



**HAL**  
open science

## Is distinctiveness gender-based ?

Jean-Yves Baudouin, Mathieu Gallay

► **To cite this version:**

Jean-Yves Baudouin, Mathieu Gallay. Is distinctiveness gender-based ?. *Journal of Experimental Psychology: Human Perception and Performance*, 2006, 32 (4), pp.789-798. 10.1037/0096-1523.32.4.789 . hal-00560077

**HAL Id: hal-00560077**

**<https://u-bourgogne.hal.science/hal-00560077>**

Submitted on 14 Apr 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Is Face Distinctiveness Gender Based?

Jean-Yves Baudouin & Mathieu Gallay

*Two experiments were carried out to study the role of gender category in evaluations of face distinctiveness. In Experiment 1, participants had to evaluate the distinctiveness and the femininity–masculinity of real or artificial composite faces. The composite faces were created by blending either faces of the same gender (sexed composite faces, approximating the sexed prototypes) or faces of both genders (nonsexed composite faces, approximating the face prototype). The results show that the distinctiveness ratings decreased as the number of blended faces increased. Distinctiveness and gender ratings did not covary for real faces or sexed composite faces, but they did vary for nonsexed composite faces. In Experiment 2, participants were asked to state which of two composite faces, one sexed and one nonsexed, was more distinctive. Sexed composite faces were selected less often. The results are interpreted as indicating that distinctiveness is based on sexed prototypes. Implications for face recognition models are discussed.*

**Keywords:** face recognition, gender, distinctiveness, face space

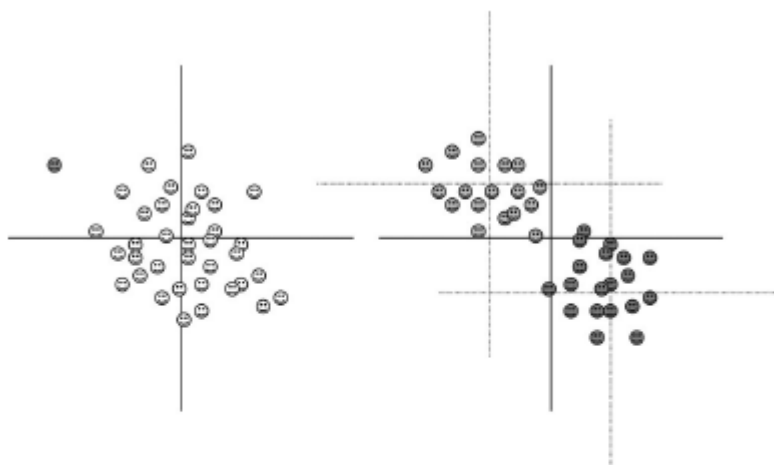
In the 1980s, various attempts to understand face recognition processing in humans showed that faces are encoded and retrieved in reference to a prototypical face (e.g., Goldstein & Chance, 1980; Rhodes, Brennan, & Carey, 1987; Valentine & Bruce, 1986a, 1986b). Such a prototype—sometimes called the facial schema or norm—is assumed to be developed through multiple exposures to facial stimuli and to capture the average properties of the face category. Individual faces are thought to be encoded in terms of their deviations from the prototype. This hypothesis has been used successfully to account for many well-documented phenomena in face recognition. For example, Rhodes et al. (1987) accounted for the caricature effect in terms of increased distance from a norm; the caricature of a face is recognized more easily than the real face because caricatures accentuate the facial characteristics that differentiate the face from the norm, that is, the features used to recognize the face. Goldstein and Chance (1980) explained the race and inversion effects by studying the facial schema on which face recognition is based. In the same way, Light, Kayra-Stuart, and Hollander (1979) and Valentine and Bruce (1986a, 1986b) proposed that distinctiveness results from the distance from a facial prototype; the farther a face is from the prototype, the more distinctive it is. This idea allowed researchers to explain many of the effects reported for typical versus distinctive faces. Notably, distinctive faces are recognized more accurately and faster than typical faces (e.g., Cohen & Carr, 1975; Going & Read, 1974; Light et al., 1979; Shepherd, Gibling, & Ellis, 1991; Valentine & Bruce, 1986a; Winograd, 1981). Typical faces tend more often to be falsely recognized (Light et al., 1979). Typical faces are also classified faster as a face (by comparison to scrambled faces) than distinctive ones (Valentine & Bruce, 1986a).

### **Valentine's (1991) Face-Space Model**

In the early 1990s, Valentine (1991) bridged the gap between these separate models by proposing to account for various effects in the framework of a multidimensional face space. He considered the representation of a face to be a point in a multidimensional Euclidean space. The dimensions of this space are the physiognomic features or facial properties used to encode faces. The origin of the space is defined as the central tendency of the dimensions, with the assumption that the values on the dimension in the population of faces are normally distributed around the central

tendency. Thus, the density of faces is higher at the central tendency and decreases as the distance from the central tendency increases. Consequently, typical faces (close to the central tendency) are more common than distinctive ones (Valentine, 1991). An illustration of this face space with two dimensions is represented on the left side of Figure 1.

Two theoretical approaches were proposed to model the coding and recognition of a face in the face-space framework. First, faces are encoded in terms of their deviation from a norm or prototype. This prototype is located at the origin of the space, and it is assumed that there is a single prototype for all faces (Valentine, 1991; Valentine & Bruce, 1986a). Valentine (1991) referred to this approach as a norm-based model. In such an approach, a face is encoded as an  $n$ -dimensional vector that originates at the origin. The discrimination of faces relies on a vector similarity measure, with the similarity of faces being a function of vector similarity. The second approach—referred to by Valentine (1991) as an exemplar-based model—is to consider that there is no extracted prototype and that the similarity of faces is a monotonic function of the distance between the faces in the space. The decision process depends on the distance between a face and its nearest neighbors. Norm-based and exemplar-based models can both account for distinctiveness effects and also the effects of caricature, race, and inversion (see Valentine, 1991). For the distinctiveness effect, Valentine (1991) suggested that because exemplar density is higher at the origin (i.e., in the area of the space where typical faces are), the discrimination of typical faces is more difficult than that of distinctive faces, with the probability of a vector or a neighbor falling close to the face being higher.



*Figure 1.* Left: The face-space model, which is based on Valentine's (1991) description. The face density is the greatest at the origin of the space. A distinctive face is located far away from the origin in a low density area (e.g., the gray face). Right: The face-space model with gender categories taken into account. There are two high density areas, one for each gender category. The origin of the space becomes a lower density area.

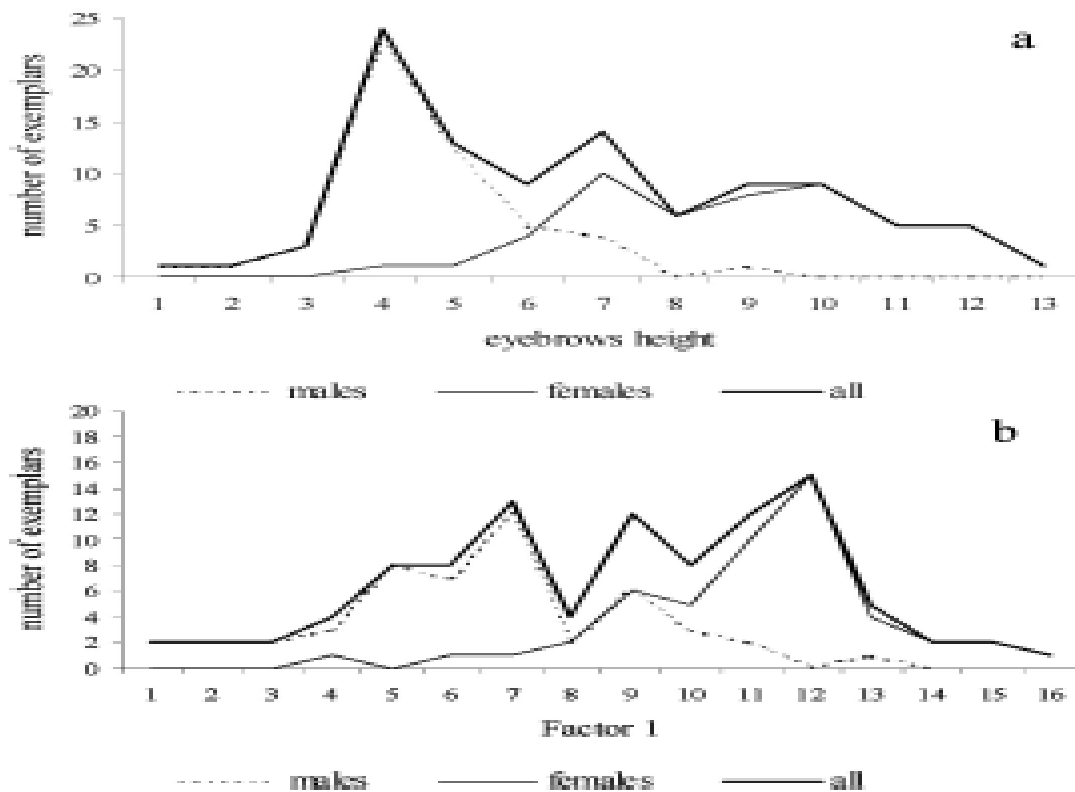
Since its proposal by Valentine in 1991, the face-space model has given rise to a number of investigations. One of the goals of these investigations was to verify that the distinctiveness rating really increases with the distance between an individual face's characteristics and the average value of those characteristics in the face population (e.g., Bruce, Burton, & Dench, 1994; Johnston, Milne, Williams, & Hosie, 1997). Another goal was to compare norm-based and exemplar-

based models, with an advantage for this last conception (e.g., Valentine & Endo, 1992; but see Levin, 1996). Attempts were also made to define the dimensions of the face space (for a review, see Valentine, 2001). Few studies have looked again at the earlier assumption that there is a single central tendency of higher density.

### **The Hypothesis of a Single Central Tendency: Consideration of Gender**

A key assumption in the initial studies based on the face-space model was that the origin of the space corresponds to the central tendency of all dimensions, where the exemplar density is the highest. This property explains the distinctiveness effect for both norm-based and exemplar-based models, even if the rationale is different (i.e., vector similarity vs. distance between exemplars). The assumption of higher density at the origin presupposes that the values of the dimension in the population of faces are normally distributed around the central tendency for *each* dimension. If any dimension in the multidimensional space is bimodal, the points plotted in the space will not be grouped around the origin but will be split in two distinct clouds. This point was already underlined for ethnicity (see Chiroro & Valentine, 1995; Valentine & Endo, 1992). Now, there are reasons to assume that some other dimensions than those coding for ethnicity are bimodal. It is the case for example of the dimensions that differentiate the gender categories. Many differences between female and male faces have been reported in the literature: Female faces have thinner eyebrows, their eyebrows are higher above the eyes, their chin is smaller, and so on (Brown & Perrett, 1993; Campbell, Benson, Wallace, Doesbergh, & Coleman, 1999; Yamaguchi, Hirukawa, & Kanazawa, 1995). One can assume that if one of the dimensions of the face space represents the value taken on by such physical characteristics, the distribution of the face population will be bimodal, with one mode for each gender. To illustrate this assumption, we measured four facial characteristics thought to differentiate female and male faces (height of eyebrows, nose size, chin length, and eyebrow thickness) on 50 female and 50 male faces using the method designed by Baudouin and Tiberghien (2004) for measuring facial features. The distribution of faces on the height of eyebrows dimension is illustrated in Figure 2a. As Figure 2a shows, the mode of each gender category is different. The result when all faces are considered together is that the face population is not normally distributed around a central tendency but quite bimodal with two “central” tendencies, one for each gender. The existence of a single dimension representing gender (Johnston, Kanazawa, Kato, & Oda, 1997) would give the same result. Figure 2b presents the factorial results for the first factor of a principal-components analysis performed on the four facial measurements. These results can be considered to code the gender category of the faces (i.e., it corresponds to a composite of the four measurements), and as Figure 2b shows, its distribution is not normal but multimodal. The insertion of this dimension in a multidimensional face space would thus result in various pools of high density areas (see right part of Figure 1 for an illustration).<sup>1</sup> The partition of the face space into two high density areas by the gender categories highlighted many problems that were not considered in the current literature (see also Benson, 1995). This question was generally tackled in studies on the properties of the face space by looking at faces of only one gender. It was nevertheless often suggested (generally implicitly) that the distinctiveness of a face is assessed by comparison to a sexed prototype. In the study by Bruce et al. (1994), for example, the authors defined the distinctive characteristics of a set of faces by comparing subjective ratings of facial distinctiveness with objective facial measurements. They reported a correlation between distinctiveness ratings and the physical deviation of faces from the average computed from faces—of the same gender. An exception is the study by Johnston, Kanazawa, et al. (1997), who suggested that both gender and age are dimensions of the face space.

However, they were interested in the role of this partition of the face space in a classification task, not in distinctiveness ratings or face recognition.



**Figure 2.** Illustration of the distribution of 50 female and 50 male faces on the eyebrow-height dimension (a) or on a gender dimension (Factor 1) derived from four facial measurements (b). The central tendency for each

It is important to consider gender categories because they partition the plot of faces in a similar way as do ethnic categories, with the main exceptions that (a) gender categories are probably closer to each other in the space than ethnic categories and (b) the different categories have probably been seen to more equal extents for gender than for ethnic origin. Many problems have thus been raised regarding the partition of faces in the face space. One of the problems concerns norm-based models: Should it be assumed that there is a single prototype for all faces (Valentine, 1991; Valentine & Bruce, 1986a, 1986b) or two prototypes, one for each gender? The assumption of a single prototype, derived from both female and male faces and lying at the origin of the space, implies that most faces are not located around the origin but lie away from it in two main directions on the dimensions that differentiate the gender categories. The prototype, which would have few gender markers, is even atypical if one considers that most faces have gender markers. A consequence of this is that deviations from the prototype (assumed to increase distinctiveness) put the face in a region of high density when the deviations concern features that differentiate the genders. In other words, some deviations from the prototype are more frequent in the population of faces than others, resulting in some distinctive faces that are seen frequently. The assumption of two prototypes raises some other questions, notably the problem of selecting the appropriate prototype. A typical female face can be considered as an atypical male face, so it may be important to compare a face with the prototype of its own gender category. But, do people select the appropriate prototype before performing encoding

and recognition processes? Or, is an encoded face compared with both prototypes at the same time? The hypothesis of a dual comparison raises the question of how the system manages the two sources of information. The existence of two high density areas is less problematic for exemplar-based models. Given that the decision process depends on the distance between a face and its nearest neighbors, the main consequence of having two high density areas is that the probability of encountering a face of the same gender in the neighborhood is higher than the probability of encountering a face of the opposite gender. A distinctive face will then be a face that deviates from the central tendency of one pool of faces without lying in the pool of faces of the other gender. The existence of two gender-defined areas nevertheless raises some questions. In particular, is gender category taken into account in decision processes—that is, is the face compared with neighbors even when they belong to the opposite gender category? Baudouin and Tiberghien (2002) showed that the assignment of a specific gender to ambiguous faces affected the distractor-comparison process: When a target face was assigned a gender opposite to the gender of the distractors, the latter were rejected faster than when the face was assigned the same gender. Thus, despite strictly equivalent distances for physical properties, the pool of exemplars (or the prototype?) used to encode and/or retrieve a face differed according to the gender attributed to the encoded face. The question raised by this finding concerns whether distinctiveness is assessed by comparison with the population of faces or with the population of same-gender faces.

Thus, there are reasons to question the role of gender category in facial distinctiveness. Notably, a major question is whether distinctiveness is evaluated by comparing an encoded face with a prototype (or a set of exemplars) that shares the same categorical properties (i.e., a prototype of the category) or, on the contrary, with a more global prototype that captures general average properties of the population of faces and thus discards properties that divide the population into distinct categories. The purpose of the present study was to test whether the distinctiveness of a face is assessed by relying on the sexed prototype (or pool of same-gender faces) of the face's gender category or by relying on a face prototype derived from both female and male faces. In Experiment 1, participants were asked to rate the distinctiveness and the gender of faces that were either real female or male faces or composite faces made with increasing numbers of faces (2, 4, 8, 16, 32, and 64). Increasing the number of faces in the composite face was supposed to put the composite face closer to the average or prototypical properties of the population of faces (Langlois & Roggman, 1990). Consequently, the distinctiveness of the composite faces should decrease as the number of faces in them increases. Two types of composite faces were used: (a) sexed composite faces, in which only faces of the same gender were mixed in, and (b) nonsexed composite faces, in which an equal number of female and male faces were used. The first kind of composite face was assumed to approximate sexed prototypes, either feminine or masculine. The second kind was assumed to approximate the (nonsexed) face prototype. If distinctiveness is determined by comparison with a sexed prototype, the decrease in distinctiveness with the increase in the number of faces should depend on or be greater for composite faces that approximate a sexed prototype. In Experiment 2, participants were shown pairs of composite faces with one sexed and one nonsexed composite face in each pair, and they were instructed to indicate which of the two faces was more distinctive. If distinctiveness is assessed on the basis of a sexed prototype, composite faces approximating this kind of prototype (i.e., sexed composite faces) should be selected less often. Even if we use the term *prototype*, these experiments were not designed to decide between norm-based or exemplar-based models but to test properties that have implications for both kinds of approaches.

## **Experiment 1**

Average composite faces were created by blending different individual faces. The number of faces in each average face was increased from 1 to 64, with five intermediate levels (2, 4, 8, 16, and 32 faces). Two types of average faces were created based on the gender of the faces of which they were composed: (a) Some of the average faces were sexed (i.e., only faces of the same gender were used), and (b) others were nonsexed (i.e., an equal number of faces of each gender were used). Participants were asked first to rate the distinctiveness of the faces and then to categorize and rate them on gender.

### ***Method***

#### *Participants*

Twenty participants (10 women and 10 men) took part in the experiment. They were between 19 and 38 years old ( $M = 22.1$ ).

#### *Materials*

The faces of 32 women and 32 men were used. No a priori selection was done: The faces were drawn from a person database (created by Jean-Yves Baudouin) according to their real sex, and the first 32 female and male faces were selected. The faces were photographed with a Kodak digital camera. The models were face front, with a neutral expression. Each face was put into an oval so that the hairstyle was not visible. The oval was 670 pixels high  $\times$  510 pixels wide, which corresponded to a size of about 14.5  $\times$  11 cm on the screen. Average faces were created using a morphing technique and Morpheus (Version 1.85; Morpheus Software, 1999–2003). With this technique, 256 points were put on the main features of a face. The points were then placed in the same locations on another face. The software computed an intermediate level between the locations of the points and created a new face for which points were at an equal distance from the two original points. It also computed an intermediate texture for surfaces delineated by points. The composite face thus had features for which position, size, and texture were at an equal distance from position, size, and texture of the features of the two original photographs. The resulting composite face can therefore be considered as the average of the two original faces. The 64 faces were divided into 16 sets of 4 faces each (2 female and 2 male). For each set, we generated 4 average faces by blending 2 faces for each. The first 2 average faces were made from the faces of the same gender (i.e., the 2 female faces and the 2 male faces in the set) and corresponded to average sexed faces (female and male, respectively). The other 2 average faces were created with faces of the opposite gender and corresponded to average nonsexed faces (Androgynous Face 1 and Androgynous Face 2). This gave us four groups of 16 averaged faces: In two groups, the faces were sexed (16 female and 16 male), and in the other two, they were nonsexed (Androgynous Face 1 and Androgynous Face 2). In each group, the faces were then averaged two by two, resulting in 8 new averaged faces per group. These 8 averaged faces thus represented the average position, size, and texture of 4 faces each. They were mixed two by two again to obtain 4 average faces that were a blend of 8 original faces, and then 2 average faces that blended 16 original faces. Finally, the two 16-face averages were blended to obtain an average of 32 faces. The 32-face averages obtained from faces of the same gender were then blended to obtain an average face of the 64 original faces. The same procedure was performed with the two 32-face averages obtained from faces of different genders. An illustration of the composite faces is

presented in Figure 3. With this procedure, we had 64 original faces (32 female and 32 male), 64 average faces from 2 original faces (16 female, 16 male, and 2 \_ 16 nonsexed), 32 average faces from 4 original faces (8 female, 8 male, and 2 \_ 8 nonsexed), 16 average faces from 8 original faces (4 female, 4 male, and 2 \_ 4 nonsexed), 8 average faces from 16 original faces (2 female, 2 male, and 2 \_ 2 nonsexed), 4 average faces from 32 original faces (1 female, 1 male, and 2 \_ 1 nonsexed), and 2 average faces from 64 original faces. The 32-face average faces created with the two groups of sexed faces were used as the sexed prototypes. The 2 average faces of 64 (created, respectively, with the 2 sexed 32-face average faces and the 2 nonsexed 32-face average faces) were used as facial prototypes.

### *Procedure*

Participants sat in front of a computer at a distance of about 80 cm. The participants performed two tasks in two distinct sessions. *Distinctiveness rating.* After a fixation point, a face was presented in the center of the screen. Participants were told to evaluate its distinctiveness on a scale ranging from 0 (*not distinctive at all*) to 9 (*very distinctive*) using the numeric pad of the keyboard. A distinctive face was said to be an atypical, unusual face that should be easy to recognize later. Participants were also told that none of the faces in the experiment were real and that they had been made using computer software. So, they had to disregard any lack of realism in the faces by imagining that they were of actual persons. Each participant performed 190 trials (i.e., all faces), presented in random order. The faces remained on the screen until the participant responded. There was no time limit.

### ***Gender categorization***

After the distinctiveness rating task, the 190 faces were presented again to participants. They were instructed to state for each face whether the person was female or male using two keys on the keyboard. After this response, they had to use the numeric pad to rate the femininity or masculinity of the face (depending on their previous response) on a 10-point scale ranging from 0 (*very unfeminine or unmasculine*) to 9 (*very feminine or masculine*).

### ***Results***

*Distinctiveness Ratings* Two analyses were performed. In the first analysis, all faces were included (real faces and 64-face composites). The type of composite face (sexed or nonsexed) was not considered. The purpose of the analysis was to test for the hypothesized decrease in distinctiveness. In the second analysis, the type of composite face was taken into account, but real faces as well as 64-face composites were not considered because they had only one type of face (sexed and nonsexed, respectively). Means and standard deviations are presented in Table 1. *All faces.* We performed a one-factor analysis of variance (ANOVA; number of faces: 1 vs. 2 vs. 4 vs. 8 vs. 16 vs. 32 vs. 64, within subjects) to determine whether distinctiveness really decreased as the number of faces in the composite face increased. The effect of number of faces was significant,  $F(6, 114) = 40.60, p = .01$ . Table 1 indicates that the distinctiveness ratings decreased as the number of faces increased, rapidly for the first step (2 faces) and slower for the other steps. The difference was significant between 1 and 2 faces (6.0 vs. 4.3),  $F(1, 19) = 66.55, p = .01$ ; 2 and 4 faces (4.3 vs. 3.3),  $F(1, 19) = 43.19, p = .01$ ; 4 and 8 faces (3.3 vs. 2.9),  $F(1, 19) = 15.96, p = .01$ ; and 8 and 16 faces (2.9 vs. 2.7),  $F(1, 19) = 5.18, p = .05$ . The difference was not significant between 16 and 32 faces,  $F(1, 19) = 0.72$ , or between 32 and 64 faces,  $F(1, 19) = 3.24$ .



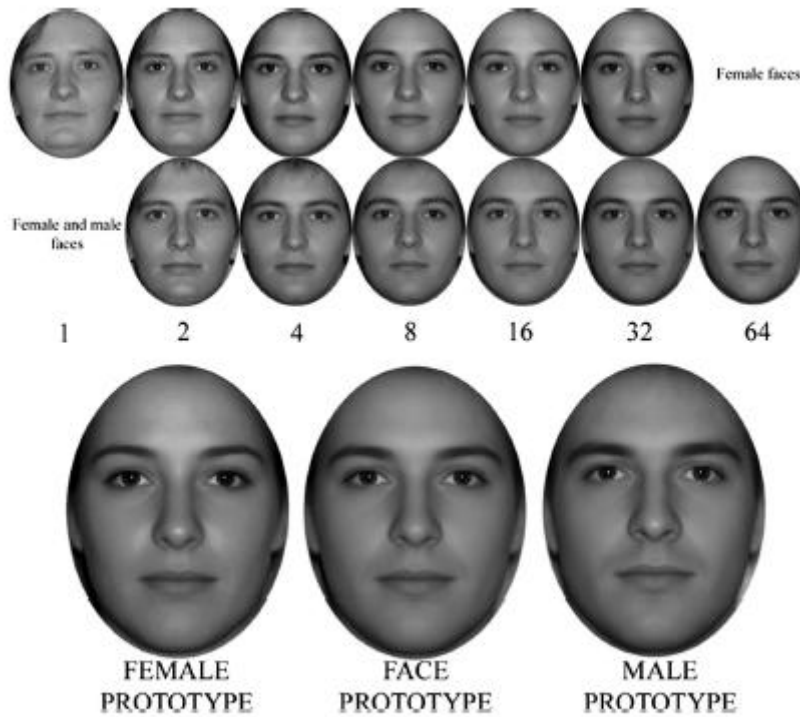


Figure 3. Illustration of composite faces. Top: A female face is blended with an increasing number of faces, either of the same gender or of both genders, to create a female or androgynous prototypical face. Bottom: The two sexed prototypes (32 faces) and the (androgynous) face prototype (64 faces).

Table 1  
Mean Distinctiveness Ratings (0 to 9) and Standard Deviations  
by Number of Blended Faces

Type of face	Number of faces						
	1	2	4	8	16	32	64
Sexed							
<i>M</i>	6.0	4.3	3.3	2.9	2.5	2.7	
<i>SD</i>	0.8	1.1	1.6	1.6	1.7	2.0	
Nonsexed							
<i>M</i>		4.4	3.4	2.8	2.8	3.0	2.5
<i>SD</i>		1.0	1.4	1.7	1.7	2.3	2.0
Overall							
<i>M</i>	6.0	4.3	3.3	2.9	2.7	2.9	2.5
<i>SD</i>	0.8	1.0	1.5	1.6	1.7	2.0	2.0

*Same versus different gender.* Preliminary analyses indicated no differential effect of gender for sexed composite faces (female vs. male) or of the group of nonsexed composite faces (Androgynous Face 1 or Androgynous Face 2). The data for each type of face were thus averaged within sexed and nonsexed composite faces. We performed a two-factor ANOVA (number of faces: 2 vs. 4 vs. 8 vs. 16 vs. 32; type of faces: sexed vs. nonsexed—both within subjects) to determine whether distinctiveness decreased in a different way for the two in the composite faces. The effect of number of faces was significant,  $F(4, 76) = 19.55$ ,  $p < .01$ , replicating the effect reported in the previous analysis. So we

will not present this analysis. The type-of-faces factor was nonsignificant,  $F(1, 19) = 1.24$ , and did not interact with the number of faces,  $F(4, 76) = 0.70$ .

### Gender Categorization

Two gender indicators were considered: percentage of “female” responses and femininity–masculinity ratings. Femininity–masculinity ratings following a “male” response were assigned a negative value, and those following a “female” response were assigned a positive value. For each indicator, we performed a  $2 \times 2 \times 5$  ANOVA with type of face (sexed vs. nonsexed), exemplar (Exemplar 1 vs. Exemplar 2; i.e., for sexed faces, female vs. male, and for nonsexed faces, Androgynous Face 1 vs. Androgynous Face 2), and number of faces (2 vs. 4 vs. 8 vs. 16 vs. 32) as within-subject factors.

*Percentage of “female” responses.* Figure 4a illustrates the results. The main effect of type of face was nonsignificant,  $F(1, 19) = 4.12$ . The main effect of exemplar was significant,  $F(1, 19) = 525.50$ ,  $p < .01$ , but it was qualified by a significant interaction between exemplar and type of face,  $F(1, 19) = 693.65$ ,  $p < .01$ , indicating that the exemplar factor had a significant effect for sexed composite faces (95.2% “female” responses for female composite faces vs. 2.4% for male composite faces),  $F(1, 19) = 1,658.73$ ,  $p < .01$ , but not for nonsexed composite faces,  $F(1, 19) = 2.00$ . The main effect of the number of faces was also significant,  $F(4, 76) = 3.81$ ,  $p < .01$ , but it was qualified by a significant interaction with the type of face,  $F(4, 76) = 3.35$ ,  $p < .05$ . This interaction indicated a significant effect of number of faces for nonsexed composite faces,  $F(4, 76) = 4.07$ ,  $p < .01$ , but not for sexed ones,  $F(4, 76) = 0.85$ . Figure 2a indicates that participants had a tendency to respond “male” for nonsexed composite faces when the number of faces was low, but this tendency disappeared when the number of blended faces was increased. The overall interaction was not significant,  $F(4, 76) = 1.36$ .

*Femininity–masculinity ratings.* Figure 4b illustrates the results. The main effect of the type-of-faces factor was nonsignificant,  $F(1, 19) = 3.22$ . The main effect of the exemplar factor was significant,  $F(1, 19) = 238.70$ ,  $p < .01$ , but it was qualified by a significant interaction between exemplar and type of face,  $F(1, 19) = 443.78$ ,  $p < .01$ . This interaction indicated that the exemplar had a significant effect for sexed composite faces (5.6 for female composite faces vs. –5.9 for male composite faces),  $F(1, 19) = 475.12$ ,  $p < .01$ , but not for nonsexed ones,  $F(1, 19) = 0.67$ . The main effect of the number of faces was also significant,  $F(4, 76) = 10.47$ ,  $p < .01$ , but it was qualified by both a Number of Faces  $\times$  Type of Face interaction,  $F(4, 76) = 4.44$ ,  $p < .01$ , and a Number of Faces  $\times$  Type of Face  $\times$  Exemplar interaction,  $F(4, 76) = 2.86$ ,  $p < .05$ . In this last interaction, the number of faces had a significant effect for male composite faces,  $F(4, 76) = 4.65$ ,  $p < .01$ , and for each kind of nonsexed composite face,  $F(4, 76) = 7.22$ ,  $p < .01$ , and  $F(4, 76) = 7.29$ ,  $p < .01$ , respectively, but not for female ones,  $F(4, 76) = 1.44$ . Figure 2b shows that there was a tendency to rate nonsexed composite faces as masculine when few faces were blended. This tendency decreased as the number of faces increased. For male faces, increasing the number of male faces decreased perceived masculinity.

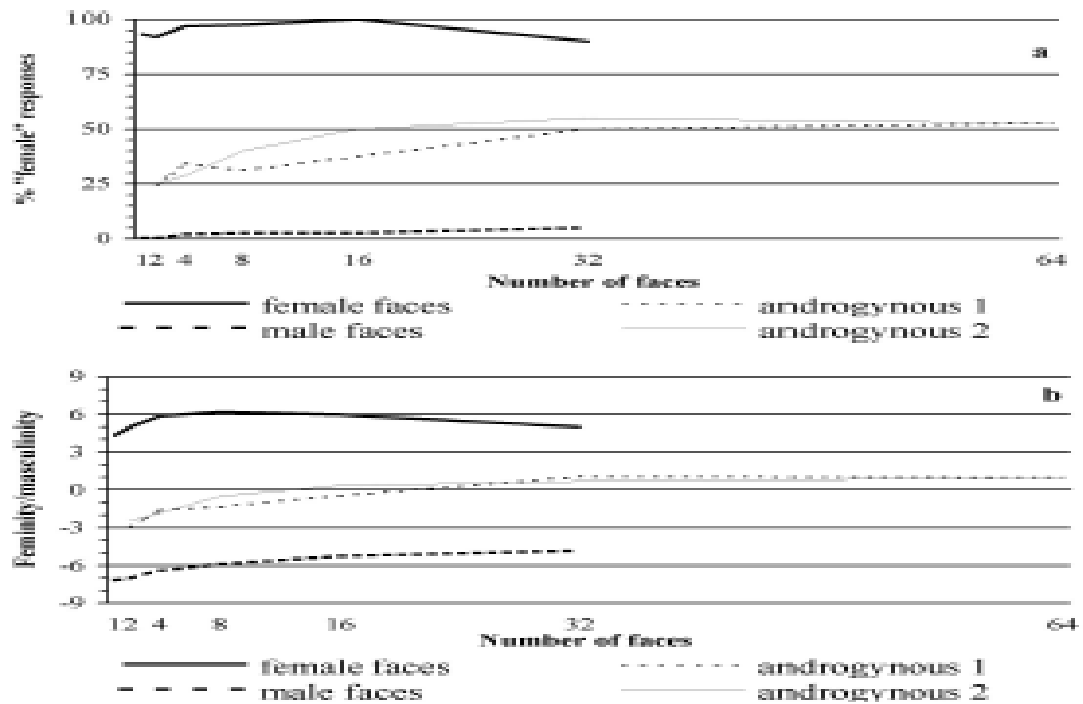


Figure 4. Performance on the gender categorization task: Percentage of "female" responses (a) and femininity–masculinity ratings (b).

#### Correlations Between Distinctiveness and GenderCategorization

Two gender categorization indicators were used. The first was an indicator of gender based on "female" responses. The percentages for each face were transformed so that percentages lower than 50% were replaced by the difference between that percentage and 100% (i.e., a face that was labeled as female in 0% of responses was assigned a percentage of 100%). This transformation gave us an indicator of the participants' agreement about gender: A strongly sexed face—either as female or male—had a score close to 100%. On the contrary, an ambiguous face had a score close to 50%. The other indicator was the femininity–masculinity ratings. For each face, we looked at the absolute value of its mean rating. A highly sexed face thus had a high value whereas an ambiguous or androgynous face had a rating close to zero. To test for a potential effect of gender on distinctiveness, Bravais–Pearson correlations were computed between the distinctiveness rating and these two gender categorization indicators. The results are summarized in Table 2. As indicated in Table 2, there was a significant correlation between the distinctiveness ratings and the two gender categorization indicators when all faces were considered together. Faces tended to be more distinctive when gender attributes were more prominent, which led to greater agreement among subjects and higher femininity–masculinity ratings. Nevertheless, when correlations were computed for real faces, sexed composite faces, and nonsexed composite faces, the link between the distinctiveness rating and the gender indicators was no longer significant for either real faces or sexed composites. It was always significant when nonsexed composite faces were considered. Before considering these observations, we examine the distribution of faces in a 2-D space, with distinctiveness and the gender indicators as dimensions. Figure 5 illustrates the distribution of the faces used in the experiment in a 2-D space with distinctiveness and the percentage of "female" responses (see Figure 5a) or the femininity–masculinity rating (see Figure 5b) as dimensions.

Table 2  
*Correlations Between Distinctiveness Ratings, Percentage of “Female” Responses, and Femininity–Masculinity Ratings (Bravais–Pearson  $r$ )*

Type of face	Coefficient of correlation
All faces ( $N = 190$ )	
Distinctiveness vs. participants’ agreement	.35**
Distinctiveness vs. femininity–masculinity ratings	.32**
Participants’ agreement vs. femininity–masculinity ratings	.91**
Sexed composite faces ( $N = 62$ )	
Distinctiveness vs. participants’ agreement	–.05
Distinctiveness vs. femininity–masculinity ratings	.14
Participants’ agreement vs. femininity–masculinity ratings	.82**
Nonsexed composite faces ( $N = 64$ )	
Distinctiveness vs. participants’ agreement	.34*
Distinctiveness vs. femininity–masculinity ratings	.34*
Participants’ agreement vs. femininity–masculinity ratings	.91**
Real faces ( $N = 64$ )	
Distinctiveness vs. participants’ agreement	.01
Distinctiveness vs. femininity/masculinity ratings	–.02
Participants’ agreement vs. femininity/masculinity ratings	.74**

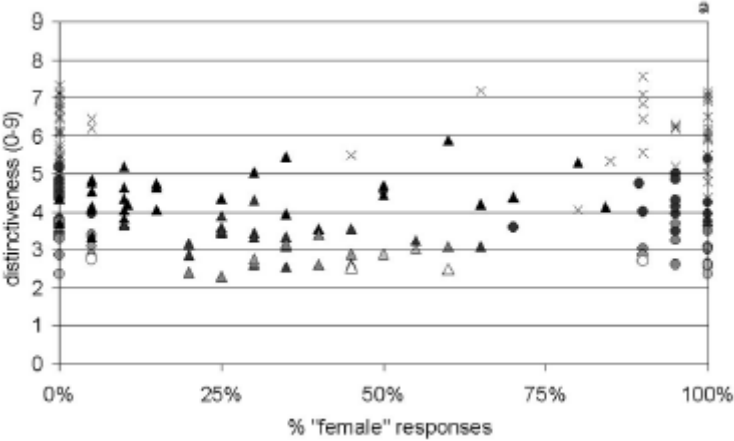
Distinctiveness, the percentage of “female” responses, and the femininity–masculinity ratings were averaged across participants for each face. If we consider only real faces (crosses in Figure 5), two pools of faces emerged along the gender dimensions, one for each gender category. Very few real faces lie between these two pools, and the distinctiveness ratings of these faces were quite high (between 4.1 and 7.6). Increasing the number of same-gender faces (circles in gradations of gray to white) made the distinctiveness ratings decrease, but the percentage of “female” responses and the femininity–masculinity ratings remained relatively constant. By adding more and more faces of the opposite gender (gray to white triangles), distinctiveness, the participants’ agreement about gender, and the femininity–masculinity ratings decreased.

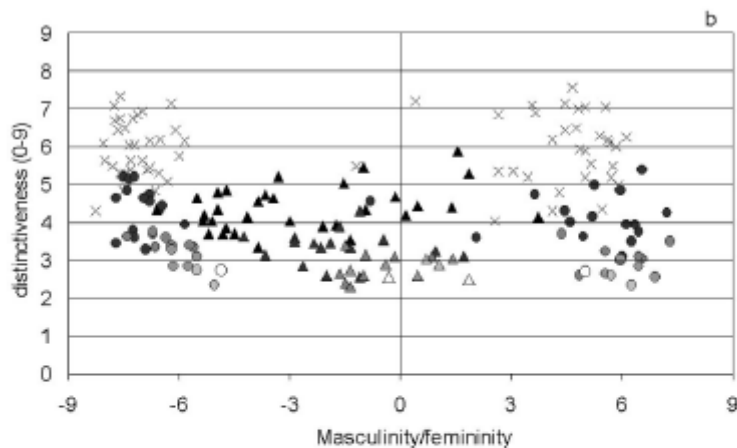
Thus, Figure 5 suggests that there are two ways in which distinctiveness can decrease: (a) The face becomes closer to the sexed prototype, or (b) the face becomes closer to the androgynous face prototype. In the first case, the decrease in distinctiveness was independent of the gender indicators. In the second case, it was related to them. The less distinctive a face, the less it will exhibit gender markers. These assumptions are in line with the coefficients of correlation reported in Table 2: The correlation between distinctiveness and the gender indicators was nonsignificant for sexed composite faces, but it was significant for nonsexed composite ones. This confirms the assumption that the relationship between distinctiveness and gender is dependent on whether the faces were closer to a sexed or nonsexed prototype. It is thus possible to find out whether the distinctiveness of real faces varies according to their closeness to a sexed or nonsexed prototype. One can hypothesize in the first case that distinctiveness and gender characteristics are unrelated, and in the second case that they are significantly related. Figure 5 suggests that there was in fact no relationship between them, which was confirmed by the nonsignificant Bravais–Pearson coefficient of correlation for real faces (see

Table 2). Thus, the distinctiveness of real faces is independent of their gender characteristics, suggesting that it is determined by comparison to a sexed prototype.

**Conclusions**

The results of Experiment 1 show that distinctiveness gradually decreased with the number of averaged faces. A floor level was reached at 16 faces, after which increasing the number of faces in the composite did not give rise to an additional decrease in the distinctiveness ratings. Moreover, the gender of the faces that were blended had no effect on the mean distinctiveness ratings. Correlation analyses nevertheless indicated that the distinctiveness ratings of the real faces, unlike the nonsexed composite faces but like the sexed ones, did not covary with gender saliency as indicated by the femininity–masculinity ratings and the participants’ agreement about gender category. This observation suggests that the distinctiveness of real faces is assessed by comparison to a sexed prototype.





*Figure 5.* Scatterplots of distinctiveness ratings versus percentage of “female” responses (a) and distinctiveness ratings and femininity–masculinity ratings (b). Real faces are represented by crosses, same-gender composites with increasing number of faces are represented by circles in gradations of gray to white, and different-genders composites with increasing number of faces are represented by triangles in gradations of gray to white.

One limitation of the results of Experiment 1 is that despite some indication of an influence of gender category on distinctiveness rating, no difference in the mean distinctiveness ratings was observed between sexed and nonsexed composite faces. The means in Table 1 suggest that sexed composite faces with 16 or 32 faces were in fact less distinctive than nonsexed ones, but the difference was not statistically significant. A possible explanation is that the measure used in this experiment—a rating scale—was not sensitive enough to allow a potential effect to show up. Real faces were presented together with composite ones in such a way that the variability in their distinctiveness was relatively high. This may have reduced fine discriminations between faces of relatively similar, low distinctiveness, like the sexed and nonsexed composite faces. To further test for a potential difference in distinctiveness for such composite faces with a more sensitive measure, we used a two-choice task. Participants were required to say which of two composite faces—a sexed and a nonsexed one—was the most distinctive. If distinctiveness is based on a nonsexed face prototype, sexed faces should be selected more often as more distinctive. If it is determined by comparison to a sexed prototype, then nonsexed prototypical faces should be selected more often.

## Experiment 2

### **Method**

#### *Participants*

Sixteen new participants (9 women and 7 men) ranging in age from 21 to 28 ( $M = 23.25$ ) participated in the study.

#### *Materials*

We used the 16 average faces that were blends of 8 real faces (4 female, 4 male, 2 \_ 4 nonsexed), the 8 average faces that were blends of 16 original faces (2 female, 2 male, 2 \_ 2 nonsexed), and the 4

average faces that were blends of 32 original faces (1 female, 1 male, 2 \_ 1 nonsexed) from Experiment 1.

### *Procedure*

After a fixation point, a composite face was displayed on the left side of the screen along with another composite face on the right side. The participants were asked to press the left or right arrow on the keyboard, depending on which face they found the most distinctive. They were told that a distinctive face was an atypical, unusual face that would be easy to recognize later. They were also told that none of the faces in the experiment were real and that they had been generated using computer software. So they had to disregard any lack of realism in the faces by imagining that they corresponded to real persons. Every sexed composite face was presented with a nonsexed composite face that blended the same number of faces (i.e., sexed composite faces made from 8 faces were displayed with nonsexed composite faces also made from 8 faces). Each participant saw 84 pairs (4 for 32-face composites, 16 for 16-face composites, and 64 for 8-face composites), presented in random order. The pairs of faces remained on the screen until the participant responded. For each participant, sexed and nonsexed composite faces were displayed equally often on the left and right sides of the screen, with the right-left location of the faces being alternated across participants.

### **Results**

For each composite face, we computed how often (in percentage) it was selected as being more distinctive than the other face in the pair. A two-factor ANOVA (number of faces: 8 vs. 16 vs. 32; type of face: sexed vs. nonsexed) by items (in which case the two factors were between-item factors) and by subjects (the two factors were within-subject factors) was conducted to find out whether distinctiveness is determined from a sexed or a nonsexed prototype. Means and standard deviations are presented in Table 3. The main effect of type of face was significant: Nonsexed faces were selected as more distinctive more often than were sexed faces (62.8% vs. 37.2%): by items,  $F(1, 22) = 104.99, p < .01$ ; by subjects,  $F(1, 15) = 6.00, p < .05$ . Neither the main effect of the number of faces nor the interaction between the number of faces and the type of face was significant ( $ps > .05$ ). To further test for the role of gender in distinctiveness evaluations, we computed Bravais-Pearson coefficients of correlation to determine whether the selection rate of each face (e.g., "this face is more distinctive") was correlated with the participants' agreement rate about gender category and with the absolute value of the mean femininity-masculinity ratings obtained for these same faces in Experiment 1. There was a strong and significant negative correlation between the selection rate and both the participants' agreement rate ( $N = 28, r = -.82, p < .01$ ) and the femininity-masculinity ratings ( $N = 28, r = -.89, p < .01$ ). Thus, the more obvious the gender, the less often the face was judged to be more distinctive.

### *Conclusions*

Experiment 2 showed that composite faces made up of same gender faces were less distinctive than composite faces made up of faces of both genders. This finding argues in favor of the idea that distinctiveness is assessed by means of a comparison to a sexed prototype rather than a nonsexed prototype. Correlation analyses further strengthened this conclusion. The composite faces were more often selected as the less distinctive face in the pair when their gender markers were more prominent. In this experiment, the gender markers corresponded to features that were averaged

across at least eight faces. So there were probably few atypical sexed features but more probably sexed features that corresponded to prototypical characteristics of each gender category. Thus, faces whose gender markers were more prominent probably tended to be closer to one of the two sexed prototypes. Consequently, the negative covariation between distinctiveness ratings and the gender indicators further indicated that the faces were even less distinctive when they were close to a sexed prototype.

**Table 3**  
*Selection Rate and Standard Deviations of Sexed Versus Nonsexed Composite Faces, by Number of Faces*

Type of face	Number of faces		
	8	16	32
<b>Sexed composite faces</b>			
Mean selection rate (%)	39.5	39.5	32.8
SD	16.9	23.0	27.0
<b>Nonsexed composite faces</b>			
Mean selection rate (%)	60.5	60.5	67.2
SD	16.9	23.0	27.0

## General Discussion

The results of Experiments 1 and 2 show that gender is a determinant of face distinctiveness. In Experiment 1, the covariation of the distinctiveness ratings and the gender indicators of real faces correspond to the covariation observed when similarity to a sexed rather than a nonsexed prototype was manipulated. In Experiment 2, sexed composite faces were selected as more distinctive less often than were nonsexed ones in a two-choice task, with the selection frequency declining as the faces became closer to the sexed prototype. Considered together, these observations show that distinctiveness ratings correspond to the distance between an individual face and a sexed rather than nonsexed prototype. Our data do not allow us to decide between a norm-based and an exemplar-based model, but they have implications for both of these types of models. First, they show that gender is a dimension of the face space, at least in an indirect way (i.e., without postulating the explicit or implicit categorization of gender). We have seen that gender categories divide any face space that represents faces of both genders (see Figures 1 and 2). This casts doubt on the assumption of a higher density on or around a central tendency and argues in favor of the existence of, at least, two central tendencies, one for each gender. The data from Experiments 1 and 2 show in addition that distinctiveness ratings are based on these sexed central tendencies rather than on an overall reference used for both genders. Thus, participants referred to a sexed prototype rather than to a nonsexed face prototype to rate distinctiveness. It would be risky to conclude that there is a perceptual effect of gender categorization in face recognition, notably from distinctiveness ratings. Some previous studies already demonstrated that a presumed perceptual effect in face processing can be accounted for by decisional operators (see Wenger & Ingvalson, 2002, 2003). Nonetheless, the present study shows that the perceptual difference between gender categories is taken into account in such ratings, at least at a decisional level. Further studies should be done to investigate the precise level (early or late) and nature (gender categorization or physical properties) of this effect.



Further modeling of face recognition processes should thus take gender category into account. A norm-based model should first solve the problem of how many norms are extracted, with different questions arising from the various conceptions. For example, is there a single norm for all types of faces or different norms for different face categories? A single norm would imply that few faces are located close to the prototype and that some kinds of deviations—notably those resulting from gender—are more frequent than others. Our results do not go along with such a conception. Considering many norms would imply considering, first, how many norms exist (for gender categories, ethnicity, others) and, second, how the appropriate norm is selected for the comparison. Regarding exemplar-based models, the existence of different high density areas is less problematic. In this view, the distinctiveness of a face is evaluated by relying on the proximity in the space of the closest neighbors. The existence of different high density areas does not change anything except for the fact that the probability of finding neighbors of the same gender is greater.

Thus, it is possible to account for our results in the framework of an exemplar-based model without postulating the prior categorization of gender. However, exemplar-based models also raise a number of questions, such as the selection of the area that will serve as the reference: Is the distinctiveness of a face evaluated only by comparison to the closest area of higher density or by comparison to all faces in the space, including faces of the other gender? Our results suggest that distinctiveness is based on faces of the same gender, which suggests that the comparison process includes only those faces. Thus, it is not currently possible to rule out that a gender categorization process influences the selection of the neighbors to which a specific face will be compared. In this vein, some data indicate that gender categorization processes affect face recognition and that the two processes are intertwined (Baudouin & Tiberghien, 2002; Ganel & Goshen-Gottstein, 2002; Goshen-Gottstein & Ganel, 2000; Rossion, 2002; but see also Bruce, Ellis, Gibling, & Young, 1987; Bruce & Young, 1986). Further investigations should clarify the role of gender categorization in locating and assessing a point in the face space.

## References

- Baudouin, J.-Y., & Tiberghien, G. (2002). Gender is a dimension of face recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 362–365.
- Baudouin, J.-Y., & Tiberghien, G. (2004). Symmetry, closeness to average, and size of features in the facial attractiveness of women. *Acta Psychologica*, *117*, 313–332.
- Benson, P. J. (1995). Perspectives on face perception: Directing research by exploiting emergent prototypes. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition: Explorations in face space* (pp. 204–224). London: Routledge.
- Brown, E., & Perrett, D. I. (1993). What gives a face its gender? *Perception*, *22*, 829–840.
- Bruce, V., Burton, A. M., & Dench, N. (1994). What's distinctive about a distinctive face? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *47(A)*, 119–141.
- Bruce, V., Ellis, H. D., Gibling, F., & Young, A. W. (1987). Parallel processing of the sex and familiarity of faces. *Canadian Journal of Psychology*, *41*, 510–520.

- Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327.
- Campbell, R., Benson, P. J., Wallace, S. B., Doesbergh, S., & Coleman, M. (1999). More about brows: How poses that change brow position affect perceptions of gender. *Perception*, *28*, 489–504.
- Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *48(A)*, 879–894.
- Cohen, M. E., & Carr, W. J. (1975). Facial recognition and the von Restorff effect. *Bulletin of the Psychonomic Society*, *6*, 383–384.
- Ganel, T., & Goshen-Gottstein, Y. (2002). Perceptual integrity of sex and identity of faces: Further evidence for a single-route hypothesis. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 854–867.
- Going, M., & Read, J. D. (1974). Effect of uniqueness, sex of subject, and sex of photograph on facial recognitions. *Perceptual and Motor Skills*, *39*, 109–110.
- Goldstein, A. G., & Chance, J. E. (1980). Memory for faces and schematic theory. *Journal of Psychology*, *105*, 47–59.
- Goshen-Gottstein, Y., & Ganel, T. (2000). Repetition priming for familiar and unfamiliar faces in a sex-judgment task: Evidence for a common route for the processing of sex and identity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1198–1214.
- Johnston, R. A., Kanazawa, M., Kato, T., & Oda, M. (1997). Exploring the structure of multidimensional face-space: The effects of age and gender. *Visual Cognition*, *4*, 39–57.
- Johnston, R. A., Milne, A. B., Williams, C., & Hosie, J. (1997). Do distinctive faces come from outer space? An investigation of the status of a multidimensional face-space. *Visual Cognition*, *4*, 59–67.
- Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, *1*, 115–121.
- Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1364–1382.
- 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.
- Morpheus Software. (1999–2003). Morpheus (Version 1.85) [Computer software]. Santa Barbara, CA: Author.
- Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology*, *19*, 473–497.
- Rossion, B. (2002). Is sex categorization from faces really parallel to face recognition? *Visual Cognition*, *9*, 1003–1020.

- Shepherd, J. W., Gibling, F., & Ellis, H. D. (1991). The effects of distinctiveness, presentation time and delay on face recognition. *European Journal of Cognitive Psychology, 3*, 137–204.
- 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, 43(A)*, 161–204.
- Valentine, T. (2001). Face-space models of face recognition. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial recognition: Contexts and challenges* (pp. 83–113). Hillsdale, NJ: Erlbaum.
- Valentine, T., & Bruce, V. (1986a). The effect of distinctiveness in recognizing and classifying faces. *Perception, 15*, 525–535.
- Valentine, T., & Bruce, V. (1986b). Recognizing familiar faces: The role of distinctiveness and familiarity. *Canadian Journal of Psychology, 40*, 300–305.
- 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, 44(A)*, 671–703.
- Wenger, M. J., & Ingvalson, E. M. (2002). A decisional component of holistic encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 872–892.
- Wenger, M. J., & Ingvalson, E. M. (2003). Preserving informational separability and violating decisional separability in facial perception and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1106–1118.
- Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 181–190.
- Yamaguchi, M. K., Hirukawa, T., & Kanazawa, S. (1995). Judgment of gender through facial parts. *Perception, 24*, 563–575.