Perception, 2006, volume 35, pages 659–670

Simulating the "other-race" effect with autoassociative neural networks: Further evidence in favor of the face-space model

Roberto Caldara* & Hervé Abdi†

* **Department of Psychology, University of Glasgow,58, Hillhead Street Glasgow, G12 8QB, UK.**
e-mail: r.caldara@psy.gla.ac.uk
† **School of Brain and Behavioral Sciences, The University of Texas at Dallas, Richardson, TX 75083-0688, USA.** e-mail: herve@utd.edu

Abstract.
Other-race (OR) faces are less accurately recognized than same race (SR) faces, but classified faster by race. This phenomenon has been often reported as the "other-race" effect (ORE). Valentine (1991) proposed a theoretical multi-dimensional face-space model that explained both of these results, in terms of variations in exemplar density between races. According to this model, SR faces are more widely distributed across the dimensions of the space when compared to OR faces. However, this model does not quantify nor states the dimensions coded within this face-space. The aim of the present study was to test the face-space explanation of the ORE with neural network simulations by quantifying its dimensions. The predicted density properties of Valentine's framework were found in the face projection spaces of the autoassociative memories. An interaction was found between the race of the learned face set and the race of the faces. In addition, the elaborated face representations showed optimal responses for SR but not for OR faces within SR face-spaces when explored at the individual level. Gender errors occurred significantly more often in OR than in SR face-space representations. Altogether, our results add further evidence in favor of a statistical exemplar density explanation of the ORE as suggested by Valentine, and question the plausibility of such coding for faces in the framework of recent neuroimaging studies.

1. Introduction

In social interactions, it is crucial to identify other people and to extract rapidly the information relevant for communication. The human face is a particular visual object composed of features arranged in definite spatial locations (e.g., two horizontally placed eyes above a centrally placed nose and mouth) that readily provides this information. The uniqueness of the facial features for a given individual, as well as their specific configuration, determines face recognition and identification. At the same time, faces provide additional (semantic) cues about the emotional state, age, sex, and race (Bruce and Young 1986). These cues can be used to classify people in various categories such as young or old, man or woman, Asian or Caucasian, *etc.* Interestingly, the latter semantic information, the race of a face, significantly modulates our proficiency in face recognition. Indeed, it is a well known phenomenon that people are less accurate in recognizing faces of a different race (Feingold 1914). This differential recognition ability is often referred to in the literature as the "other-race" effect (ORE) and has been well documented under laboratory and field studies during the last thirty years (for a review, see Meissner and Brigham 2001; for a review, see Valentine *et al.,* 1995). The robustness and the reliability of this phenomenon across different racial groups has also been confirmed in meta-analytic studies of face recognition tasks (Bothwell *et al.,* 1989; Meissner and Brigham 2001; Shapiro and Penrod 1986), which established that same-race (SR) faces are recognized better and faster than other-race (OR) faces. Nevertheless, even though several hypotheses have been proposed to explain the ORE, the underlying mechanisms of this phenomenon are still matter of an active debate (Meissner and Brigham 2001).

The impairment in recognizing OR faces, is not the unique facet of the ORE. For example, when faces are classified on the criterion of race, OR faces are faster classified than SR faces (Caldara *et al.,* 2004; Levin 1996; Valentine and Endo 1992). For instance, Caucasian subjects need less time to categorize a face as Asian than they need to categorize a Caucasian face as Caucasian (the reverse pattern is observed for Asian subjects). These results are provocative because they seem paradoxical: why are we faster at classifying the faces which we find the hardest to recognize?

Valentine (1991; see also Valentine and Endo 1992), inspired by the categorization literature (Medin and Shaffer 1978; Nosofsky 1986), proposed a theoretical framework based on an exemplar model which specifies the coding and representation of faces (and also facial distinctiveness and the face inversion effect), in order to account for the recognition and classification observations of the ORE. In this psychological model, a face is encoded and positioned as a point in an n-face dimensional space which is constructed through experience with the encountered faces. The dimensions represent factors (physiognomic features, such as color of the hair, eye width, etc) that maximize the discrimination between encountered faces. The origin of the face-space is the mean on all the dimensions and it is positioned at the place with the maximum exemplar density. Typical faces are encountered more frequently, and therefore are located near the origin. In contrast, distinctive faces are located far from the origin (but for a critical review on this point see, Burton and Vokey 1998).

In the face-space model, the dimensions of the space are elaborated with visual experience, and therefore a small number of OR face exemplars (due the lack of experience with these faces) are located in a high-density cluster of faces compared to a large number of exemplars of SR faces, which are widely distributed across the dimensions (see figure 1).

The high density pattern for OR faces results from the small amount of variation across dimensions that are optimal for discriminating SR faces. For instance, the color of the hair and the eyes represent useful diagnostic (discriminative) visual information in a Caucasian face-space, but the same information is not diagnostic for Asian faces. As a consequence, Caucasian faces will be more distributed across those dimensions than Asian faces. This particular differential density, between the races in the face-space representations, results in a high perceived similarity between OR faces, which is also in turn reflected in the popular belief that OR faces "all look alike" (Feingold 1914). More importantly, such patterns explain the poor face recognition performance for OR faces, because the higher density for OR faces increase the difficulty of the discrimination of different exemplars. At the same time, the classification advantage for OR faces could be accounted for faster activation of nearby individuals, all leading to the same and located in a small, high density cluster compared to a larger one (for SR faces). Thus, this model provides a useful and elegant account of the ORE and on how faces might be stored in human memory; it also posits the differential level of experience, or contact, with OR faces as the cause of the ORE (Valentine and Endo 1992). Nevertheless, the critical limitation of the theoretical face-space model (Valentine 1991; Valentine and Endo 1992) is that it does not clearly state what is encoded in the dimensions of the face-space. Indeed, in this model the dimensions are arbitrary and so are neither specified nor quantified. Thus, such a model can explain different results by positing

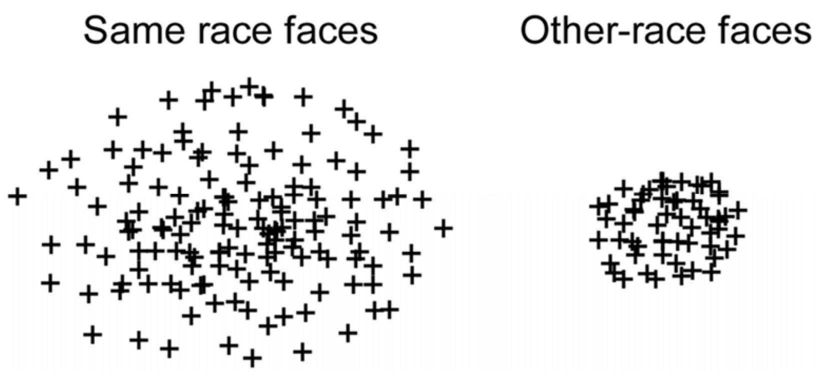

FIGURE 1. A two-dimensional representation of the face-space model suggested by Valentine (1991). On the left, a large number of exemplars of same-race faces widely distributed across diagnostic information compared to a small number of other-race face exemplars (on the right), that are located in a high-density space.

a variation in dimension weights or a shift in space location. For example, Valentine and Endo (1992) discussed these changes in the multidimensional face-space in terms of selective attention. In addition, in the face-space model, the projections of the faces on the dimensions are assumed to be normally distributed and having the same variance. As pointed by Burton and Vokey (1998), these last two assumptions are somewhat problematic because they imply that the distribution of the distances of the faces to their center follows a chi-square distribution (with the number of degrees of freedom begin the number of dimensions of the model). This means, in such a model, that most faces are *far* from their common center when the number dimensions is large. Burton and Vokey suggest that one way of addressing this problem is to specify the dimensions by using the statistical properties of *real* faces. Such a procedure will also fix the values of the weights of the dimensions because these weights as well as the dimensions will be dependent upon the sample of faces learned. An additional benefit of this approach is to suggest that perceptual learning can act as the *psychological* mechanism responsible for the acquisition of these dimensions.

Recently, the concept of *perceptual learning* has been proposed as an important mechanism for understanding how differential experience affects the way we process SR and OR faces (O'Toole *et al.,* 1995). Classically, this concept has been defined by Gibson (1969, p. 3) as *"an increase in the ability to extract information from the environment, as a result of practice and experience with stimulation coming from it".* According to the differential experience hypothesis, as face recognition skills develop, individuals learn to use the perceptual dimensions that are optimal for discriminating among individual human faces. Evidence in favor of this hypothesis has been reported by O'Toole and colleagues (O'Toole *et al.,* 1995; O'Toole *et al.,* 1991; O'Toole *et al.,* 1994) with computational models based on face-pixel autoassociative neuronal networks (Abdi 1988; Abdi *et al.,* 1999). When their neural networks were trained to recognize a majority of faces of one race (Caucasian or Asian), they better reconstructed novel faces from the majority race than faces from the other-race. More recently, Furl *et al.,* (2002) further investigated this hypothesis of the ORE by training 13 face-recognition algorithms with Caucasian faces as the majority race. Overall, these authors refined and improved previous results from the ORE simulations by showing that this effect is present only in algorithms that generate representations in the face-space distorted to emphasize features that permit the individuation of faces.

Taken together, the theoretical face-space model (Valentine 1991; Valentine and Endo 1992) and perceptual learning theory have at least two major points in common. First, as previously suggested in the literature (e.g. Brigham 1986; Brigham *et al.,* 1982; Brigham and Malpass 1985; Platz and Hosch 1988), they both consider that differential experience is a crucial factor in the ORE explanation. Second, they both support a difference at the level of internal representations as a theoretical explanation of the ORE. Surprisingly, despite such similarities, the relationship between these two theoretical positions of the ORE has not yet been directly assessed. More precisely, none of the previous studies based on perceptual learning theory analyzed the structure of the internal representation put forward by the neural network simulations of the ORE (Furl *et al.,* 2002; O'Toole *et al.,* 1995; O'Toole *et al.,* 1991; O'Toole *et al.,* 1994), in terms of the predicted density properties of the Valentine's model (Valentine 1991; Valentine and Endo 1992). The spatial distribution of face representations was never measured, and therefore the differential density hypothesis as an explanation of the ORE remains an open question. Moreover, to our knowledge, the structure of the representations created by the neuronal network simulations was never investigated at individual level (one face compared with the others).

From this logic, the purpose of the present study was to quantify the multidimensional theoretical face-space proposed by Valentine, on the basis of statistical neuronal network computations, to measure the spatial distribution of face representations and to verify its predictions of the ORE. We took advantage of the PCA approach, as this technique represents an optimal tool for analyzing the perceptual and statistical information for faces (Abdi 1988), and clearly identifies the projections for each single face. First, we hypothesized that the face-space will be more scattered for SR than OR faces. Obviously, these predictions will be sensitive to the set of learned faces (*i.e.,* Caucasian as SR faces; Asian as OR faces), and the reverse pattern of results should be expected for the converse learned set of faces (*i.e.,* Asian as SR faces; Caucasian as OR faces). Second, because the faces will be distributed in the face-space as a function of their similarity (Valentine 1991), the distance between any two points is analogous to the similarity between two faces. Thus, for SR faces it should be expected that faces of the same gender would share a similar location in the space (Abdi *et al.,* 1995). However, the neuronal network face memory space is shaped through the experience by the race of the learned faces, and thus it is better constructed and tailored for the SR faces than OR faces (Furl *et al.,* 2002; O'Toole *et al.,* 1995; O'Toole *et al.,* 1991; O'Toole *et al.,* 1994). As a consequence, we hypothesized that SR face-space would be more adapted on such perceptual dimensions (gender) for SR faces, but its representations would be defective for OR faces. Indeed, behavioral studies reinforce such hypothesis, since it has been demonstrated that we are better at classifying by gender (O'Toole *et al.,* 1996) or at estimating the age (Dehon and Bredart 2001) of SR faces than OR faces. These observations suggest SR faces have a more adapted psychological representations than OR faces.

2. Methods

2.1. *Stimuli.* Three hundred and sixteen grayscale photographs of Caucasian (79 male, 79 female) and Asian faces (79 male, 79 female) were digitally scanned using 16 gray levels for use in our simulations. This face database was created by authors from The University of Texas at Dallas and has been previously used in various neuronal network simulations (*e.g.,* Abdi *et al.,* 1999; O'Toole *et al.,* 1995; O'Toole *et al.,* 1991; O'Toole *et al.,* 1994). It is worth noting that the same stimuli have been used also in behavioral studies that demonstrated an ORE in Asian and Caucasian participants for their respective OR faces (O'Toole *et al.,* 1994; O'Toole *et al.,* 1996).

All the pictures were cropped around the face to remove clothing. Male faces were clean-shaven and none had particularly distinctive features. All images showed a frontal view, with the eyes being roughly aligned on the horizontal midline of the image.

2.2. *Procedure.*

2.2.1. *Principal component analysis approach.* We simulated a *pure* Caucasian and an Asian developmental experience by training two neural networks with only their respective races. For the Caucasian learning, 158 grayscale photographs of Caucasian faces were coded in a pixel-based I-dimensional vector, concatenated from the columns of the face image (with I representing the number of pixels in the images, and k indexing the $K = 158$ faces). The vectors are normalized so that $\mathbf{x}_k^T \mathbf{x}_k = 1$. The set of K faces is represented by an $I \times K$ matrix in which the k-th column is equal to $\mathbf{x}_k$. The faces are stored in an autoassociative memory, in which each unit is connected to all the other units, and the intensity of the connections is represented by an $I \times I$ matrix $\mathbf{W}$. We used a Widrow-Hoff learning rule, which iteratively corrects the weights of the network as a function of the quality of the response as follows:

$$\mathbf{W}_{[t+1]} = \mathbf{W}_{[t]} + \eta \left(\mathbf{X} - \mathbf{W}_{[t]}\mathbf{X}\right)\mathbf{X}^T \tag{1}$$

where is η a small positive constant (typically smaller than one).

As previously demonstrated (Anderson *et al.*, 1977; Kohonen 1977), since the weight matrix is a cross-product matrix, it is positive semi-definite (*i.e.*,, all its eigenvalues are positive or zero, and all its eigenvectors are real). As a consequence, $\mathbf{W}$ can be expressed as a weighted sum of its eigenvectors:

$$\mathbf{W} = \sum_{\ell=1}^{L} \lambda_\ell \mathbf{u}_\ell \mathbf{u}_\ell^T = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T \quad \text{with } \mathbf{U}^T\mathbf{U} = \mathbf{I} \tag{2}$$

where $\mathbf{u}_\ell$is the ℓ-th eigenvector of $\mathbf{W}$, λ_ℓ the ℓ-th eigenvalue, $\mathbf{I}$ represents the identity matrix, $\boldsymbol{\Lambda}$ represents the $L \times L$ diagonal matrix of the eigenvalues, $\mathbf{U}$is the $I \times L$ matrix of eigenvectors, and L is the rank of the matrix $\mathbf{W}$. The eigenvectors in $\mathbf{U}$ are generally ordered according to their eigenvalues. In what follows, the eigenvector with the largest eigenvalue is referred to as the first eigenvector; the eigenvector with the second largest value is referred to as the second eigenvector, and so on. The eigenvectors and eigenvalues of the weight matrix can be obtained directly using the singular value decomposition (*cf.*, *e.g.*,, Horn & Johnson 1985) of the face matrix. It is formally expressed as:

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T \tag{3}$$

where $\mathbf{U}$ represents the matrix of eigenvectors of $\mathbf{X}\mathbf{X}^T$, $\mathbf{V}$ represents the matrix of eigenvectors $\mathbf{X}^T\mathbf{X}$, and $\boldsymbol{\Delta}$ stands for the diagonal matrix of singular values, which are equal to the square roots of the eigenvalues of $\mathbf{X}\mathbf{X}^T$ and $\mathbf{X}^T\mathbf{X}$ (they are the same).

The estimation of a face by the system can, thus, be represented as a weighted sum of eigenvectors:

$$\widehat{\mathbf{x}}_k = \sum_{\ell=1}^{L} \lambda_\ell \mathbf{u}_\ell \mathbf{u}_\ell^T \mathbf{x}_k = \sum_{\ell=1}^{L} \lambda_\ell \gamma_\ell \mathbf{u}_\ell \quad \text{with } \gamma_\ell = \mathbf{u}_\ell^T \mathbf{x}_k \tag{4}$$

where the weights γ_ℓ are the projections of the faces onto the eigenvectors. These weights can be interpreted as an indication of the extent to which a given eigenvector (or "macrofeature") characterizes a particular face. When the Widrow-Hoff learning is used, this process corresponds to:

$$\mathbf{W} = \mathbf{U}\mathbf{U}^T \tag{5}$$

and the estimation of a face is obtained by dropping the eigenvalues in:

$$\widehat{\mathbf{x}}_k = \sum_{\ell=1}^{L} \gamma_\ell \mathbf{u}_\ell \quad \text{with } \gamma_\ell = \mathbf{u}_\ell^T \mathbf{x}_k \tag{6}$$

This is equivalent to giving the same importance to each eigenvector in the reconstruction of a face. More formally, Widrow-Hoff learning amounts in sphericizing the weights matrix $\mathbf{W}$.

We used a Widrow-Hoff learning rule followed by a *jackknife* procedure: we took off one face from the set in an iterative way for the entire Caucasian face database, and computed the values of its projections on the eigenvectors computed from all the other faces. Then, the entire set of Caucasian faces was memorized and a set of 158 pixel-based Asian faces was tested with the Caucasian autoassociative memory. In this step we computed the values of the projections of all Asian faces. We limited our analysis into a three-dimensional space with the axis defined by the first three eigenvectors. We decided this, mainly, for three reasons. First, more than 90% of the variance is captured by the three first eigenvectors. Second, it has already been demonstrated that the macro-characteristics of the faces are captured in the first eigenvectors, the last eigenvectors encoding characteristics of individual faces (O'Toole *et al.,* 1993; Valentin and Abdi 1996; Valentin *et al.,* 1994). Thus, the important information for the race of the faces lies on the first eigenvectors and taking into account all the space projections would shape the face-space into individual rather than global dimensions of the faces. Finally, keeping the eigenvectors in a 3D space, offers also a greater convenience for the illustrations of the face-space. Thus, the barycenters on the first three *eigenvectors* were calculated across all the faces in their respective face-spaces, by separately computing the mean of the projections on each *eigenvector.* Finally, squared Euclidian distances d^2 and the cosine values, were computed as measures of similarity between each face projections of both races and their respective barycenters.

For the Asian learning, the same procedure was applied using the reverse set of faces.

The calculated squared Euclidian distances and cosine values, for both race faces in both learning sets, were separately compared using an ANOVA with the Learned face set (Caucasian/Asian) and Race 3D face projection space (Caucasian/Asian) as factors.

For all the face projections on the first three *eigenvectors* and for both face races in both learning sets, we located the nearest point in the space and identified its gender. A change between the selected face and its nearest neighbor was considered as an error. The number of errors committed by the artificial neuronal networks was compared using an ANOVA with the learned face set (Caucasian/Asian) and Race 3D face projection space (Caucasian/Asian) as factors.

3. **Results**

Figure 2 illustrates the face-space projections for the encoding of each face of both races on the first three eigenvectors after the training with Caucasian or Asian faces.

The Euclidian distances and cosine values were calculated between each face position and their respective barycenters, and the observed means are reported in table 1.

TABLE 1. Means and standard deviations (± SD) of the squared Euclidian distances and cosine values calculated between each face position in the 3D face projection spaces and the respective barycenters.

	Euclidian distances		Cosine values	
	Race of the faces		Race of the faces	
	Caucasian	Asian	Caucasian	Asian
Caucasian learning	0.059 (0.043)	0.028 (0.027)	0.964 (0.027)	0.982 (0.018)
Asian learning	0.027 (0.028)	0.057 (0.050)	0.982 (0.018)	0.965 (0.031)

For the Euclidian distances and the cosine values, the ANOVAs revealed a highly significant, full, crossover interaction between the race of the learned face set and the race of the 3D face projection space, respectively [$F(1,628) = 94.50$, $MSE = .142$, <.0001] for the Euclidian distances and [$F(1,628) = 76.97$, $MSE = .046$, $p < .0001$] for the cosine values. SR faces were widely distributed and more dissimilar compared to OR faces.

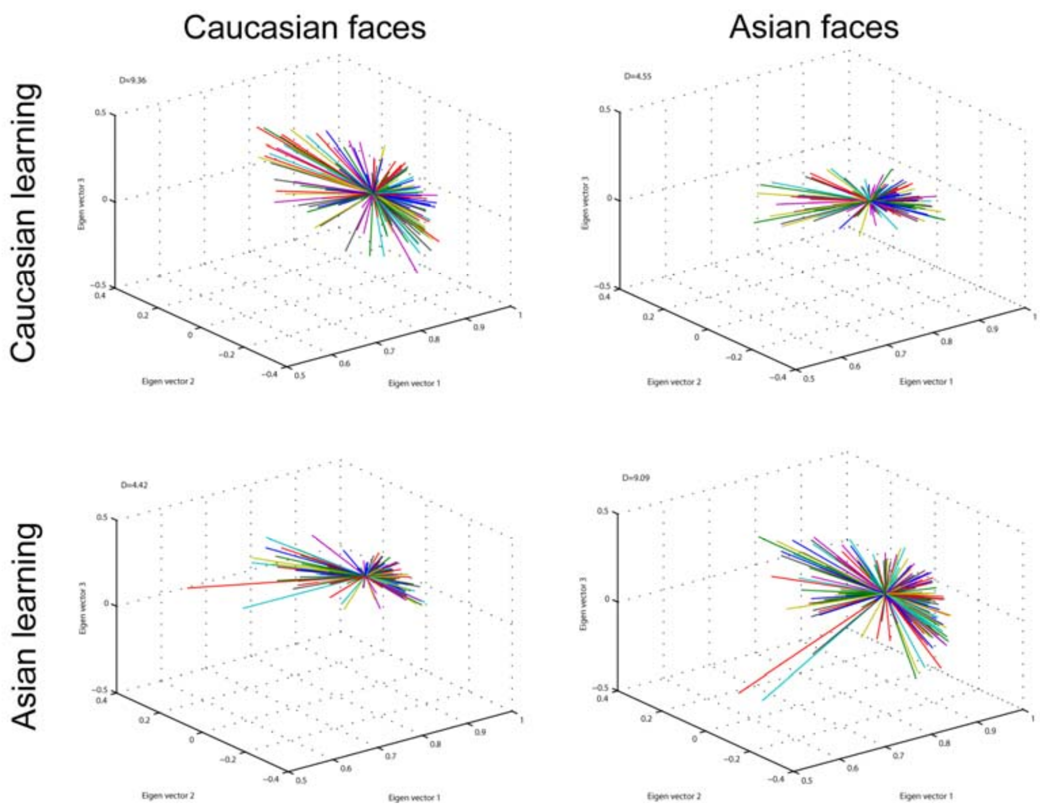

FIGURE 2. Top: 3D Face projection space on the three first eigenvectors after a Caucasian learning. The cumulated distances (D) from the barycenter of each space are reported. Bottom: 3D Face projection space on the three first eigenvectors after an Asian learning. The face-spaces are widely distributed for same- rather than other-race faces.

In the face-space, faces are represented as points. Hence, the distances between faces can be computed and the nearest neighbor of each face can identified. For example, figure 3 illustrates this procedure and displays two selected targets of each race and their neighbors in the SR or OR face-space.

Both race faces were better represented in their SR face-space learning set than in OR one. The visual similarity between neighbors was higher, and the neighbors were of the same gender, in the SR learning conditions than OR learning conditions (see table 2).

These observations were statistically confirmed, by a significant interaction between the race of the learned face set and the errors in the gender of the neighbor [$F(1,628) = 10.18$, $MSE = 1.72$, $p < .01$].

4. **Discussion**

The purpose of the present study was to quantify, on the basis of statistical neuronal network computations, the multidimensional theoretical face-space proposed by Valentine (1991) as a possible explanation of the ORE. Our simulation results showed that the predicted density properties of Valentine's model (see figure 1) are present on the 3D face projection space of the autoassociative memories for both race faces (see figure 2). A significant interaction was found

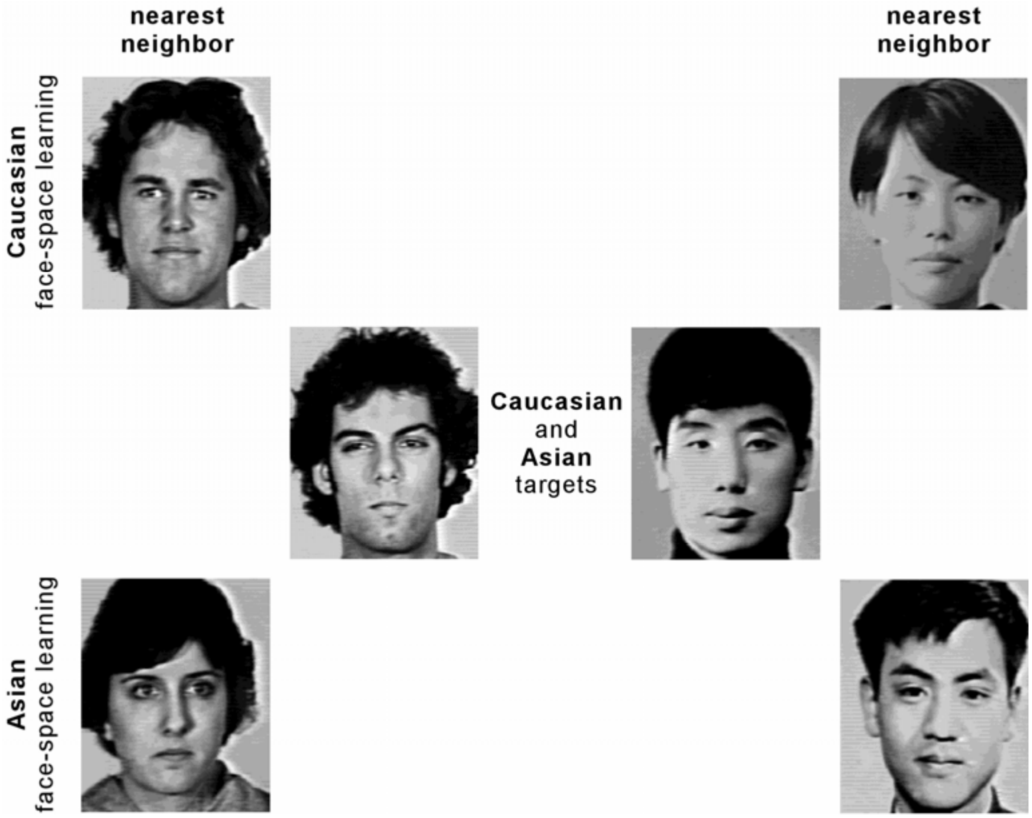

FIGURE 3. A Caucasian and an Asian target faces (middle) and their nearest neighbors found in the face-space representations after the Caucasian (top) and an Asian learning (bottom). Note the high visual similarity between faces in the same-race learning conditions, and gender errors in the other-race learning conditions.

TABLE 2. : Means and standard deviations (± SD) of the percentage of gender errors in the 3D face projection spaces.

	Race of the faces	
	Caucasian	Asian
Caucasian learning	19.6 % (0.39)	24.6 % (0.43)
Asian learning	29.7 % (0.45)	13.8 % (0.34)

between the race of the learned faces and the face projections' density on the three first eigenvectors: SR faces were widely distributed and were less similar compared to OR faces. The face-space created with the experience of one particular race (SR) widely distributed representations for the learned faces, by capturing diagnostic information and a larger amount of the variance on learned faces compared to OR faces. As demonstrated by the cosine values, this process leads to a larger difference in similarity between SR faces compared to OR faces. Altogether, these observations refine the state of knowledge of the role played by these eigenvectors in terms of computational, statistical, perceptual learning (O'Toole *et al.,* 1994; O'Toole *et al.,* 1995), and highlight their crucial role in the explanation of the ORE. It is worth noting that these properties emerged spontaneously from the memory representations of our neuronal network simulations, providing direct evidence—as suggested by Valentine (1991)—that a statistical exemplar density, derived from visual experience, plays an important role in the explanation of the ORE

At this stage, however, in order to estimate the density of the face-spaces we uniquely focused on the distribution of each single face from the barycenter: the *global* properties of face-space representations. Our findings suggest that OR faces are more typical (due to their higher similarity) than SR faces. Nevertheless, as pointed out by Burton and Vokey (1998), for understanding the typicality of the faces, it is also necessary to explore the *local* properties of their representations (one single face compared to the others regardless the barycenter). Indeed, it would be possible to observe a modulation of density properties in function of the race of the faces uniquely when such a measure is calculated by comparison with the barycenter. This result would indicate that crucial factors in the explanation of the ORE rely only on *global* dimensions. The exploration of *local* properties in the 3D face projection space, however, lead to a comparable interaction between the Caucasian (mean cosine value for Caucasian faces = .93; Asian faces = .964) and the Asian learning sets (mean cosine value for Caucasian faces = .965; Asian faces = .932) and their respective SR and OR face-spaces. SR faces were more similar in the OR compared to the SR face-space. In a nutshell, we found that SR faces were widely distributed and were less similar (and typical) at the *global* and the *local* level.

Our artificial models refined the model of Valentine by encoding and placing each face in the space. Crucially, this quantification process offers the advantage to select one face and identify its neighbors. As illustrated in figure 3, we selected one Caucasian and one Asian face and their nearest neighbors in the SR and OR face-spaces. The exploration of the face-spaces at the individual level revealed a better gender discrimination for SR faces, for both Caucasian and Asian faces. More precisely, the neighbors looked alike the target faces and were of the same gender as the target faces significantly more often in the SR than the OR learning conditions. Indeed, when the target faces were projected in their respective OR faces space, their neighbors did not look alike the target face anymore, and also they were significantly more often from the opposite gender than the target face. This pattern of results illustrates how the internal representations created with experience are adapted for SR faces, but not for OR faces. Like humans in gender classification (O'Toole *et al.,* 1996) the network created better representations for SR than OR faces. In other words, face-spaces do not optimally respond to OR faces (Furl *et al.,* 2002), but also they do not distribute OR faces for an optimal gender perception. Typically, because individuals have more experience with SR faces, humans become more expert with the feature dimensions that are diagnostic for distinguishing SR faces (Ellis *et al.,* 1975). As a consequence, perceptual learning enhances individuals' ability to process and recognize SR faces, but it also results in a decreased ability to process and recognize OR faces. These basic assumptions are reflected in the face-space representations created by our neuronal network simulations. To conclude on this point, it is important to stress that this approach can be used in future studies to determine the similarity of a set of faces and verify the validity of some predictions issued from the Valentine's theoretical framework (1991). Firstly, because neighbour faces share a high similarity in the face-space, in a face recognition task, for example, the nearest neighbour of one selected face would be more confusable with this face compared to a distant neighbour. Such a pattern of results would predict longer reaction times and larger errors (e.g. Davies *et al.,* 1979; Light *et al.,* 1979). Secondly, the faces near to the barycenter will be more difficult to recognize than the distant ones, because (as defined by Valentine, 1991) they are typical. Behavioural studies have shown that distinctive faces are better recognized than typical faces (e.g. Going and Read 1974; Light *et al.,* 1979; Valentine and Bruce 1986).

The overall pattern of results also leads to other considerations in the theoretical framework of the OR effect. First, we found that experience plays an important role in the explanation of the ORE. These findings are in line with previous results from neuronal network simulations (Furl *et al.,* 2002; O'Toole *et al.,* 1991; O'Toole *et al.,* 1994; O'Toole *et al.,* 1996) as well as behavioral results (Brigham 1986; Brigham *et al.,* 1982; Brigham and Malpass 1985; Platz and Hosch 1988). Second, our results demonstrated that visual experience with one race modulates the similarity on the macro-dimensions (encoded in the three first eigenvectors) encoding race information, by leading to an increased visual similarity for OR faces in the face-space. These findings confirm the predictions of the Valentine's model (1991). Incidentally, the adequacy of the face model has been recently questioned by Levin (1996; Levin 2000). This author suggested that a feature selection process, driven by social cognitive mechanisms (*e.g.,* social stereotypes applied to out-group members), causes race-specifying information to be coded as a visual feature[1] in OR faces but not in SR faces. However, our findings show that a face-space model can explain the occurrence of the ORE. It is however possible that social cognition factors interact also with perceptual process and might contribute to a part of the ORE, but crucially, the unique visual experience is likely to be a critical factor in the explanation of the ORE. Third, the more distributed representations of SR faces compared to OR faces, modulated the quality of the encoded diagnostic information in favor of SR faces, as showed by gender errors in the OR face-space. This modulation in the quality of the encoded representations might also account for an increased ability to encode configural information for SR faces (Michel *et al.,* in press; Rhodes *et al.,* 1989; Tanaka *et al.,* 2004). Finally, our simulation results suggest that the ORE occurs independently of the race of the faces. In our simulations we found a qualitatively comparable ORE effect for the two sets of faces, which is the expression of a comparable heterogeneity within the faces regardless of race. This observation is in line with an objective anthropometric investigation of facial dimensions conducted by Goldstein (1979a; 1979b) within different races (Black, Caucasian and Asian faces), that yielded no evidence for racial differences in facial heterogeneity. Altogether, these findings do not support the inherent difficulty hypothesis as a possible explanation of the ORE (Malpass and Kravitz 1969). This hypothesis postulates that discrimination difficulty for OR faces is due to the lack of physiognomical variations for OR faces2. Our results add a new evidence to this question and show that OR faces in fact do look alike in the SR learning situation (the Euclidian distances for OR faces are smaller and denser compared to those of the SR face-space), but since this pattern of results is [2]fully reversed when exactly the same faces are used as SR faces, it can be definitely and objectively stated that OR faces don't all look alike! As previously suggested by Valentine (1991), this perceptual effect is largely due to our larger experience with SR faces.

Finally, in the field of cognitive neuroscience there is currently a debate about the emergence of representations in the brain and the computational role played by the activations identified with neuroimaging techniques. In the domain of face perception, as revealed by functional Magnetic Resonance Imaging (*f*MRI) studies, there is a region of the brain that produces at least twice as many responses to faces than to other visual objects (Kanwisher *et al.,* 1997; McCarthy *et al.,* 1997): the right middle fusiform gyrus, the so called 'Fusiform Face Area' (FFA—Kanwisher *et al.,* 1997). Recently, the neural basis of the ORE has been investigated by using such techniques (Golby *et al.,* 2001). The results showed a greater activation of the FFA for SR compared to OR faces during the encoding stage of a face recognition task. Because the visual expertise for faces and non face-objects plays an important role in the activity of the FFA (Gauthier *et al.,* 2000; Gauthier *et al.,* 1999), the greater activity for SR faces was related to the greater visual expertise and experience for those faces (Golby *et al.,* 2001). Of particular interest, the interaction observed at the level of brain activations is in someway comparable to the interaction observed at the level

[1]Note, that Levin (1996, 2000) do not precise the nature of this visual feature.

[2]If faces of a certain race were inherently more homogeneous, the individuals of this race must present a difficulty in the recognition of the other members of their race. Currently, there is no evidence that supports this assumption.

of the face-space representations in our neuronal networks findings. Note that the encoding task used in the functional neuroimaging study is also comparable to our neuronal network simulations procedure, which consisted in the encoding of novel faces. Furthermore, recent neuroimaging studies that investigated the nature of the representations of faces at the brain level, in humans (Loffler *et al.,* 2004) and monkeys (Giese *et al.,* 2004), revealed the existence of a prototype-referenced encoding (Giese *et al.,* 2004; Loffler *et al.,* 2004), a pattern of results that is line with the face-space model. Altogether, these observations suggest that the activity of the FFA might be directly related to the recruitment of perceptual representations elaborated with experience, which are in turn modulated by the race of the faces. Further studies combining neural network simulations and neuroimaging techniques will help to assess and better understand their nature and their exact relationship.

Acknowledgements. Thanks are due to reviewers for helpful comments on an earlier draft of this paper. Correspondence can be sent to: Roberto Caldara Department of Psychology University of Glasgow 58, Hillhead Street Glasgow, G12 8QB UK, `email: r.caldara@psy.gla.ac.uk`, or Hervé Abdi, The University of Texas at Dallas, School of Behavioral and Brain Sciences, MS: Gr.4.1., Richardson TX 75083–0688, `email: herve@utd.edu`. Roberto Caldara was supported by a post-doctoral fellowship provided by the Swiss National Science Foundation.

References

[1] Abdi, H. (1987). A generalized approach for connectionist auto-associative memories: Interpretation, implication, and illustration for face processing. In J. Demongeot (Ed.) *Artificial intelligence and cognitive science.* Manchester: Manchester University Press.

[2] Abdi, H., Valentin, V., Edelman B., (1999). *Neural Networks.* Thousand Oaks (CA): Sage.

[3] Abdi H, Valentin D, Edelman B, O'Toole A J, 1995 "More about the difference between men and women: evidence from linear neural networks and the principal-component approach" *Perception* **24** 539-62.

[4] Anderson J A, Silverstein J W, Ritz S A, Jones R S, 1977 "Distinctive features, categorical perception, and probability learning: some applications of a neural model" *Psychological Review* **84** 413-451

[5] Bothwell R K, Brigham J C, Malpass R S, 1989 "Cross-racial identification" *Personality and Social Psychology Bulletin* **15** 19-25

[6] Brigham J C, 1986 "The influence of race on face recognition", in *Aspects of face Processing* Eds H D Ellis, M A Jeeves, F Newcombe and A W Young (Dordrecht: Dordrecht) pp 170-177

[7] Brigham J C, Maass A, Snyder L D, Spaulding K, 1982 "The accuracy of eyewitness in a field setting" *The Journal of Personality and Social Psychology* **442** 673-678

[8] Brigham J C, Malpass R S, 1985 "The role of experience and contact in the recognition of faces of own- and other-race persons" *Journal of Social Issues* **41** 139-155

[9] Bruce V, Young A, 1986 "Understanding face recognition" *British Journal of Psychology* **77** 305-327

[10] Burton A M, Vokey J R, 1998 "The face-space typicality paradox: Understanding the face-space metaphor" *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* **51** 475-483

[11] Caldara R, Rossion B, Bovet P, Hauert C A, 2004 "Event-related potentials and time course of the 'other-race' face classification advantage" *Neuroreport* **15** 905-910

[12] Davies G M, Shepherd J W, Ellis H D, 1979 "Similarity effects in face recognition" *American Journal of Psychology* **92** 507-23

[13] Dehon H, Bredart S, 2001 "An 'other-race' effect in age estimation from faces" *Perception* **30** 1107-13

[14] Ellis H D, Deregowsky J B, Shepherd J W, 1975 "Description of white and black faces by white and black subjects" *International Journal of Psychology* **10** 119-123

[15] Feingold C A, 1914 "The influence of the environment on identification of persons and things" *Journal of Criminal Law & Police Science* **5** 39-51

[16] Furl N, Jonathon P, O'Toole A J, 2002 "Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis" *Cognitive Science* **26** 797-815

[17] Gauthier I, Skudlarski P, Gore J C, Anderson A W, 2000 "Expertise for cars and birds recruits brain areas involved in face recognition" *Nature Neuroscience* **3** 191-197

[18] Gauthier I, Tarr M J, Anderson A W, Skudlarski P, Gore J C, 1999 "Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects" *Nature Neuroscience* **2** 568-73

[19] Gibson E J, 1969 *Principles of perceptual learning and development* (New York: Appleton-Century-Crofts)

[20] Giese M A, Sigala R, Wallraven C, Leopold D A, 2004 "Physiologically inspired neural model for the prototype-referenced encoding of faces" *Journal of Vision* **4** 213a

[21] Going M, Read J D, 1974 "Effects of uniqueness, sex of subject, and sex of photograph on facial recognition" *Perceptual and Motor Skills* **39** 109-110

[22] Golby A J, Gabrieli J D, Chiao J Y, Eberhardt J L, 2001 "Differential responses in the fusiform region to same-race and other-race faces" *Nature Neuroscience* **4** 845-50

[23] Goldstein A G, 1979a "Facial feature variation: Anthropometric data II" *Bulletin of the Psychonomic Society* **13** 191-193

[24] Goldstein A G, 1979b "Race-related variation of facial features: Anthropometric data I" *Bulletin of the Psychonomic Society* **13** 187-190

[25] Horn R A, Johnson C R, 1985 *Matrix analysis* (Cambridge: C.U.P.) Kanwisher N, McDermott J, Chun M M, 1997 "The fusiform face area: a module in human extrastriate cortex specialized for face perception" *Journal of Neuroscience* **17** 4302-11

[26] Kohonen T, 1977 *Associative memory - A system-theoretical approach* (Berlin: Springer-Verlag)

[27] Levin D T, 1996 "Classifying faces by race: The structure of face categories" *Journal of Experimental Psychology: Learning, Memory, & Cognition* **22** 1364-1382

[28] Levin D T, 2000 "Race as a visual feature: using visual search and perceptual discrimination tasks to understand face categories and the cross race recognition deficit" *Journal of Experimental Psychology: General*

[29] 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

[30] Loffler G, Wilkinson F, Yourganov G, Wilson H R, 2004 "Effect of Facial Geometry on the fMRI signal in the Fusiform Face Area" *Journal of Vision* **4** 136a

[31] Malpass R S, Kravitz J, 1969 "Recognition for faces of own and other race" *Journal of Personality & Social Psychology* **13** 330-334

[32] McCarthy G, Puce A, Gore J C, Allison T, 1997 "Face-specific processing in the human fusiform gyrus" *Journal of Cognitive Neuroscience* **9** 604-609

[33] Medin D L, Shaffer M M, 1978 "Context theory of classification learning" *Psychological Review* **85** 207-238

[34] Meissner C A, Brigham J C, 2001 "Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review" *Psychology, Public Policy and Law* **7** 3-35

[35] Michel C, Caldara R, Rossion B, in press "Same-race faces are perceived more holistically than other-race faces" *Visual Cognition*

[36] Nosofsky R M, 1986 "Attention, similarity, and the identification-categorization relationship" *Journal of Experimental Psychology General* **115** 39-61

[37] O'Toole A J, Abdi H, Deffenbacher K A, Valentin D, 1993 "A low dimensional representation of faces in the higher dimensions of the space" *Journal of the Optical Society of America, series A* **10** 405-411

[38] O'Toole A J, Abdi H, Deffenbacher K A, Valentin D, 1995 "A perceptual learning theory on the information in faces", in *Cognitive and Computational Aspects of Face Recognition* Ed T Valentine (London: London)

[39] O'Toole A J, Deffenbacher K A, Abdi H, Bartlett J C, 1991 "Simulating the 'Other-race Effect' as a Problem in Perceptual Learning" *Connection Science* **3** 163-178

[40] O'Toole A J, Deffenbacher K A, Valentin D, Abdi H, 1994 "Structural aspects of face recognition and the other-race effect" *Memory and Cognition* **22** 208-224

[41] O'Toole A J, Peterson J, Deffenbacher K A, 1996 "An 'other-race effect' for categorizing faces by sex" *Perception* **25** 669-676 Platz S J, Hosch H M, 1988 "Cross-racial/ethnic eyewitness identification: A field study" *Journal of Applied Social Psychology* **18** 972-984

[42] Rhodes G, Tan S, Brake S, Taylor K, 1989 "Expertise and configural coding in face recognition" *British Journal of Psychology* **80** 313-331 Shapiro P N, Penrod S, 1986 "Meta-analysis of facial identification studies" *Psychological Bulletin* **100** 139-156

[43] Tanaka J W, Kiefer M, Bukach C M, 2004 "A holistic account of the own-race effect in face recognition: evidence from a cross-cultural study" *Cognition* **93** B1-9

[44] Valentin D, Abdi H, 1996 "Can a linear autoassociator recognize faces from new orientations?" *Journal of the Optical Society of America, series A* **13** 717-724

[45] Valentin D, Abdi H, O'Toole A J, Cottrell G W, 1994 "Connectionist models of face processing: a survey" *Pattern Recognition* **27** 1208-1230

[46] Valentine T, 1991 "A unified account of the effects of distinctiveness, inversion, and race in face recognition" *Quarterly Journal of Experimental Psychology Human Experimental Psychology* **43a** 161-204

[47] Valentine T, Bruce V, 1986 "Recognizing familiar faces: The role of distinctiveness and familiarity" *Canadian Journal of Psychology* **40** 300-305

[48] Valentine T, Chiroro P, Dixon R, 1995 "An account of the own-race biais and the contact hypothesis based on a 'face space' model of face recognition", in *Cognitive and computational aspects of face recognition: explorations in face space* Ed T Valentine (London: London) pp 69-94

[49] Valentine T, Endo M, 1992 "Towards an exemplar model of face processing: The effects of race and distinctiveness" *Quarterly Journal of Experimental Psychology Human Experimental Psychology* **44a** 671-703