COGNITION

Cognition 68 (1998) 199–220

Mapping attractor fields in face space: the atypicality bias in face recognition

James Tanaka[a,]*, Michael Giles[a], Sarah Kremen[b], Valerie Simon[c]

[a]Severance Laboratory, Department of Psychology, Oberlin College, Oberlin, OH 44074-1086, USA
[b]Johns Hopkins University, Baltimore, MD, USA
[c]Duke University, Durham, NC, USA

Received 9 June 1998; accepted 20 July 1998

Abstract

A familiar face can be recognized across many changes in the stimulus input. In this research, the many-to-one mapping of face stimuli to a single face memory is referred to as a face memory's 'attractor field'. According to the attractor field approach, a face memory will be activated by any stimuli falling within the boundaries of its attractor field. It was predicted that by virtue of its location in a multi-dimensional face space, the attractor field of an atypical face will be larger than the attractor field of a typical face. To test this prediction, subjects make likeness judgments to morphed faces that contained a 50/50 contribution from an atypical and a typical parent face. The main result of four experiments was that the morph face was judged to bear a stronger resemblance to the atypical face parent than the typical face parent. The computational basis of the atypicality bias was demonstrated in a neural network simulation where morph inputs of atypical and typical representations elicited stronger activation of atypical output units than of typical output units. Together, the behavioral and simulation evidence supports the view that the attractor fields of atypical faces span over a broader region of face space that the attractor fields of typical faces. © 1998 Elsevier Science B.V. All rights reserved

Keywords: Face recognition; Atypicality bias; Attractor field; Memory

1. Introduction

In face recognition, a face is recognized when the stimulus activates a stored

* Corresponding author.

0010-0277/98/$19.00
PII S0010-0277(98)00048-1

representation in memory. While recognition is optimal when there is a good fit between the face stimulus and its memorial representation, the recognition system is sufficiently robust to identify a familiar face across deviations in the face input. For example, the behavioral evidence has shown that faces are recognizable despite changes due to age (Bruck et al., 1991), orientation (Hill et al., 1998), expression (Ekman et al., 1972), and caricaturization (Rhodes et al., 1987). These findings suggest that a single face representation can attract and be activated by multiple face inputs. The many-to-one mapping of stimulus inputs to a face memory is referred to as a representation's attractor field. In the described experiments and computer simulation, we explore the factors that might determine the boundaries of a face representation's attractor field. Our working hypothesis is that the attractor fields of atypical faces are determined by the density of nearby representations in similarity space. The density hypothesis predicts that by virtue of their locations in a sparse region of face space, the attractor fields of atypical faces will be relatively large compared to attractor fields of typical faces. To test this hypothesis, subjects were shown face images that were morphed together from equal contributions of atypical and typical faces. The main finding of these experiments was that consistent with the density hypothesis, subjects judged the morph face as bearing a stronger resemblance to the atypical face than the typical face.

2. Attractors in face space

The set of all possible faces that exist in the world can be represented as an *n*-dimensional Euclidean space or 'face space'. It is assumed that the dimensions of face space, while not explicitly specified, are continuous and correspond to a physical property of a face (e.g., length of nose, inter-eye distance) (Valentine, 1991). At the origin of face space lies the most typical or average face in the population, referred to as the norm face. Face representations in face space are normally distributed around the norm face such that the density of face representations decreases proportionately as distance from the origin increases. In a face space model, face typicality can be shown as a vector distance between the norm face and the target face such that, as the vector distance increases, typicality decreases. Given the normal distribution of representations in face space, this model predicts that typicality and distinctiveness should be inversely correlated such that, as typicality decreases, distinctiveness increases. Typical faces are less distinctive than atypical faces because they lie in a region of higher face density where there is more competition from nearby representations. Atypical faces, on the other hand, are more distinctive because they are located in a sparser region of face space where there is less competition. Consistent with the face space approach, the behavioral evidence has shown that faces rated as atypical in appearance are learned more easily and are identified more quickly and accurately than typical faces (Light et al., 1979; Bartlett et al., 1984; Valentine and Bruce, 1986).

In an attractor field model, memory representations, such as familiar faces, form stable points in the system that carve valleys or basins on the surface of the similarity space (Tank and Hopfield, 1987). In recognition, analogous to a raindrop moving downhill to minimize its gravitational energy, the to-be-recognized face stimulus decreases its computational energy by following the path that leads to the nearest attractor basin in representational space. Following this approach, the stimulus need not be a perfect fit with its underlying representation; only a close enough approximation to fall within the boundaries of the attractor field's basin. The potential activation of any given representation will therefore be directly proportional to the size or span of its attractor field (e.g., representations with large attractor fields will capture a broader range of stimulus inputs).

In the current study, the attractor field boundaries of atypical and typical faces are tested by presenting face stimuli that lie equidistant from the two competing representations in face space. Equidistant stimuli are generated by averaging or morphing pairs of atypical and typical faces. As shown in Fig. 1, the morph face can be represented as a vector that connects two face representations in face space. Points

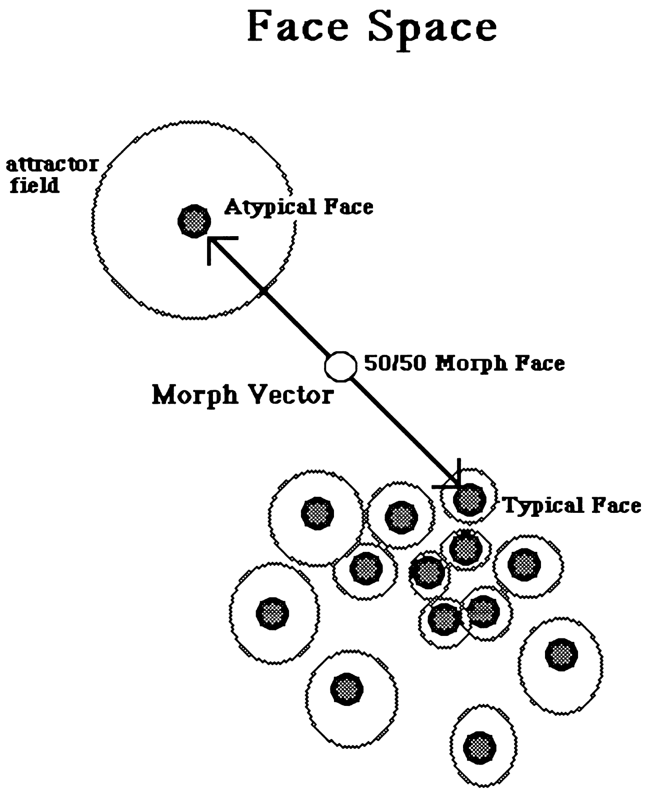

Fig. 1. Diagram of theoretical face space. Face representations are indicated by the small filled circles and their attractor fields by the larger circles. The morph vector connects an atypical face and typical face representation. The 50/50 morph face containing 50% contribution from the atypical face parent and 50% contribution from the typical face parent is indicated by the unfilled circle on the morph vector.

along the morph vector correspond to the continuum of faces that can be derived by differentially weighting the contributions of two parent faces to the morph face. Hence, the point that is equal distant from two representations along the morph vector represents the 50/50 morph face.

Should the 50/50 morph face more strongly resemble the atypical or the typical parent face? A strict Euclidean model holds that similarity judgments are computed on the basis of interpoint distance in the metric space (Shepard, 1962). According to this view, representations that are closer neighbors in Euclidean space should be psychologically more similar than representations that are farther apart. Given the 50/50 morph face is equidistant from the atypical and typical representations, the Euclidean-based, nearest-neighbor hypothesis predicts that the morph face should bear equal resemblance to the atypical face and the typical face.

Alternatively, Krumhansl (1978) has suggested that similarity is jointly determined by interpoint distance and the spatial density of representations in a surrounding area of similarity space. Atypical faces reside in a sparser region of face space where there is less competition from neighboring representations. As depicted in Fig. 1, less competition results in broader attractor fields for the atypical face representations. In contrast, typical representations are located in a spatially dense area of face space where their attractor fields are more constrained by neighboring representations. Although a 50/50 morph is equidistant from the atypical and typical face in similarity space, the density hypothesis predicts that the morph face should more strongly activate the atypical parent than the typical face given its larger attractor field.

3. Experiment 1: the effects of typicality on likeness judgments

In Expt. 1, the predictions of the nearest-neighbor and density hypotheses were examined by asking subjects to make likeness judgments of morph faces derived from atypical and typical parents. In previous face research, morphing has been used as an experimental technique to investigate the perception of facial expression (Etcoff and Magee, 1992; Calder et al., 1996) and identity (Beale and Keil, 1995). To examine the perception of facial expression, line drawings (Etcoff and Magee, 1992) and grey-scale images (Calder et al., 1996) of individuals displaying two of the basic facial expressions (e.g. happy, sad, fear, anger and surprise) were averaged together along a linear continuum. Although the images changed in equal increments, subjects perceived the expressions as changing abruptly from one expression to another at a certain point along the continuum. Similar results have been reported by Beale and Keil (1995) where faces of well-known celebrities (e.g. Bill Clinton, John Kennedy) were averaged. Beale and Keil (1995) found that morph faces that varied along a continuum were perceived and recognized discontinuously by subjects and argued that recognition of facial identity, similar to the recognition of facial expression, was categorical.

As shown in Fig. 2, typical and atypical faces were averaged to produce a

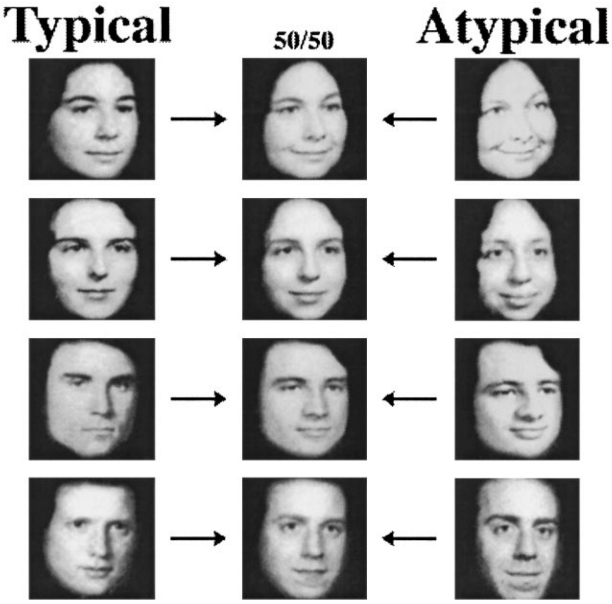

Fig. 2. Examples of the atypical and typical female and male faces and their 50/50 morph faces.

morphed image. Following previous studies (Etcoff and Magee, 1992; Beale and Keil, 1995), we employed a delayed-match-to-sample task in which an atypical and a typical face were presented for 2.5 s, followed by their morphed image. The subjects were asked to decide whether the morph face more strongly resembled the atypical face or the typical face. Judging the likeness of a morph face provides a good test between the competing claims of the nearest-neighbor and the density hypotheses. Given that the image is the same distance from both representations in Euclidean space, the nearest-neighbor hypothesis predicts that the morph image should produce equivalent activation of the typical and the atypical representations. Hence, subjects should be just as likely to judge the morph image as resembling the typical face as the atypical face. According to the density hypothesis, the size of an attractor field varies as a function of the representation's position in face space. If the attractor fields of atypical faces extend over a larger region of face space than typical faces, the morph face, despite being the geometric average of the two faces, would fall within the attractor field of the atypical representation. The density hypothesis thus predicts that subjects should prefer the atypical parent over the typical parent.

3.1. Method

3.1.1. Subjects

Twenty-seven undergraduates from the introductory psychology class at Oberlin College participated in this experiment for course credit. All subjects had normal or corrected normal vision.

3.1.2. Materials

The face stimuli were selected from several college yearbooks. The selected faces contained no facial hair or glasses, little hair covering the forehead, and a closed mouth. Faces were displayed in frontal or near-frontal viewing pose. Seventy-six male faces and sixty-six female faces meeting the above criteria were Xeroxed, cropped around the face outline, and glued onto eight $1/2 \times 11$-inch sheets of paper (approx. 12 faces/page). Packets were assembled from the face sheets, with the male and female pages mixed together in a random order. Forty subjects rated the distinctiveness of the faces sheet on a scale from 1 to 5, with 1 being non-distinctive and 5 being very distinctive. Subjects were instructed to rate the faces according to how distinctive looking they appeared relative to the general population.

From the original set of faces, 16 faces, eight female and eight male, were selected on the basis of being either high or low in rated typicality. The four atypical female faces had a mean distinctiveness rating of 3.95 and the four typical faces had a mean distinctiveness rating of 2.10. The four atypical male faces had a mean distinctiveness rating of 4.00 and the four typical faces had a mean distinctiveness rating of 1.88. The 16 faces were digitized using a MicroTech Z flatbed scanner and Adobe Photoshop™. Using the Morph™ 2.5 program, a 50/50 morph face was generated by averaging each same gender, atypical face with each same gender, typical face. This process yielded a total 16 female morphs and 16 male morphs. Key points for facial features were kept constant, with 12 points on the mouth, 7 points on each eye, 6 points on the nose, 5 points on each eyebrow, and 22 points for the outline of the face. Occasionally, a few extra points were required to eliminate shadows on the forehead region.

The morphing procedure was equivalent to the process described by Beale and Keil (1995). In the morphing process, a Delaunay tessellation process was applied in which neighboring control points were connected to form non-crossing triangular regions on a planar surface. The triangular regions were optimized such that pixels within a given region were closer to the control points at the triangle's vertices than to any other control points on the surface. Applying a warping algorithm, control points for the morph face were generated by moving 50% of the total distance along the vector that connected corresponding control points in parent face 1 (x_1, y_1) and parent face 2 (x_2, y_2). The locations of intervening pixels were linearly interpolated across the surface based on the position of the nearest control point (see Wolberg, 1990). A fade process was then employed in which the brightness values for each corresponding pixel were weighted according to the contribution of each parent image. In a 50/50 morph face, pixel values were determined based on the equal contribution of each parent image. Faces subtended a visual angle of 11° and 7° in the vertical and horizontal dimension, respectively. Face stimuli were presented on a computer monitor with a resolution of 72 dots per inch.

3.1.3. Procedure

Each trial consisted of the subject first seeing the atypical and typical face probes presented side-by-side on the screen for 2.5 s The face probes were then replaced by the morph face for 1 s and then replaced by the words ‘Left or Right?’. The subject was asked to decide whether the morph face more strongly resembled the face probe on the left or the face probe on the right. They indicated their decision by pressing the appropriately marked key. After their response, the next pair of probe faces was presented after a 2-s delay. The left and right positions of the typical and atypical faces were counterbalanced across trials. Each of the 32 face pairs and morph faces were presented randomly twice for a total of 64 experimental trials. Subjects were tested on Macintosh computers using the SuperLab 1.5 experimental package.

3.2. Results and discussion

Fig. 3 shows the distribution of subjects according to the overall percentage of atypical and typical responses. As shown in the figure, subjects demonstrated a strong bias toward selecting the atypical face over the typical face. Overall, the atypical face was selected on 60% of the trials. The obtained atypical face responses were compared against the 50% level predicated by the nearest-neighbor hypothesis. The atypical face preference was significantly above the level of 50% as tested by subject, $t(26) = 6.07$, $P < 0.01$, and by item, $t(31) = 3.70$, $P < 0.01$. Separate analysis of the female and male faces revealed that both kinds of faces demonstrated the atypical bias. Subjects preferred the atypical female face on 64.6% of the trials, $t(26) = 6.67$, $P < 0.01$, and the atypical male face on 56%, $t(26) = 2.50$, $P < 0.02$.

In Expt. 1, subjects demonstrated a strong preference for the atypical parent over

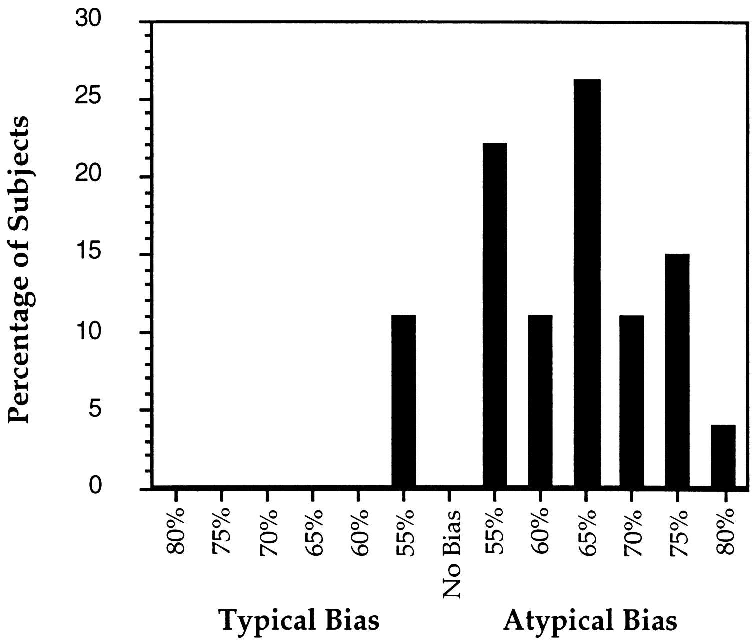

Fig. 3. Percentage of subjects showing an atypical and typical bias as indicated by their overall percentage of responses.

the typical parent. This finding indicates that contrary to the nearest-neighbor hypothesis, the representational boundary between an atypical and typical face does not correspond to the Euclidean midpoint that separates the two representations. Consistent with the density hypothesis, these results indicate that as indexed by subjects' likeness judgments, the attractor fields of atypical faces are larger than the attractor fields of typical faces.

4. Experiment 2: related and unrelated face pairs

In the match-to-sample paradigm where two probe faces are simultaneously presented and subsequently replaced by a morph face, subjects might select the atypical face not because the test face more strongly resembles the atypical face, but because the previously presented atypical face is more memorable than the typical face. In other words, the atypical bias might reflect differences in memory strength of atypical and typical faces rather than differences in the size of their attractor fields. In Expt. 2, this alternative explanation is examined by presenting an atypical and a typical face, followed by either a related morph face or an unrelated morph face. In the related morph face condition, the morph face was derived from the atypical and the typical parent shown in that trial. In the unrelated morph face condition, the morph was generated from a different pair of typical and atypical parents. If subjects continue to select the atypical face in the unrelated condition, this result would indicate that their decision is based on an encoding advantage for atypical faces. Alternatively, if subjects are equally likely to select the typical face as the atypical face in this condition then this would suggest that obtained atypicality bias reflects representational differences.

4.1. Method

4.1.1. Subjects

Twenty-three Oberlin undergraduates from the introductory psychology class participated in this experiment for course credit. All subjects had normal or corrected normal vision.

4.1.2. Stimuli

The stimuli used in this study were the sixteen faces and their respective morphs from Expt. 1.

4.1.3. Procedure

Following the match-to-sample procedure used in Expt. 1, subjects viewed an atypical and a typical face presented simultaneously for 2.5 s. The face pair was replaced by either a related morph face or an unrelated morph face. In the related condition, the morph face was derived from the presented atypical and typical parent faces. In the unrelated condition, the morph face was not derived from the presented pair of atypical and typical parent faces. In the unrelated condition, each morph face item was assigned to an unrelated atypical and typical face pairing of the same gender as the morph face (e.g., female morph faces were matched with other unrelated female atypical and typical parents). Subjects were instructed to decide whether the morph face more strongly resembled the face previously presented on the left or right of the screen and to indicate their decision via a key press response. There were a total of 32 related morph face items (16 males and 16 females) and 32 unrelated morph face items (16 males and 16 females), and each item was presented twice yielding a total of 128 trials. Trials were randomly presented.

4.2. Results and discussion

Replicating the findings of Expt. 1, likeness judgments in the related condition more strongly favored the atypical face over the typical face. Overall, the atypical face was preferred on 63% of the trials, a preference that was reliably above the 50% level as tested by subject, $t(22) = 9.52$, $P < 0.01$, and by item, $t(31) = 5.49$, $P < 0.01$. Separate analysis of the female and male faces again showed that both sexes demonstrated the atypical bias. Subjects selected the atypical female face on 67% of trials, $t(22) = 10.50$, $P < 0.01$, and the atypical male face on 58%, $t(22) = 4.461$, $P < 0.01$.

In the unrelated condition, the opposite pattern was found where subjects demonstrated a typicality rather than an atypicality bias. That is, on those trials where the morph face was not derived from the preceding ‘parent’ faces, subjects preferred the typical face on 60% of the trials, a preference that was significantly above the 50% chance level, tested by subject, $t(22) = 5.30$, $P < 0.01$, and by item, $t(31) = 2.89$, $P < 0.01$. Presence of the typicality bias seemed to vary according to the gender of the face stimuli. When viewing female faces in the unrelated condition, subjects demonstrated no preference for either the typical or the atypical female face; both faces were selected on 50% of the trials, $P > 0.10$. In contrast, when presented with male faces, subjects preferred the typical male face on 70% of the trials, $t(22) = 10.32$, $P < 0.001$. The difference in typicality judgments between the male and female faces was significant, $t(22) = 6.65$, $P < 0.01$.

To summarize, results from Expt. 2 demonstrated that a 50/50 morph face was perceived to more strongly resemble the atypical parent than the typical parent. Thus, Expt. 2 supports the predictions of the density hypothesis in which the resemblance judgments are claimed to be determined by the relative positions of representations in face space. These results are not compatible with the nearest-neighbor claims that a 50/50 morph face should lie at the attractor boundary between the typical and atypical face and therefore bear equal resemblance to both parent faces.

Interestingly, in the unrelated condition where the morph face was randomly paired with an atypical and typical parent, subjects showed a preference for the typical parent rather than the atypical parent. The bias toward the typical face in the unrelated condition makes sense within the framework of a multi-dimensional face space. An atypical face can be extreme on one or many separate dimensions pushing it further out in face space. It is unlikely that the dimensions of atypicality found in the morph face will be the same atypical dimensions found in unrelated atypical parent. However, typical faces are similar in that they are located near the origin of face space. By virtue of its location in face space then, a typical face (or morph face containing a mixture of atypical and typical contributions) will be more likely to resemble an arbitrarily selected typical face than an atypical face. Simply put, while there are many ways in which a face can be distinctive, there is only one way in which it can be ordinary.

Alternatively, it is possible that memorability effects might have also influenced subjects' responses in the unrelated condition. Given that the morph face did not match either of the probe faces, subjects might be more likely to reject the more strongly encoded atypical face in favor of the more weakly encoded typical face. Therefore, memorability effects might explain why preference for the atypical face in the related condition switched to the typical face in the unrelated condition. To rule out memory accounts of the atypicality bias, a perceptual version of the matching task was employed in the next experiment.

5. Experiment 3: perceptual basis of atypicality bias

In Expts. 1 and 2, subjects judged the morph face as bearing a stronger likeness to the atypical parent than the typical parent. The likeness task employed in these experiments tapped short term memory processes in that subjects were required to remember the previously shown morph face in order to compare it to the subsequently presented parent faces. In Expt. 3, likeness judgments were obtained while the morph and parent faces were simultaneously present. There were two primary reasons for testing the atypicality bias in the simultaneous presentation paradigm. First, simultaneous presentation of morph and parent faces removes all memory demands of the task and, therefore, excludes accounts of the bias effect related to differential memorability of atypical and typical faces. Secondly, this paradigm directly examines the perceptual basis of the atypicality bias. An atypicality bias finding in this paradigm would indicate that the bias occurs at the initial perceptual encoding stage. On the other hand, an absence of an atypicality bias in this paradigm would suggest that the bias occurs during storage and retrieval of face representations.

In Expt. 3, the atypicality bias was also tested for inverted faces. Previous research has shown that inversion disproportionately impairs recognition of faces more than other objects (i.e. face inversion effect) (Yin, 1969). It has been speculated that a face is perceived as atypical based on a single, distinctive feature (e.g. long nose, bushy eyebrows). Because a single local feature is less susceptible to inversion effects than a configuration of features (Rhodes et al., 1993; Leder and Bruce, 1998), it is plausible that atypical faces are less prone to inversion effects than typical faces that are characterized by their configural properties. Thus, if the encoding of distinctive features is not affected by inversion, then the inverted atypical parent should still be perceived as bearing a stronger resemblance to the morph face than the inverted typical parent. Alternatively, the magnitude of the atypicality bias would be expected to be diminished or abolished if the atypicality bias relies on the processing of configural face information.

5.1. Method

5.1.1. Subjects

Thirty-two Oberlin College undergraduates from the introductory psychology class participated in this experiment for course credit. All subjects had normal or corrected normal vision.

5.1.2. Stimuli

The stimuli used in this study were the photographs of the eight male and eight female faces used in Expts. 2 and 3 and their respective morphs.

5.1.3. Procedure

A similar match-to-sample paradigm used in Expts. 1 and 2 was employed in the present experiment with the exception that the morph face and its atypical and typical parent faces were presented simultaneously. The morph face was presented in the middle of the computer screen and was flanked by the atypical and typical parent faces. The left and right positions of the atypical and typical parent faces were counterbalanced across trials. Subjects were instructed to decide whether the morph face more strongly resembled the face on the left or the right of the screen and to indicate their decision via a key press response. The stimuli remained on the screen until the subject responded. The test items were blocked according to orientation and each item was tested twice in its upright and its inverted orientations. Half of the subjects were tested with the upright faces first and then saw the inverted faces; the other half of the subjects were tested in the reverse orientation order. There were a total of 32 upright morph face items (16 males and 16 females) and 32 inverted morph face items (16 males and 16 females), and each item was presented twice yielding a total of 128 trials. Trials were randomly presented within orientation blocks.

5.2. Results and discussion

When presented in the upright orientation, the atypical face was selected on 62% of the trials. The obtained atypical face responses were compared against the 50% level predicated by the nearest-neighbor hypothesis. The atypical face preference was significantly above the 50% level as tested by subject, $t(31) = 5.91$, $P < 0.001$, and by item, $t(31) = 3.70$, $P < 0.01$. When shown in their inverted orientation, the morph face was judged to be more similar to the atypical face than the typical face on 55% of the trials. The preference for the atypical face was significant by subject $t(31) = 3.51$, $P < 0.01$, but not by item, $t(31) = 1.70$, $P = 0.10$. A comparison of the atypical responses to upright versus inverted faces revealed that the magnitude of the atypicality bias was significantly greater for upright faces than inverted faces as assessed by subject, $t(31) = -2.91$, $P < 0.01$, and by item, $t(31) = 2.47$, $P < 0.02$.

While Expts. 1 and 2 established the atypicality bias effect in an immediate memory paradigm, Expt. 3 demonstrated an atypical bias in a strictly perceptual task. In this experiment, a reliable atypicality bias was found under conditions where subjects were allowed to view both the morph and parent faces at the same time and without time restriction. This result indicates that atypical bias is not attributable to the relative memorability of atypical and typical faces, but suggests that atypical face representations attract a broader range of face inputs than typical face representations. Hence, these results indicate that atypicality bias occurs during the initial encoding stage of face processing rather than at the later memory stage.

Results from Expt. 3 also identify two potential sources of the atypicality bias effect. Because an atypical face is frequently specified by a single distinctive feature and featural information is less disrupted by inversion than relational information (Rhodes et al., 1993; Leder and Bruce, 1998), it was predicted that an atypical bias effect should be observed when faces are presented in their inverted orientations. Consistent with this view, it was found that atypical faces produced a bias effect whether displayed in their upright or inverted orientations. However, the present results also indicate that the encoding of featural information does not provide a full account of the atypicality bias. Although inverted faces demonstrated a reliable atypicality bias effect, upright faces produced a significantly larger atypical bias than inverted faces. The difference in the atypicality bias between upright and inverted faces suggests that perceived atypicality is also determined by the configural information in a face. Distinctive configural information (e.g. especially wide or narrow set eyes) is salient when the face is viewed in its upright orientation, but may be disrupted when a face is turned upside-down (Tanaka and Sengco, 1997) thereby producing a diminished atypical bias effect. One interpretation of these results is that face atypicality can be defined by a distinctive isolated feature or by the distinctive configural context of its features.

6. Experiment 4: probing attractor boundaries in face space

The attractor field boundary between the atypical and the typical face is assumed to correspond to the point at which activation of the competing representations is equally probable. Results from Expts. 1 and 2 indicated that the Euclidean midpoint between atypical and typical representation did not reflect the attractor boundary because there was a bias to judge the 50/50 face as more similar to the atypical face. These findings raise the question of where along the atypical-typical continuum does the attractor field boundary lie. In a fourth experiment, we tested for the attractor boundary between the typical and atypical parent faces. As shown in Fig. 4, a linear continuum of faces can be produced by parametrically varying the contributions of the atypical and typical parent faces. Following the same match-to-sample procedure described in Expt. 1, subjects were simultaneously presented with the atypical and typical faces, followed by a morphed image that contained anywhere from 35% to 65% contribution from the atypical (typical) parent. Like the previous experi-

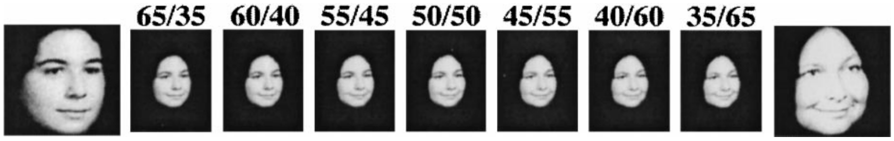

Fig. 4. Example of the continuum of morph faces produced by image averaging a typical and atypical face pair. Morph faces in the continuum were produced in 5% intervals ranging from 35% contribution from the atypical (typical) parent to 65%.

ments, the subjects' task was to decide whether the morphed image more strongly resembled the typical or atypical face. The attractor boundary was defined as the point at which there was an equal preference for atypical and typical parent in the morph face.

6.1. Method

6.1.1. Subjects

Thirty-two Oberlin undergraduates from the introductory psychology class participated in this experiment for course credit. All subjects had normal or corrected normal vision.

6.1.2. Stimuli

The four typical and four atypical females from the previous study served as parent images. New morphs were created using the same key points, with varying degrees of contribution from the typical and atypical faces. As shown above, morph faces were generated at 5% intervals beginning with a 65/35 typical-to-atypical face ratio of contribution and ending with a 35/65 typical-to-atypical ratio. This procedure yielded a total of seven morph faces for each of the 16 typical–atypical face pairings.

6.1.3. Procedure

The same match-to-sample procedure described in Expt. 1 was used in the current experiment. The seven morph faces produced from the 16 typical–atypical face pairings yielded a total of 112 experimental trials. Trials were randomly presented.

6.2. Results and discussion

As shown in Fig. 5, there was a positive linear relationship between the amount of atypical face contribution and the percentage of atypical face responses, $r^2 = 0.975$, $P < 0.001$. Consistent with the previous subjects demonstrated a bias toward the atypical face. In the 50/50 condition, subjects selected the atypical face on 56% of the trials which was reliably different from the 50% chance level performance, $t(31) = 2.09$, $P < 0.05$. The atypical bias was further demonstrated by directly comparing atypical and typical responses in conditions where the percentage contributions from the atypical and typical faces were equivalent. In the 65/35 condi-

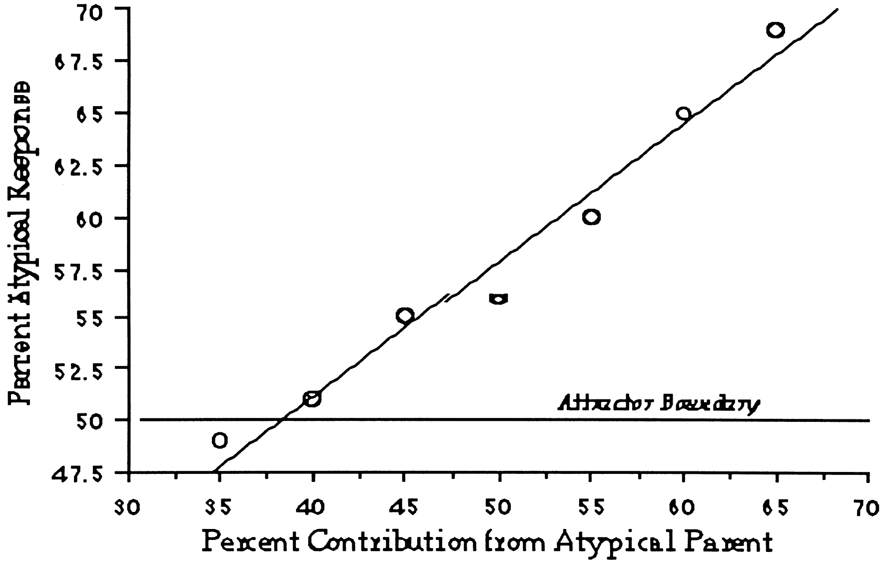

Fig. 5. Graph showing the relation between percent contribution from the atypical parent to the morph face and the percent of atypical responses. The attractor boundary (i.e. the point at which the atypical and typical faces equally preferred by subjects) was interpolated as the 37/63 (% atypical contribution/% typical contribution) morph face.

tion, subjects selected the atypical face on 69% of the trials in comparison to the 35/65 condition where subjects selected the typical face on 53% of the trials. In the 60/40 condition, subjects preferred the atypical face on 65% of the trials in contrast to the 40/60 condition where the typical face was selected on 49% of the trials. In the 55/45 condition, subjects favored the atypical face on 60% of the trials whereas in the 45/55 condition, the typical face on 45% of the trials. In each comparison where the level of contribution from the atypical and typical parent was equated, subjects reliably preferred the atypical face over the typical face, $P < 0.001$.

Based on the derived regression equation, the attractor boundary between the atypical and typical face was interpolated as the 37/63 morph face. That is, according to these data, a morph face containing 37% contribution from the atypical parent and 63% contribution from the typical parent would be perceived to bear equal resemblance to the atypical and the typical parent. Thus, the boundary separating competing attractor fields in recognition can be manipulated depending on the typicality of the representations vying for recognition.

Interestingly, contrary to the findings of Beale and Keil (1995), subjects in Expt. 4 did not perceive the images discontinuously. Indeed, there was a strong linear relationship between the weighted contribution of the parent images and their perceived likeness. This finding raises some question concerning the conditions under which faces are recognized categorically, as in the Beale and Keil study, and continuously, as in the current experiment. We return to the issue of categorical and continuous recognition in the General Discussion.

7. Computer simulation: testing the nearest-neighbor and density hypotheses in a neural network

In the following simulation, a computational test of the nearest-neighbor and density hypotheses was performed using a neural network model. Previous research has shown that neural networks are well suited for exploring the 'microstructure' of face recognition processes. For example, neural networks have been used to simulate caricature recognition (Lewis and Johnston, 1998; Tanaka and Simon, 1996) and atypicality effects in face recognition processes (Burton et al., 1990; Valentine and Ferrara, 1991). In the current simulation, the structural properties of a face, such as its outline, internal features, and spatial configuration are represented as a face vector which in turn, is used as input to the neural network model. According to connectionist implementations of face space (Burton et al., 1990; Valentine and Ferrara, 1991; Tanaka and Simon, 1996; Lewis and Johnston, 1998), encoded faces occupy unique points in multi-dimensional space such that structurally similar faces are located closer together in face space whereas structurally dissimilar faces are more spatially separated.

One advantage of the modeling approach is that face typicality can be explicitly represented by the structure of the input vectors presented to the model. In this simulation, three typical vectors were generated that differed from each other by two feature units whereas the one atypical vector differed from the typical vectors by six feature units. According to the structure of the input, the typical vectors were located nearer the origin of face space whereas the atypical vector was located further away from the origin. After learning, morph vectors were created by averaging the values of atypical and typical feature vectors. Thus, the morph vectors represented the Euclidean midpoints between the atypical and typical vectors in similarity space.

The network's response to the morph vectors provided the critical test between the nearest-neighbor and density hypotheses. If the network organized representations in a strict Euclidean space as claimed by the nearest-neighbor hypothesis, the morph vector should equally activate the typical and atypical representations. However, according to the density hypothesis, if the network was sensitive to the density of representations in similarity space, the morph should more strongly activate the atypical face than the typical face. The competing predictions of the nearest-neighbor and density hypotheses were examined in the following simulation.

7.1. Description of the neural network model

The neural network was based upon the back propagation program of the McClelland and Rumelhart (1986) software package. The network consisted of three layers: a ten-unit input layer, a five-unit middle layer, and a four-unit output layer. The layers were interconnected such that all input units were connected to all hidden units and all hidden units were connected to all face outputs. Face representations were comprised of ten feature units whose values were either +1 or 0. Each feature vector was associated with a unique output vector whose activation values were either 0 or +1.

The network learned to associate an input vector with its corresponding output vector using the back propagation algorithm. The network was trained to recognize the three Typical Feature Vectors_{1-3}, [1,0,1,0,1,0,1,0,1,0], [1,1,0,0,1,0,1,0,1,0], [1,0,0,0,1,1,1,0,1,0], and one Atypical Feature Vector_4, [0,0.5,0.5,1,0,0.5,1,0, 0.5, 1]. Each typical feature vector differed from the other typical feature vectors by 2 units, while the atypical feature vector differed from the typical feature vectors by 6 units. Instead of presenting the veridical feature vectors for learning, permutations of the veridical vectors were created by flipping the value of one of the ten input units to either 0 or 1. For example, a permutation of the first feature of Typical Feature Vector_1 is the vector [0,0,1,0,1,0,1,0,1,0]. A total of 32 feature vectors, eight permutations of the four veridical feature vectors, were presented randomly for learning to the network. Each typical and atypical feature vector was associated with a unique output vector of either [1,0,0,0], [0,1,0,0] or [0,0,1,0] for Typical Face_1, Typical Face_2, and Typical Face_3, respectively, or [0,0,0,1] for Atypical Face_4. Training continued for 100 epochs, and ten simulations were run. Before the start of each simulation, the connection weights were initialized to small, random values.

At the conclusion of training, three typical–atypical morph vectors were produced by averaging the feature values of each typical face vector with the corresponding values of the atypical face vector. For example, the features of Typical Feature Vector_1 [1,0,1,0,1,0,1,0,1,0] and Atypical Feature Vector_4 [0,0.5,0.5, 1,0,0.5, 1,0,0.5,1] were averaged to produce the Morph_{1-4} vector of [0.5, 0.25,0.75, 0.5,0.5,0.25,1, 0,0.75,0.5]. After averaging, the three morph vectors were presented to the network and their output activations recorded. Output activation ranged from 0.0 to 1.0.

7.2. Results and discussion

When presented to the network, the morph vectors elicited differential amounts of activation in the four face outputs as confirmed by the one-way ANOVA (for Morph_{1-4}, $F(3,36) = 16.31$, $MSe = 227.4$, $P < 0.001$; for Morph_{2-4}, $F(3,36) = 16.54$, $MSe = 258.9$, $P < 0.001$; for Morph_{3-4}, $F(3,36) = 11.83$, $MSe = 197.9$, $P < 0.001$). Consistent with the density hypothesis, the morph vector produced reliably higher levels of activation in the atypical face output than the typical face outputs (see Fig. 6). When Morph_{1-4} was presented to the network, Typical Face_1 captured 30% of output activation as compared to 40% of the activation captured by the Atypical Face_4, $P < 0.05$. Similarly, Morph_{2-4} generated a level of 27% activation in Typical Face_2 and 43% in Atypical Face_4, $P < 0.05$. When Morph_{3-4} was presented, Typical Face_3 produced 28% of the output activation versus 41% of the activation in Atypical Face_4, $P < 0.10$. In all three cases when a morph of a typical and an atypical vector was presented as input to the network, the atypical face output unit was more strongly activated than the typical face output unit.

By averaging the vector values of atypical and typical representations, morph vectors were constructed such that they were located equidistant from the atypical

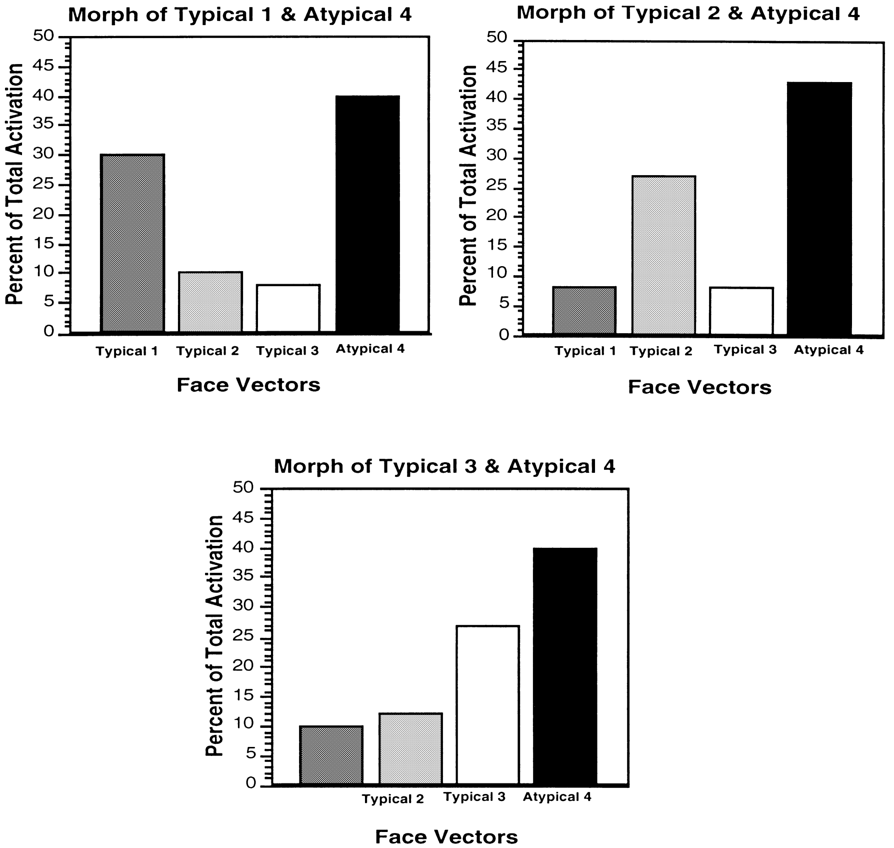

Fig. 6. Results of neural network simulation. In a–c, it is shown that when the morph vector was presented as input to the network, a larger percentage of the output activation was captured by the atypical units than the typical units.

and typical representations in Euclidean space. The main finding of the simulation was that the morph vectors produced stronger activation in the atypical face output unit than of the typical face output units. Hence, similar to the behavioral results, the network demonstrated a preference for atypical representations over typical representations. Recognition in a neural network appears to be determined by two factors; the match between the input and the learned representations and the density of the representations. As shown in these simulations, when the match between competing representations was equated, the representation located in a sparser area of similarity space won out. It is important to point out that this simulation was performed on a relatively simple network composed of only three layers of units. Further simulations are advised to test the scalability of these results (Hinton and Shallice, 1991). Nevertheless, the present simulation demonstrates that neural network recognition is sensitive to the clustering and distribution of representations in the similarity space.

8. General discussion

We began by defining an attractor field as the range of face inputs that are captured by a given face representation. In these experiments, the attractor fields of competing representations were tested in a task involving recognition of a morph face. In Expt. 1, it was found that morph faces containing equal contributions from atypical and typical parent faces were perceived to bear a stronger resemblance to the atypical parent than the typical parent. Contrasting effects were obtained in Expt. 2 where an unrelated morph face was perceived to bear a stronger resemblance to the typical rather than the atypical parent image. The perceptual basis of the atypicality bias was established in Expt. 3 where participants demonstrated a preference for the atypical face in a simultaneous presentation paradigm In Expt. 4, we found that the attractor boundary (i.e., the point at which that morphed image bore equal resemblance to the typical and atypical parent face) was interpolated to correspond to morph image of 37% contribution from the atypical face and 63% contribution from the typical face. Finally, the computational basis of the atypicality bias was demonstrated in a neural network simulation where the morphed vectors more strongly activated atypical face representations than typical face representations. Thus, when interpoint distance is equated through morphing, the behavior and simulation findings indicate that atypical representations attract a wider range of inputs than typical representations.

The obtained atypicality bias seems not to reflect a difference in the relative memorability of atypical and typical faces based on several lines of evidence. First, when the memory load of the likeness was eliminated in Expt. 3, subjects nevertheless demonstrated a bias for the atypical face. Second, In Expt. 4, subjects' likeness judgments varied linearly as a function of the relative contributions of the atypical and typical parent to its derived morph. This finding indicates that typical face representations could be accessed if information in the morph face was sufficiently weighted in the direction of the typical face. Finally, recent work in our laboratory shows that when the memorability for atypical and typical faces was equated through training, the biases toward the atypical face in the 50/50 morph are still evident. The atypicality bias therefore seems not to stem from the more strongly encoded memories for atypical faces, but from the organization of face representations in face space.

These results are compatible with the density hypothesis that allows for some representations to exert a recognition advantage over other representations depending on their locations in face space. Specifically, atypical representations by virtue of their locations in a sparser region of face space possess larger attractor fields than typical representations. Consequently, larger attractor fields enable atypical representations to be activated by a broader class of inputs (e.g. 50/50 morph inputs). In contrast, these results are not consistent with the nearest-neighbor hypothesis that claims that face stimuli will activate the closest representation in Euclidean space. Indeed, these studies showed that in some cases, morph representations that were closer in interpoint distance to their typical parents were more likely to activate their more distant atypical parents.

These results appear to violate a central assumption of the Euclidean model that similarity is characterized by a ‘a linear or monotonic decreasing function to the interpoint distances in the metric space, that is, the larger, the measure of similarity between two objects, the smaller the distance between the corresponding points in the metric space’ (Krumhansl, 1978, p. 445). It follows from this assumption that a representation that lies at the midpoint between two endpoint representations should bear equal resemblance to each of the endpoint representations. Contrary to the strict Euclidean view, the current experiments demonstrate that morph representation despite being located equal distant from the atypical and typical representations was judged to be more similar to the atypical representation. These results are consistent with the similarity model initially proposed by Krumhansl (1978) in which spatial density as well as interpoint distance is claimed to play a role in perceived similarity.

However, Krumhansl (1978) predicted that the spatial density would influence similarity judgments in the opposite direction to the direction revealed in the present experiments. Given the greater density of representations surrounding typical representations, Krumhansl speculated that similarity judgments should be biased toward the typical rather than atypical representations. Support for the typicality bias effect comes from recent research in the speech perception domain. In what they referred to as the *magnet effect*, Kuhl (1991) and colleagues (Iverson and Kuhl, 1995) found that when subjects hear a speech sound that is equidistant from a phonetic prototype and an atypical phonetic exemplar, it is perceived as more similar to the phonetic prototype. Why might subjects show a bias toward typical exemplars in the speech experiments by Kuhl and colleagues and an atypicality bias in our face recognition study? The difference may depend on whether the underlying cognitive task requires that stimuli be grouped into a common category or differentiated into separate representations. In speech perception, for example, the listener must classify slightly different speech tokens as belonging to the same phonetic category and ignore variations in the speech signal that are not phonetically relevant.

Successful face recognition, on the other hand, requires that perceptually similar faces be discriminated and identified as unique individuals. The face recognition system is therefore acutely tuned to the distinctive properties of a face (i.e. atypical features) that differentiates it from other competing faces in memory. Whereas the goal of speech perception is to find commonalties between the speech signal and its phonetic prototype, the objective of the face recognition system is to differentiate face exemplars from each other with respect to the central face prototype. Thus, the contrasting goals of the grouping versus differentiation may account for the typicality bias in speech perception and the atypicality bias in face recognition.

Several questions emerged from the present set of experiments that warrant empirical investigation. First, these experiments leave open the issue of whether faces are perceived categorically or continuously. On one side of the issue, Beale and Keil (1995) found that subjects perceived a continuum of famous faces as shifting abruptly from one famous face (e.g. Bill Clinton) to another (e.g. John Kennedy). In contrast, the results of Expt. 4 indicated that a continuum of atypical and typical faces was perceived continuously. What might account for the discrepancy between these two studies? Whereas Beale and Keil generated continua using well-known, famous faces, we used faces that were previously unfamiliar to subjects. In the attractor field model, the continuous or categorical face recognition is reflected in the slope of the attractor field. The weakly encoded representations of unfamiliar faces would be expected to have less clearly defined boundaries with graded activation slopes whereas the boundaries of familiar face representations should be better defined with steeper activation slopes. Similar effects of familiarity have been reported in the speech perception literature where linguistic experience has been shown to affect the continuous or categorical perception of speech sounds (Miyawaki et al., 1975). In support of the continuous-to-categorical shift, Beale and Keil found that categorical perception positively correlated with face familiarity.

A related issue is whether the familiarity of the parent faces influences the way the morph face is perceived. For example, suppose a morph face was generated from a familiar parent face that was ordinary-looking and a less familiar parent face that was atypical in appearance. The question is whether subjects will perceive the morph as bearing a stronger resemblance to the more familiar face or the more atypical face. The morph paradigm provides a promising experimental technique for pitting the influences of familiarity against the influences of typicality.

Finally, these results raise the question of whether the atypicality bias is unique to faces or might other objects show similar effects. As members of the same stimulus class, all faces share the same internal features (i.e. eyes, nose and mouth) arranged in a similar spatial configuration (i.e. the eyes are above the nose which is above the mouth). Individuation of faces must therefore be based on subtle differences between the shape of the features and their spatial configuration or what Diamond and Carey (1986) have referred to as ‘second order relational properties’. It is possible that other objects whose members are differentiable on the basis of second-order relational properties might show a similar atypicality bias (e.g. birds, automobiles). This question awaits further empirical test.

In conclusion, the present work stresses how recognition processes are influenced by the category structure of competing representations in memory. While important work in object recognition has focused on the mechanisms by which an image-based description is matched to an object representation (Hoffman and Richards, 1984; Biederman, 1987), the current experiments suggest that image-based theories may not tell the whole story behind face and object recognition. Recent work examining the interactions between perception, recognition and categorization (Goldstone, 1994, 1995; Schyns and Rodet, 1998) suggests that a complete theory of object and face recognition should consider the context in which the recognition process develops. Consistent with the contextual viewpoint of recognition, our experiments suggest that how an object is perceived and recognized is influenced by the experience of the observer and the category structure of the object to be recognized.

Acknowledgements

This research has benefited from the suggestions and advice of Al Porterfield. We would also like to thank Bill Friedman for his comments on a earlier draft of the manuscript and Marjorie Henderson for her technical assistance. This research was supported by a Hughes Research Grant, Keck Foundation Faculty Research Award and National Science Foundation (SBR-9729030).

References

Bartlett, J.C., Hurry, S., Thorley, W., 1984. Typicality and familiarity of faces. Memory and Cognition 12, 219–228.

Beale, J.M., Keil, F.C., 1995. Categorical effects in the perception of faces. Cognition 57, 217–239.

Biederman, I., 1987. Recognition-by-components: a theory of human image understanding. Psychological Review 94, 115–147.

Bruck, M., Cavanagh, P., Ceci, S., 1991. Fortysomething: Recognizing faces at one’s 25th reunion. Memory and Cognition 19, 221–228.

Burton, M., Bruce, V., Johnston, R.A., 1990. Understanding face recognition with an interactive activation model. British Journal of Psychology 81, 361–380.

Calder, A.J., Young, A.W., Perrett, D.I., Etcoff, N.L., Rowland, D., 1996. Categorical perception of morphed facial expressions. Visual Cognition 3, 81–117.

Diamond, R., Carey, S., 1986. Why faces are and are not special: An effect of expertise. Journal of Experimental Psychology: General 115, 107–117.

Ekman, P., Friesen, W.V., Ellsworth, P., 1972. Emotion in the Human Face. Pergamon, New York.

Etcoff, N.L., Magee, J.J., 1992. Categorical perception of facial expressions. Cognition 44, 227–240.

Goldstone, R.L., 1994. Influences of categorization on perceptual discrimination. Journal of Experimental Psychology: General 123, 178–200.

Goldstone, R.L., 1995. Effects of categorization on color perception. Psychological Science 6, 298–304.

Hill, H., Schyns, P.G., Akamatsu, S., 1998. Information and viewpoint dependence in face recognition. Cognition 62, 201–222.

Hinton, G.E., Shallice, T., 1991. Lesioning an attractor network: Investigations of acquired dyslexia. Psychological Review 98, 74–95.

Hoffman, D.D., Richards, W., 1984. Parts in recognition. Cognition 18, 65–96.

Iverson, P., Kuhl, P.K., 1995. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. Journal of the Acoustical Society of America 97, 553–562.

Krumhansl, C.L., 1978. Concerning the applicability of geometric models to similarity date: the interrelationship between similarity and spatial density. Psychological Review 85, 445–463.

Kuhl, P.K., 1991. Human adults and human infants show a ‘perceptual magnet effect’ for the prototypes of speech categories, monkeys do not. Perception and Psychophysics 50, 93–107.

Leder, H., Bruce, V., 1998. Local and relational aspects of face distinctiveness. Quarterly Journal of Experimental Psychology, in press.

Lewis, M.B., Johnston, R.A., 1998. Caricature effects from exemplar-based models. Visual Cognition, in press.

Light, L.L., Kayra-Stuart, F., Hollander, S., 1979. Recognition memory for typical and unusual faces. Journal of Experimental Psychology: Human Learning and Memory 5, 212–228.

McClelland, J.L., Rumelhart, D.E., 1986. Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises. Bradford Books, Cambridge, MA.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J., Fujimura, O., 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Perception and Psychophysics 18, 331–340.

Rhodes, G., Brake, S., Atkinson, A., 1993. What’s lost in inverted faces?. Cognition 17, 25–57.

Rhodes, G., Brennan, S., Carey, S., 1987. Identification and ratings of caricatures: Implications for mental representations of faces. Cognitive Psychology 19, 473–497.

Schyns, P.G., Rodet, L., 1998. Categorization creates functional features. Journal of Experimental Psychology: Learning, Memory and Cognition 23, 681–696.