as described below. There were 11 males and 21 females in each condition. Their ages ranged from 19 to 33 years (median age 23 years).

Photographic material

Seventy-two color photographs, each depicting a child from the neck to the top of the head, were used. The pictures had been taken by the experimenters or their assistants under controlled lighting conditions. Of the 72 children depicted in the pictures, half were girls and half were boys. We used Adobe Photoshop® to obliterate all background detail, replacing it by a uniform dark grey field (33% of maximum intensity in each of R, G, and B channels). A sample photograph is shown in Figure 3. The ages of the children ranged from 17 months to 15 years. The distribution of age differences for related and unrelated pairs was matched. The distribution of gender was also counterbalanced across related and unrelated pairs. The facial expressions were neutral or close to neutral. All came from three adjacent provinces of Northern Italy: Padova, Mantova, and Vicenza. All were Caucasian in appearance. The parents of each child gave appropriate permission for their child’s photograph to be used in scientific experiments. We asked for and received separate parental permission to use the photograph in Figure 3 as an illustration here. For privacy reasons, we did not verify by DNA fingerprinting that sibling pairs shared two parents. Recent research using DNA fingerprinting shows that the median rate of ‘‘extrapair paternity’’ is much lower than previously thought, under 2% (for a review, see Simmons, Firman, Rhodes, & Peters, 2004). In any case, the presence of half-siblings (who share a mother but not a father) would have little effect on the outcome of our experiment. Such half-sibling would have 25% of their DNA in common rather than 50%, but would still be more closely related than nonsibling pairs.

Picture pairs

Sixty of the photographs chosen at random were used in the main experiment. Twelve of the photographs were used only in the familiarization and training parts of the experiment, described below. The sixty experimental photographs comprised 30 pairs, half of which depicted children who were biological siblings. The remaining 15 pairs depicted children who were not biological siblings. We refer to the pairs in the first group as related and in the second as unrelated. Within each group of 15, five pairs depicted two boys, five pairs depicted two girls and five pairs depicted a boy and a girl. The twelve nonexperimental photographs were selected at random from the full set, subject to the constraint that they included three pairs of biological siblings.

Procedure

The experiment was conducted in a computer classroom at the University of Padova. Observers viewed all stimuli on a computer monitor and responded by marking forms provided. The experiment was self-paced and consisted of three phases. (1) Familiarization: The observer was first asked to perform a simple recognition memory task that involved all of the experimental stimuli. All 72 photographs of faces were shown in groups of six per display in random order. The purpose of this part of the experiment was to familiarize the observer with the range of faces he would see in the main experiment. The observer was asked to study the display and told that, immediately after studying each group, he would be shown a probe photograph and asked to report whether this photograph had been among the group of six just studied. The probe photographs were the nonexperimental photographs described above which were not used in the main part of the experiment. (2) Training: The observer practiced the response for his condition (similarity or kinship) on six pairs of photographs that did not overlap with the photographs used in the main part of the experiment. These pairs were drawn from the nonexperimental photographs organized so that there were three pairs that were biological siblings and three that were not. The purpose of this part of the experiment was simply to let the observer become comfortable with the procedure and response. (3) Main: The observer then performed the task appropriate
for his condition (similarity or kinship) on 30 pairs of photographs, presented in random order.

Tasks

Participants in the experiment were alternately assigned to one of two main experimental conditions (“similarity” or “kinship”). There were 32 participants in each condition and all participants were initially familiarized with the stimuli and allowed to practice the task for their condition. In each main experimental condition, the participants viewed 30 pairs of photographs of children, half of whom were genetic siblings. The pairs were presented in random order. In the similarity condition, participants rated the facial similarity of each pair. They were not told that half of the children were genetic siblings nor did we mention age differences and gender differences. Observers were free to interpret the term “similarity” as they chose. In the kinship condition, participants were told that half the pairs portrayed genetic siblings and were asked to classify each pair as siblings or nonsiblings.

Results

We first compared the mean similarity ratings across observers given to each of the 30 pairs of pictures in the similarity condition and the proportion of observers who judged the same pair to be related in the kinship condition. These measures are plotted against each other in Figure 4. Points plotted in red correspond to pairs that were related, points plotted in blue correspond to pairs that were not. The Pearson product-moment correlation is .92 (and variance accounted for is therefore .84). The regression line for the regression of mean similarity on proportion judged related is drawn in black. There are no evident outliers and the pairs of ratings are in good agreement with one another. Children judged to be more similar in the similarity condition are more likely to be judged to be siblings in the kinship condition and v.v.

We then performed an analysis in which we in effect ask, given that one is told that the mean rating for a particular pair of photos is s, what has one learned about the condition probability $P[R|s]$ that the children portrayed are siblings? We will find that the relationship is remarkably simple.

The red-filled circles connected by solid red lines in Figure 5A are a plot of the relative frequency of occurrence $P[s|R]$ of the similarity ratings 0, 1, 2, ..., 10 evoked by pairs of children who were actually siblings. The blue-filled circles connected by dashed blue lines in Figure 5A are a plot of the relative frequency of occurrence $P[s|R]$ of the ratings for pairs who were not. As the notation suggests, these relative frequencies are estimates of $P[s|R]$ and $P[s|R]$, respectively. It is clear that related pairs are typically assigned higher ratings. About 25% of the ratings assigned to unrelated pairs were 0 but some unrelated pairs received ratings of 10. Conversely, about one in six of the ratings of related pairs were 10 but some related pairs received ratings of 0.

The ratio between corresponding points of the curve on Figure 5A,

$$D(s) = \frac{P[s|R]}{P[s|R]},$$

is an estimate of the log posterior odds that children rated as having similarity s are related. We reach Equation 4 by substituting Equation 2 into Equation 3 and then replacing $P[s|R]/P[s|R]$ by its estimate. The resulting estimate is a maximum likelihood estimate of $D(s)$.

Figure 5B is a plot (black circles) of these estimates of the log posterior odds $D(s)$ as defined in Equation 3 versus similarity rating $s$. The line is the maximum likelihood regression fit to the log-odds. Its equation is,

$$D(s) = 0.489 (s - 4.79).$$

The proportion of variance accounted for by the linear regression ($R^2$) is .96. The relationship is remarkably linear and can be interpreted as follows. A rating of 4.79
Figure 5. The log-odds for related and unrelated pairs. (A) The estimated likelihood functions for similarity ratings of related pairs of children, \( \hat{P}[s|R] \), are plotted as red-filled circles connected by solid red lines, and for similarity ratings of unrelated pairs of children, \( P[s|R] \), as blue-filled circles connected by dashed blue lines. (B) The filled circles in the plot on the right mark the natural logarithms of the ratios of the estimated likelihood for related pairs to the estimated likelihood for unrelated pairs for each similarity rating, the estimated log likelihood ratio (log posterior odds) \( D(s) \). See text. The solid line is the maximum likelihood regression fit to the remaining log posterior odds. It has slope 0.489 and horizontal intercept 4.79. The proportion of variance accounted for \( (R^2) \) is .96.

(near the middle of the scale, 5) corresponds to even odds that the children are related. Each increase in rating by one step corresponds to an increase in \( D(s) \) of 0.489 (the slope of Equation 5). That increase is equivalent to a multiplication of the posterior odds by \( e^{0.489} = 1.63 \). Each decrease in rating by one step corresponds to a multiplicative increase of 0.61 = 1/1.63. A rating of 8 corresponds to odds of almost 8:1 that the children are related; a rating of 1 corresponds to odds of more than that 6:1 that they are not.

The proportion of variance accounted for in the linear regression of Equation 5 is 0.96, suggesting that similarity primarily conveys information about kinship. Indeed, any other objective attribute of pairs of children that is uncorrelated with degree of kinship (such as gender difference and age difference) could account for no more than 0.04 of the variance. Consequently, the thresholding of similarity employed in previous studies is an effective way to turn similarity ratings into judgments of degree of kinship (Brédart & French, 1999; Bressan & Dal Martello, 2002; Christenfeld & Hill, 1995). We return to this point below when we attempt to predict age difference and gender difference of pairs of children in the sample from observers’ similarity ratings and find that we cannot.

The observers as a group are using the similarity scale in a remarkably consistent manner to signal the log posterior odds that children are related. Although individual observers may deviate from this rule, these deviations are idiosyncratic in that they cancel when all observers’ data are combined. What all observers’ ratings share is an encoding of a kin recognition signal as log posterior odds in a condition of the experiment where kinship was never mentioned in the instructions to the observers. The children also differed in age and in gender (relatedness, gender difference and age difference were counterbalanced across the pairs of photographs). Yet the correlations between similarity ratings and these other factors were not significantly different from 0. Observers chose to ignore obvious differences in age or gender and base their ratings almost exclusively on perceived genetic relatedness.

We then computed signal detection measures of performance for participants in the kinship condition. These are reported as sensitivity \( d' \) and likelihood criterion \( \beta \) in Table 1. In our terminology above, \( \Delta = \log \beta \). We use \( \beta \) in Table 1 rather than \( \Delta \) as it is customary in signal detection analyses.

The value for \( \beta \) in the kin recognition condition is slightly less than 1, indicating that participants have a slight bias to judge pairs of children to be related. The \( d' \) value is significantly greater than 0 (see table). We then computed signal detection measures for the TSO with threshold 4.79 on the similarity scale. This choice of threshold is the threshold estimated by linear regression

\[
\text{Table 1. Signal detection analysis. Estimates of } d' \text{ and likelihood criterion } \beta \text{ for both the TSO and the kin recognition conditions. To compute a } d' \text{ value for the TSO, the similarity ratings were converted to effective signal detection judgments by thresholding them. The threshold used for the TSO (4.79) was chosen so that the likelihood criterion } \beta \text{ for the TSO matched the likelihood criterion estimate in the kin recognition as closely as possible. See text. The values preceded by } \pm \text{ are standard deviations estimated by a Bootstrap procedure (Efron & Tibshirani, 1993). The } z \text{ statistic is used in testing the hypothesis that the corresponding } d' \text{ measure (in the same row) is significantly different from 0. It is computed as the estimate of } d' \text{ divided by the Bootstrap estimate of its } SD. \ A bound on the corresponding } p \text{ value is reported in the last column.}
\]

\[
\begin{array}{cccccc}
\text{ } & d' & \beta & z & p \\
\text{TSO} & 1.057 \pm 0.084 & 0.936 \pm 0.044 & 12.642 < .001 \\
\text{Kin recognition} & 0.999 \pm 0.084 & 0.867 \pm 0.039 & 11.926 < .001 \\
\end{array}
\]
above and it also resulted in the closest match between $\beta$ for the kinship condition and the resulting estimate of $\beta$ for the TSO. With this choice of threshold, the $d'$ values for the kinship condition and the TSO are very close and that for the TSO is slightly higher than that for the kinship condition, but not significantly so ($z = 0.469; p > .05$). The TSO, given the participants’ ratings in the similarity condition, is as effective in discriminating genetic relatedness as the participants who directly judged the degree of kinship of the children.

What if we attempted to predict the gender difference of pairs of children from the similarity ratings, using a computation analogous to the TSO? The signal is no longer genetic relatedness but whether the children are of the same gender. We estimate $d'$ and $\beta$ for this new “TSO (gender).” We arbitrarily designated pairs of the same gender as signal. The results of this analysis are shown in Table 2: the estimate of $d'$ for the TSO (Gender) is small and is not significantly different from 0. We conclude that we cannot use an analogue of the TSO to reliably predict whether children differ in gender given rated similarity.

We also examined whether age difference can be predicted from similarity ratings. We labeled each pair of children as signal if their age difference was less than the median age difference across the sample. Otherwise, the pair counted as “not signal.” The results of this analysis are also shown in Table 2: the estimate of $d'$ is also small and is not significantly different from 0. Although similarity ratings can be used to predict genetic relatedness, we cannot reject the hypotheses that they carry no information about gender and age differences.

This outcome is consistent with the analysis above where we found that similarity ratings accounted for .96 of the variance in the estimated log likelihood ratios, leaving little variance to be accounted for by any other measure.

Evidently our results depend upon the observer’s interpretation of the instructions to judge similarity (cf. Medin, Goldstone, & Gentner, 1993). We could, for example, ask observers to judge “similarity in age” and expect to find that similarity now conveyed considerable information about age difference. What is therefore of interest is that observers judging “unconstrained similarity” of children’s faces chose to judge, in effect, degree of kinship and not age difference or gender difference. This outcome cannot be a property of all similarity judgments but only of similarity judgments of objects that are, or can be interpreted as, biological organisms. It remains to be seen whether this same bias is specific to children’s faces or whether it is present in judgments of the similarity of adults’ faces, or in adult–child pairs of faces or in judgments of the similarity of members of other species.

**Discussion**

In this article, we examined the relationship between ratings of facial similarity of pairs of children and observers’ ability to discriminate pairs of children that were biological siblings from those that were not. The results and analyses just described support the claim that the perceived similarity of pairs of children is little more than a summary of the (log) likelihood ratio that they are close biological relatives. Adults, in judging the facial similarity of children, are in effect estimating their degree of kinship expressed as the log-odds that they are near genetic relatives.

This article reports the first demonstration of collateral kin recognition (here, recognition of siblings) in human observers. As we noted in the introduction, researchers often use facial similarity judgments as a measure of kin recognition (Brédart & French, 1999; Bressan & Dal Martello, 2002; Christenfeld & Hill, 1995). Our results support this practice when observers are judging pairs of children (for an analogous experiment concerning adults and infants, see Bressan & Grassi, 2004).

We began by noting that similarity leads a double life, as something we experience, and as an explanatory construct in a wide variety of theories. Our results are consistent with the claim that kin recognition signals are available that could guide human social interactions (Hamilton, 1964a, 1964b). Hamilton’s inclusive fitness theory postulates that the organisms are capable of kin recognition via phenotype matching. That is, they should be able to make “implicit evaluation of relatedness on the basis of some trait-based assessment of phenotypic similarity” (DeBruine, 2002). Our results indicate that they do. Our participants in fact judge relatedness when asked only to assess similarity of children’s faces.

The results also support the hypothesis that our representation of the world around us is naturally framed in the language of mathematical statistics (Gold & Shadlen, 2001; Rao, Lewicki, & Olshausen, 2002). Gold and Shadlen (2001), in particular, have recently proposed that log likelihood ratio (log-odds) rather than probability is the neural representation of uncertainty (an idea that can be traced to Peirce & Jastrow, 1885). It is intriguing to consider that, in the similarity condition, observers may be
translating their estimates of the log-odds that each pair of children is related into a similarity rating by a simple linear transformation of the output of a neural mechanism that encodes estimated log-odd of genetic relatedness (Equation 4).

Our results also are consistent with the claim that similarity is not an arbitrary construct but rather reflects aspects of the physical, biological and social worlds (Shepard, 1987) that are important to the survival of the organism. The weighted mixture of features that constitute “unconstrained similarity” gives little weight to features that signal age or gender and great weight to features that signal genetic relatedness. The resulting “odd” mixture of features is easily explained once we realize that it generates as good a kin recognition measure (estimate) as the observer can manufacture when given explicit instructions to judge kinship and assess genetic relatedness in our social world.

We end by raising two questions. The first is what features of children’s faces does the observer use in judging either degree of kinship or of similarity? Our results suggest that both judgments rely on parts of the face that are largely genetically determined and that are expressed early in development but these results do not allow us to identify the features themselves. In Dal Martello and Maloney (2006), we begin to address this question by masking regions of the face to determine where the features signaling kinship lie. We plan analogous studies for judgments of similarity and to test the conjecture that the same facial features are used with the same weighting in the two kinds of judgment.

The second question is would the observer continue to use the same features with the same weighting in judging kinship, age, gender, or similarity between adults or between an adult and a child? If a cue becomes difficult to estimate or unreliable over the course of normal growth, an ideal observer would give it less weight, in effect changing over to other cues (Landy, Maloney, Johnston, & Young, 1995). It is perfectly plausible that useful age and gender cues change over the life span. For example, the ratio of upper face height to lower is a good candidate cue to the age of a child (compare the four faces in Figure 1), but this cue would be of little value in judging age differences beyond adolescence when development of the bony structure of the face is completed (Kohn, 1991).

The bony structure of the upper and lower face are genetically determined to roughly the same degree (Kohn, 1991) but the lower face does not reach final form until early adulthood. In Dal Martello and Maloney (2006), we find that observers rely primarily on features in the upper face in judging kinship of children. Would we find that observers make greater use of features in the lower face in judging kinship between adults, now that these (fully expressed) features are informative? We can frame this question as the conjecture that the rules employed by the visual system in making judgments of age, gender, and kinship/similarity from the appearance of faces are dynamic and predictable: For any judgment, it makes use of the best available relevant information at any point in normal facial development.

Acknowledgments

LTM was supported by NIH grant EY08266, MFDM was supported by funds from the Italian Ministero dell’Università e della Ricerca Scientifica e Tecnologica. We thank Davide Bianconi, Francesca Constantini, and Carla Scagliarini for assistance in collection of data and taking of photographs. We thank Marisa Carrasco, Michael Landy, Gary Marcus, and Richard Power for comments on earlier drafts.

Commercial relationships: none.
Corresponding author: Laurence T. Maloney.
E-mail: Laurence.Maloney@nyu.edu.
Address: Department of Psychology, Center for Neural Science, New York University, 6 Washington Place, New York 10003, USA.

References


