Author manuscript, published in "Journal of Experimental Psychology. Human Perception and Performance 32, 4 (2006) 789-798"
DOI : 10.1037/0096-1523.32.4.789

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–masculinityof real or artificial composite faces. The composite faces were created by blending either faces of thesame 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 distinctivenessratings decreased as the number of blended faces increased. Distinctiveness and gender ratings did notcovary for real faces or sexed composite faces, but they did vary for nonsexed composite faces. InExperiment 2, participants were asked to state which of two composite faces, one sexed and onenonsexed, was more distinctive. Sexed composite faces were selected less often. The results areinterpreted 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 recognitionprocessing in humans showed that faces are encoded and retrievedin 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 ornorm—is assumed to be developed through multiple exposures tofacial stimuli and to capture the average properties of the facecategory. Individual faces are thought to be encoded in terms oftheir deviations from the prototype. This hypothesis has been usedsuccessfully to account for many well-documented phenomena inface recognition. For example, Rhodes et al. (1987) accounted forthe caricature effect in terms of increased distance from a norm;the caricature of a face is recognized more easily than the real facebecause caricatures accentuate the facial characteristics that differentiatethe face from the norm, that is, the features used torecognize the face. Goldstein and Chance (1980) explained therace and inversion effects by studying the facial schema on whichface 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 afacial prototype; the farther a face is from the prototype, the moredistinctive it is. This idea allowed researchers to explain many ofthe effects reported for typical versus distinctive faces. Notably,distinctive faces are recognized more accurately and faster thantypical 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 tobe falsely recognized (Light et al., 1979). Typical faces are alsoclassified faster as a face (by comparison to scrambled faces) thandistinctive ones (Valentine & Bruce, 1986a).

Valentine's (1991) Face-Space Model

In the early 1990s, Valentine (1991) bridged the gap betweenthese separate models by proposing to account for various effectsin the framework of a multidimensional face space. He consideredthe representation of a face to be a point in a multidimensionalEuclidean space. The dimensions of this space are the physiognomicfeatures or facial properties used to encode faces. Theorigin of the space is defined as the central tendency of thedimensions, with the assumption that the values on the dimensionin the population of faces are normally distributed around thecentral tendency. Thus, the density of faces is higher at the centraltendency and decreases as the distance from the central tendencyincreases. 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 representedon the left side of Figure 1.

Two theoretical approaches were proposed to model the codingand recognition of a face in the face-space framework. First, facesare encoded in terms of their deviation from a norm or prototype.This prototype is located at the origin of the space, and it isassumed that there is a single prototype for all faces (Valentine,1991; Valentine & Bruce, 1986a). Valentine (1991) referred to thisapproach as a norm-based model. In such an approach, a face isencoded 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 anexemplar-based model—is to consider that there is no extractedprototype and that the similarity of faces is a monotonic functionof the distance between the faces in the space. The decisionprocess depends on the distance between a face and its nearestneighbors. 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 distinctivenesseffect, Valentine (1991) suggested that because exemplar densityis higher at the origin (i.e., in the area of the space where typicalfaces are), the discrimination of typical faces is more difficult thanthat of distinctive faces, with the probability of a vector or aneighbor falling close to the face being higher.

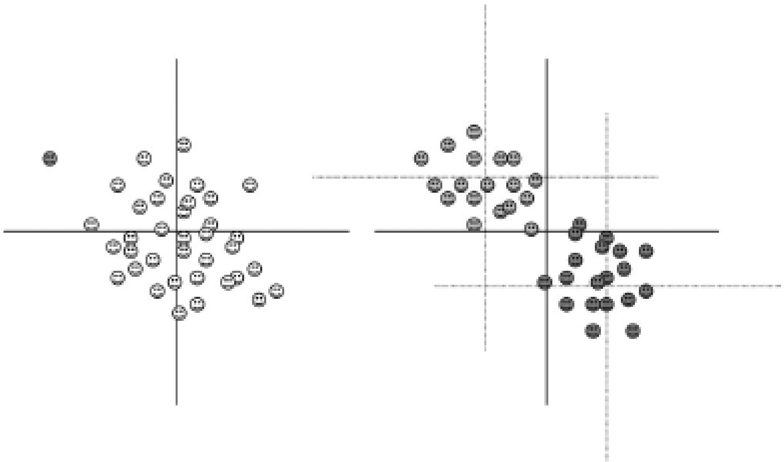


Figure 1. Left: The face-space model, which is based on Valentine's (1991) description. The face density is the greatest at the origin of the space. A distinctive face is located far away from the origin in a low density area (e.g., the gray face). Right: The face-space model with gender categories taken 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 modelhas given rise to a number of investigations. One of the goals ofthese investigations was to verify that the distinctiveness ratingreally increases with the distance between an individual face'scharacteristics and the average value of those characteristics in theface population (e.g., Bruce, Burton, & Dench, 1994; Johnston,Milne, Williams, &Hosie, 1997). Another goal was to comparenorm-based and exemplar-based models, with an advantage forthis last conception (e.g., Valentine & Endo, 1992; but see Levin,1996). Attempts were also made to define the dimensions of theface space (for a review, see Valentine, 2001). Few studies havelooked again at the earlier assumption that there is a single centraltendency of higher density.

The Hypothesis of a Single Central Tendency: Consideration of Gender

A key assumption in the initial studies based on the face-spacemodel was that the origin of the space corresponds to the centraltendency of all dimensions, where the exemplar density is thehighest. This property explains the distinctiveness effect for bothnorm-based and exemplar-based models, even if the rationale isdifferent (i.e., vector similarity vs. distance between exemplars).The assumption of higher density at the origin presupposes that thevalues of the dimension in the population of faces are normallydistributed around the central tendency for *each* dimension. If anydimension in the multidimensional space is bimodal, the pointsplotted in the space will not be grouped around the origin but willbe split in two distinct clouds. This point was already underlinedfor 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 forexample of the dimensions that differentiate the gender categories.Many differences between female and male faces have been reportedin the literature: Female faces have thinner eyebrows, theireyebrows 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 spacerepresents the value taken on by such physical characteristics, thedistribution of the face population will be bimodal, with one modefor each gender. To illustrate this assumption, we measured fourfacial 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 designedby Baudouin and Tiberghien (2004) for measuring facial features.The distribution of faces on the height of eyebrows dimension isillustrated in Figure 2a. As Figure 2a shows, the mode of eachgender category is different. The result when all faces are consideredtogether is that the face population is not normally distributedaround a central tendency but quite bimodal with two “central”tendencies, one for each gender. The existence of a single dimensionrepresenting gender (Johnston, Kanazawa, Kato, &Oda,1997) would give the same result. Figure 2b presents the factorialresults for the first factor of a principal-components analysisperformed on the four facial measurements. These results can be considered to code the gender category of the faces (i.e., it correspondsto a composite of the four measurements), and as Figure 2bshows, its distribution is not normal but multimodal. The insertionof this dimension in a multidimensional face space would thusresult in various pools of high density areas (see right part ofFigure 1 for an illustration).1The partition of the face space into two high density areas by thegender categories highlighted many problems that were not consideredin the current literature (see also Benson, 1995). Thisquestion was generally tackled in studies on the properties of theface space by looking at faces of only one gender. It was neverthelessoften suggested (generally implicitly) that the distinctivenessof a face is assessed by comparison to a sexed prototype. Inthe study by Bruce et al. (1994), for example, the authors definedthe distinctive characteristics of a set of faces by comparingsubjective ratings of facial distinctiveness with objective facialmeasurements. They reported a correlation between distinctivenessratings and the physical deviation of faces from the averagecomputed from faces—of the same gender. An exception is thestudy by Johnston, Kanazawa, et al. (1997), who suggested thatboth gender and age are dimensions of the face space.

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

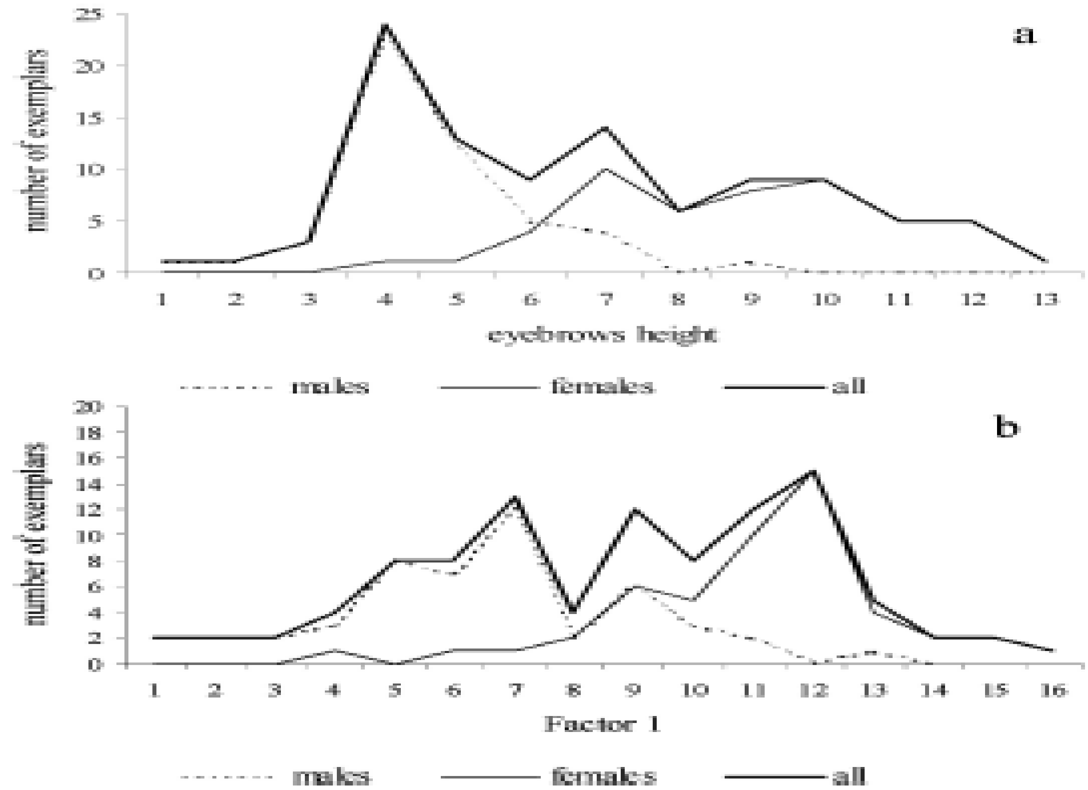


Figure 2. Illustration of the distribution of 50 female and 50 male faces on the eyebrow-height dimension (a) or on a gender dimension (Factor 1) derived from four facial measurements (b). The central tendency for each

It is important to consider gender categories because they partitionthe plot of faces in a similar way as do ethnic categories, withthe main exceptions that (a) gender categories are probably closerto each other in the space than ethnic categories and (b) thedifferent categories have probably been seen to more equal extentsfor gender than for ethnic origin. Many problems have thus beenraised regarding the partition of faces in the face space. One of theproblems concerns norm-based models: Should it be assumed thatthere 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 femaleand male faces and lying at the origin of the space, implies thatmost faces are not located around the origin but lie away from it intwo main directions on the dimensions that differentiate the gendercategories. The prototype, which would have few gender markers,is even atypical if one considers that most faces have gendermarkers. A consequence of this is that deviations from the prototype(assumed to increase distinctiveness) put the face in a regionof high density when the deviations concern features that differentiatethe genders. In other words, some deviations from theprototype are more frequent in the population of faces than others,resulting in some distinctive faces that are seen frequently. Theassumption of two prototypes raises some other questions, notablythe problem of selecting the appropriate prototype. A typicalfemale face can be considered as an atypical male face, so it maybe important to compare a face with the prototype of its owngender category. But, do people select the appropriate prototypebefore performing encoding and recognition processes? Or, is anencoded face compared with both prototypes at the same time?The hypothesis of a dual comparison raises the question of how thesystem manages the two sources of information.The existence of two high density areas is less problematic forexemplar-based models. Given that the decision process dependson the distance between a face and its nearest neighbors, the mainconsequence of having two high density areas is that the probabilityof encountering a face of the same gender in the neighborhoodis higher than the probability of encountering a face of theopposite gender. A distinctive face will then be a face that deviatesfrom the central tendency of one pool of faces without lying in thepool of faces of the other gender. The existence of two genderdefinedareas nevertheless raises some questions. In particular, isgender category taken into account in decision processes—that is,is the face compared with neighbors even when they belong to theopposite gender category? Baudouin and Tiberghien (2002)showed that the assignment of a specific gender to ambiguousfaces affected the distractor-comparison process: When a targetface was assigned a gender opposite to the gender of the distractors,the latter were rejected faster than when the face was assignedthe same gender. Thus, despite strictly equivalent distances forphysical properties, the pool of exemplars (or the prototype?) usedto encode and/or retrieve a face differed according to the genderattributed to the encoded face. The question raised by this findingconcerns whether distinctiveness is assessed by comparison withthe population of faces or with the population of same-genderfaces.

Thus, there are reasons to question the role of gender categoryin facial distinctiveness. Notably, a major question is whetherdistinctiveness is evaluated by comparing an encoded face with aprototype (or a set of exemplars) that shares the same categoricalproperties (i.e., a prototype of the category) or, on the contrary,with a more global prototype that captures general average propertiesof the population of faces and thus discards properties thatdivide the population into distinct categories. The purpose of thepresent study was to test whether the distinctiveness of a face isassessed by relying on the sexed prototype (or pool of same-genderfaces) of the face’s gender category or by relying on a faceprototype derived from both female and male faces. In Experiment1, participants were asked to rate the distinctiveness and the genderof faces that were either real female or male faces or compositefaces made with increasing numbers of faces (2, 4, 8, 16, 32, and64). Increasing the number of faces in the composite face wassupposed to put the composite face closer to the average orprototypical properties of the population of faces (Langlois&Roggman, 1990). Consequently, the distinctiveness of the compositefaces should decrease as the number of faces in themincreases. Two types of composite faces were used: (a) sexedcomposite faces, in which only faces of the same gender weremixed in, and (b) nonsexed composite faces, in which an equalnumber 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 toapproximate the (nonsexed) face prototype. If distinctiveness isdetermined by comparison with a sexed prototype, the decrease indistinctiveness with the increase in the number of faces shoulddepend on or be greater for composite faces that approximate asexed prototype. In Experiment 2, participants were shown pairs ofcomposite faces with one sexed and one nonsexed composite facein each pair, and they were instructed to indicate which of the twofaces was more distinctive. If distinctiveness is assessed on thebasis of a sexed prototype, composite faces approximating thiskind of prototype (i.e., sexed composite faces) should be selectedless often. Even if we use the term *prototype*, these experimentswere not designed to decide between norm-based or exemplarbasedmodels but to test properties that have implications for bothkinds of approaches.

Experiment 1

Average composite faces were created by blending differentindividual faces. The number of faces in each average face wasincreased from 1 to 64, with five intermediate levels (2, 4, 8, 16,and 32 faces). Two types of average faces were created based onthe gender of the faces of which they were composed: (a) Some ofthe average faces were sexed (i.e., only faces of the same genderwere used), and (b) others were nonsexed (i.e., an equal number offaces of each gender were used). Participants were asked first torate the distinctiveness of the faces and then to categorize and ratethem 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 wasdone: The faces were drawn from a person database (created by Jean-YvesBaudouin) according to their real sex, and the first 32 female and malefaces were selected. The faces were photographed with a Kodak digitalcamera. The models were face front, with a neutral expression. Each facewas put into an oval so that the hairstyle was not visible. The oval was 670pixels high _ 510 pixels wide, which corresponded to a size of about14.5 _ 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, 256points were put on the main features of a face. The points were then placedin the same locations on another face. The software computed an intermediatelevel between the locations of the points and created a new face forwhich points were at an equal distance from the two original points. It alsocomputed an intermediate texture for surfaces delineated by points. Thecomposite face thus had features for which position, size, and texture wereat an equal distance from position, size, and texture of the features of thetwo original photographs. The resulting composite face can therefore beconsidered as the average of the two original faces.The 64 faces were divided into 16 sets of 4 faces each (2 female and 2male). For each set, we generated 4 average faces by blending 2 faces foreach. The first 2 average faces were made from the faces of the samegender (i.e., the 2 female faces and the 2 male faces in the set) andcorresponded to average sexed faces (female and male, respectively). Theother 2 average faces were created with faces of the opposite gender andcorresponded to average nonsexed faces (Androgynous Face 1 and AndrogynousFace 2). This gave us four groups of 16 averaged faces: In twogroups, the faces were sexed (16 female and 16 male), and in the other two,they were nonsexed (Androgynous Face 1 and Androgynous Face 2). Ineach group, the faces were then averaged two by two, resulting in 8 newaveraged faces per group. These 8 averaged faces thus represented theaverage position, size, and texture of 4 faces each. They were mixed twoby 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 two16-face averages were blended to obtain an average of 32 faces. The32-face averages obtained from faces of the same gender were thenblended to obtain an average face of the 64 original faces. The sameprocedure was performed with the two 32-face averages obtained fromfaces of different genders. An illustration of the composite faces is presentedin 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 _ 16nonsexed), 32 average faces from 4 original faces (8 female, 8 male, and2 _ 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, 2male, and 2 _ 2 nonsexed), 4 average faces from 32 original faces (1female, 1 male, and 2 _ 1 nonsexed), and 2 average faces from 64 originalfaces. The 32-face average faces created with the two groups of sexed faceswere used as the sexed prototypes. The 2 average faces of 64 (created,respectively, with the 2 sexed 32-face average faces and the 2 nonsexed32-face average faces) were used as facial prototypes.

Procedure

Participants sat in front of a computer at a distance of about 80 cm. Theparticipants performed two tasks in two distinct sessions.*Distinctiveness rating.*After a fixation point, a face was presented inthe center of the screen. Participants were told to evaluate its distinctivenesson 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 anatypical, unusual face that should be easy to recognize later. Participantswere also told that none of the faces in the experiment were real and thatthey had been made using computer software. So, they had to disregard anylack of realism in the faces by imagining that they were of actual persons.Each participant performed 190 trials (i.e., all faces), presented in randomorder. The faces remained on the screen until the participant responded.There was no time limit.

Gender categorization

After the distinctiveness rating task, the 190faces were presented again to participants. They were instructed to state foreach face whether the person was female or male using two keys on thekeyboard. After this response, they had to use the numeric pad to rate thefemininity 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 faceswere included (real faces and 64-face composites). The type ofcomposite face (sexed or nonsexed) was not considered. Thepurpose of the analysis was to test for the hypothesized decrease indistinctiveness. In the second analysis, the type of composite facewas taken into account, but real faces as well as 64-face compositeswere 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 decreasedas 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 decreasedas the number of faces increased, rapidly for the first step (2 faces)and slower for the other steps. The difference was significantbetween 1 and 2 faces (6.0 vs. 4.3), $F(1, 19)$ _ 66.55, p _ .01; 2and 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 (2.9vs. 2.7), $F(1, 19)$ _ 5.18, p _ .05. The difference was notsignificant between 16 and 32 faces, $F(1, 19)$ _ 0.72, or between32 and 64 faces, $F(1, 19)$ _ 3.24.

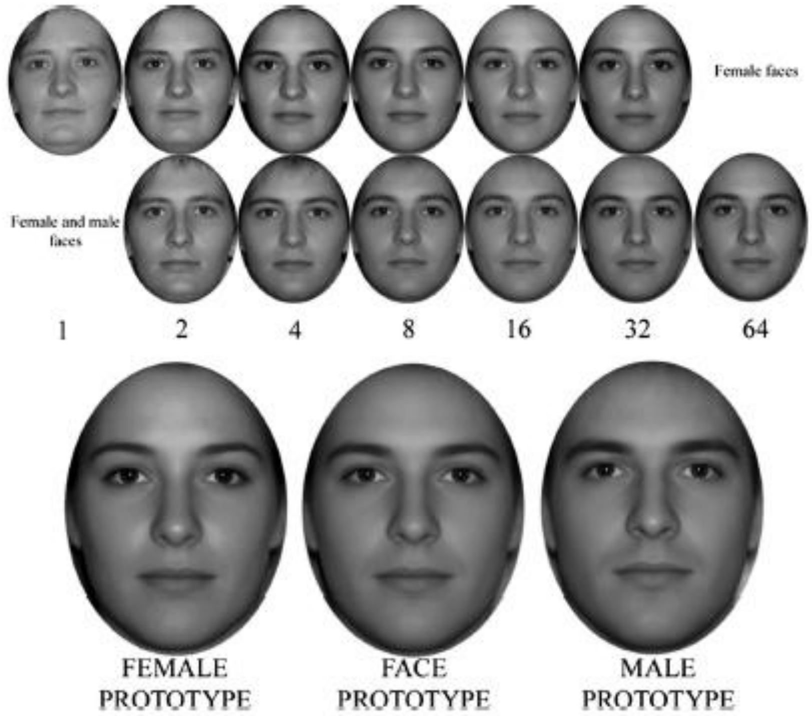


Figure 3. Illustration of composite faces. Top: A female face is blended with an increasing number of faces, either of the same gender or of both genders, to create a female or androgynous prototypical face. Bottom: The two sexed prototypes (32 faces) and the (androgynous) face prototype (64 faces).

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							
M	6.0	4.3	3.3	2.9	2.5	2.7	
SD	0.8	1.1	1.6	1.6	1.7	2.0	
Nonsexed							
M		4.4	3.4	2.8	2.8	3.0	2.5
SD		1.0	1.4	1.7	1.7	2.3	2.0
Overall							
M	6.0	4.3	3.3	2.9	2.7	2.9	2.5
SD	0.8	1.0	1.5	1.6	1.7	2.0	2.0

*Same versus different gender.*Preliminary analyses indicatedno differential effect of gender for sexed composite faces (femalevs. male) or of the group of nonsexed composite faces (AndrogynousFace 1 or Androgynous Face 2). The data for each type offace were thus averaged within sexed and nonsexed compositefaces. We performed a two-factor ANOVA (number of faces: 2 vs.4 vs. 8 vs. 16 vs. 32; type of faces: sexed vs. nonsexed—bothwithin subjects) to determine whether distinctiveness decreased ina different way for the two in the composite faces. The effect ofnumber of faces was significant, F(4, 76) _ 19.55, p _ .01,replicating the effect reported in the previous analysis. So we willnot present this analysis. The type-of-faces factor was nonsignificant,F(1, 19) _ 1.24, and did not interact with the number offaces, 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 wereassigned a positive value. For each indicator, we performed a 2 _2 _ 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. AndrogynousFace 2), and number of faces (2 vs. 4 vs. 8 vs. 16 vs. 32) aswithin-subject factors.

*Percentage of “female” responses.*Figure 4a illustrates theresults. 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 significantinteraction between exemplar and type of face, *F*(1, 19) _ 693.65,*p* _ .01, indicating that the exemplar factor had a significant effectfor sexed composite faces (95.2% “female” responses for femalecomposite 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 alsosignificant, *F*(4, 76) _ 3.81, *p* _ .01, but it was qualified by asignificant interaction with the type of face, *F*(4, 76) _ 3.35, *p* _.05. This interaction indicated a significant effect of number offaces for nonsexed composite faces, *F*(4, 76) _ 4.07, *p* _ .01, butnot for sexed ones, *F*(4, 76) _ 0.85. Figure 2a indicates thatparticipants had a tendency to respond “male” for nonsexed compositefaces when the number of faces was low, but this tendencydisappeared when the number of blended faces was increased. Theoverall 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 asignificant interaction between exemplar and type of face, *F*(1,19) _ 443.78, *p* _ .01. This interaction indicated that the exemplarhad a significant effect for sexed composite faces (5.6 for femalecomposite 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. Themain 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 _Type of Face interaction, *F*(4, 76) _ 4.44, *p* _ .01, and a Numberof Faces _ Type of Face _ Exemplar interaction, *F*(4, 76) _ 2.86,*p* _ .05. In this last interaction, the number of faces had asignificant 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 notfor female ones, *F*(4, 76) _ 1.44. Figure 2b shows that there wasa tendency to rate nonsexed composite faces as masculine whenfew faces were blended. This tendency decreased as the number offaces increased. For male faces, increasing the number of malefaces decreased perceived masculinity.

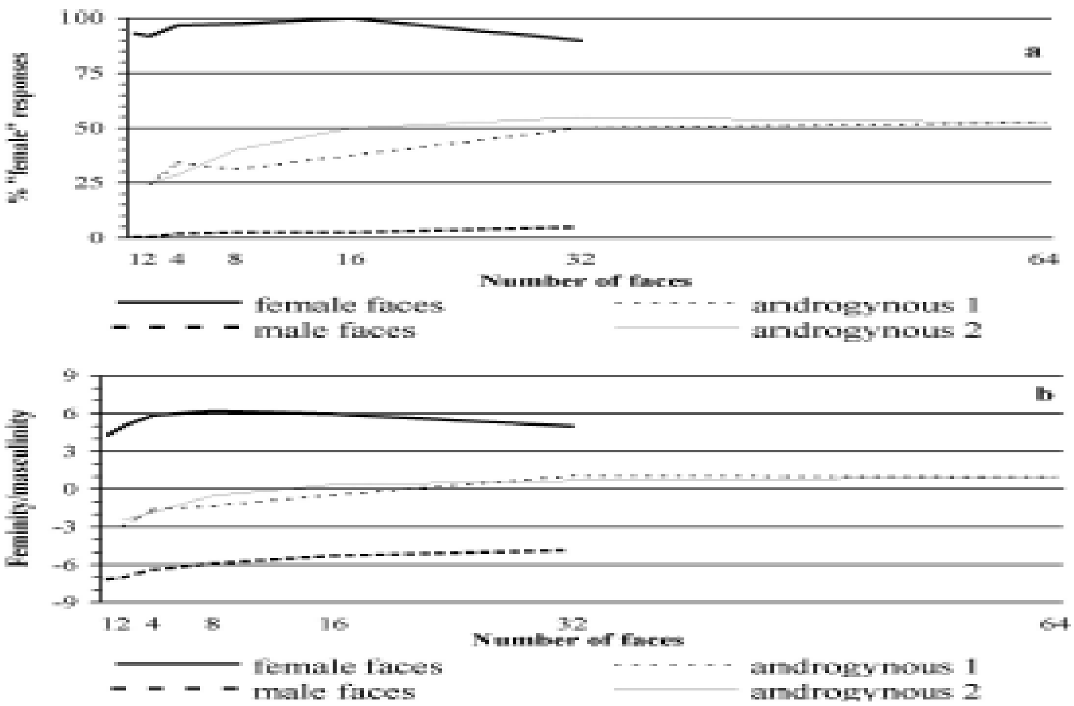


Figure 4. Performance on the gender categorization task: Percentage of "female" responses (a) and femininity–masculinity ratings (b).

Correlations Between Distinctiveness and GenderCategorization

Two gender categorization indicators were used. The first wasan indicator of gender based on "female" responses. The percentagesfor each face were transformed so that percentages lower than50% were replaced by the difference between that percentage and100% (i.e., a face that was labeled as female in 0% of responseswas assigned a percentage of 100%). This transformation gave usan indicator of the participants' agreement about gender: Astrongly sexed face—either as female or male—had a score closeto 100%. On the contrary, an ambiguous face had a score close to50%. 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 ambiguousor androgynous face had a rating close to zero. To test for apotential effect of gender on distinctiveness, Bravais–Pearson correlationswere computed between the distinctiveness rating andthese two gender categorization indicators. The results are summarizedin Table 2.As indicated in Table 2, there was a significant correlationbetween the distinctiveness ratings and the two gender categorizationindicators when all faces were considered together. Facestended to be more distinctive when gender attributes were moreprominent, which led to greater agreement among subjects andhigher femininity–masculinity ratings. Nevertheless, when correlationswere computed for real faces, sexed composite faces, andnonsexed composite faces, the link between the distinctivenessrating and the gender indicators was no longer significant for eitherreal faces or sexed composites. It was always significant whennonsexed composite faces were considered. Before consideringthese observations, we examine the distribution of faces in a 2-Dspace, with distinctiveness and the gender indicators asdimensions.Figure 5 illustrates the distribution of the faces used in theexperiment in a 2-D space with distinctiveness and the percentageof "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**

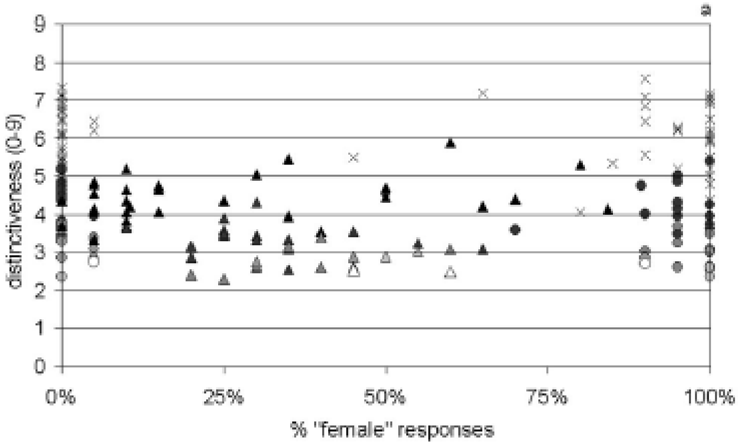
Distinctiveness,the percentage of "female" responses, and the femininity– masculinity ratings were averaged across participants for eachface. If we consider only real faces (crosses in Figure 5), two poolsof faces emerged along the gender dimensions, one for each gendercategory. Very few real faces lie between these two pools, and thedistinctiveness ratings of these faces were quite high (between 4.1and 7.6). Increasing the number of same-gender faces (circles ingradations of gray to white) made the distinctiveness ratings decrease,but the percentage of "female" responses and thefemininity–masculinity ratings remained relatively constant. Byadding more and more faces of the opposite gender (gray to whitetriangles), distinctiveness, the participants' agreement about gender,and the femininity–masculinity ratings decreased.

Thus, Figure 5 suggests that there are two ways in whichdistinctiveness can decrease: (a) The face becomes closer to thesexed prototype, or (b) the face becomes closer to the androgynousface prototype. In the first case, the decrease in distinctiveness wasindependent of the gender indicators. In the second case, it wasrelated to them. The less distinctive a face, the less it will exhibitgender markers. These assumptions are in line with the coefficientsof correlation reported in Table 2: The correlation between distinctivenessand the gender indicators was nonsignificant for sexedcomposite faces, but it was significant for nonsexed compositeones. This confirms the assumption that the relationship betweendistinctiveness and gender is dependent on whether the faces werecloser to a sexed or nonsexed prototype. It is thus possible to findout whether the distinctiveness of real faces varies according totheir closeness to a sexed or nonsexed prototype. One can hypothesizein the first case that distinctiveness and gender characteristicsare unrelated, and in the second case that they are significantlyrelated. Figure 5 suggests that there was in fact no relationshipbetween them, which was confirmed by the nonsignificantBravais–Pearson coefficient of correlation for real faces (see

Table2). Thus, the distinctiveness of real faces is independent of theirgender characteristics, suggesting that it is determined by comparisonto a sexed prototype.

Conclusions

The results of Experiment 1 show that distinctiveness graduallydecreased with the number of averaged faces. A floor level wasreached at 16 faces, after which increasing the number of faces inthe composite did not give rise to an additional decrease in thedistinctiveness ratings. Moreover, the gender of the faces that wereblended had no effect on the mean distinctiveness ratings. Correlationanalyses nevertheless indicated that the distinctiveness ratingsof the real faces, unlike the nonsexed composite faces but likethe sexed ones, did not covary with gender saliency as indicated bythe femininity–masculinity ratings and the participants' agreementabout gender category. This observation suggests that the distinctivenessof real faces is assessed by comparison to a sexedprototype.



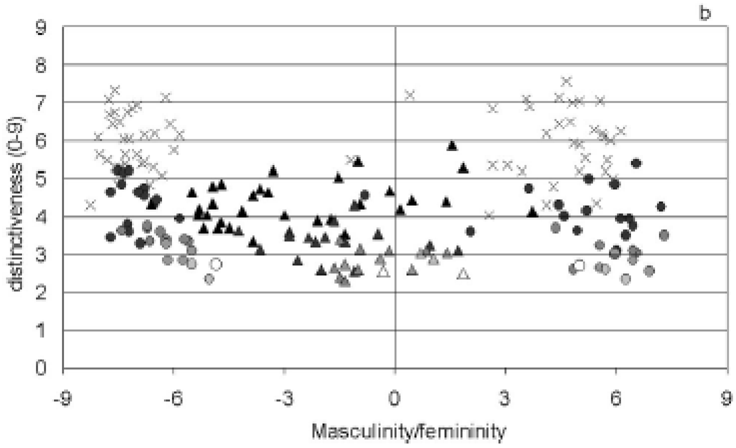


Figure 5. Scatterplots of distinctiveness ratings versus percentage of "female" responses (a) and distinctiveness ratings and femininity–masculinity ratings (b). Real faces are represented by crosses, same-gender composites with increasing number of faces are represented by circles in gradations of gray 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 despitesome indication of an influence of gender category on distinctivenessrating, no difference in the mean distinctiveness ratings wasobserved between sexed and nonsexed composite faces. The means in Table 1 suggest that sexed composite faces with 16 or 32faces were in fact less distinctive than nonsexed ones, but thedifference was not statistically significant. A possible explanationis that the measure used in this experiment—a rating scale—wasnot sensitive enough to allow a potential effect to show up. Realfaces were presented together with composite ones in such a waythat the variability in their distinctiveness was relatively high. Thismay have reduced fine discriminations between faces of relativelysimilar, low distinctiveness, like the sexed and nonsexed compositefaces. To further test for a potential difference in distinctivenessfor such composite faces with a more sensitive measure, we useda two-choice task. Participants were required to say which of twocomposite faces—a sexed and a nonsexed one—was the mostdistinctive. 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, thenonsexed prototypical faces should be selected more often.

Experiment 2

Method

Participants

Sixteen new participants (9 women and 7 men) ranging in age from 21to 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 16original faces (2 female, 2 male, 2 _ 2 nonsexed), and the 4 average facesthat 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 ofthe screen along with another composite face on the right side. Theparticipants were asked to press the left or right arrow on the keyboard,depending on which face they found the most distinctive. They were toldthat a distinctive face was an atypical, unusual face that would be easy torecognize later. They were also told that none of the faces in the experimentwere real and that they had been generated using computer software. Sothey had to disregard any lack of realism in the faces by imagining that theycorresponded to real persons. Every sexed composite face was presentedwith a nonsexed composite face that blended the same number of faces(i.e., sexed composite faces made from 8 faces were displayed withnonsexed composite faces also made from 8 faces). Each participant saw84 pairs (4 for 32-face composites, 16 for 16-face composites, and 64 for8-face composites), presented in random order. The pairs of faces remainedon the screen until the participant responded. For each participant, sexedand nonsexed composite faces were displayed equally often on the left andright sides of the screen, with the right–left location of the faces beingalternated across participants.

Results

For each composite face, we computed how often (in percentage)it was selected as being more distinctive than the other face inthe 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 twofactors were between-item factors) and by subjects (the two factorswere within-subject factors) was conducted to find out whetherdistinctiveness 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 faceswere selected as more distinctive more often than were sexed faces(62.8% vs. 37.2%): by items, $F(1, 22)$ _ 104.99, p _ .01; bysubjects, $F(1, 15)$ _ 6.00, p _ .05. Neither the main effect of thenumber of faces nor the interaction between the number of facesand 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 todetermine whether the selection rate of each face (e.g., “this faceis more distinctive”) was correlated with the participants’ agreementrate about gender category and with the absolute value of themean femininity–masculinity ratings obtained for these same facesin Experiment 1. There was a strong and significant negativecorrelation 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 moreobvious the gender, the less often the face was judged to be moredistinctive.

Conclusions

Experiment 2 showed that composite faces made up of samegenderfaces were less distinctive than composite faces made up offaces of both genders. This finding argues in favor of the idea thatdistinctiveness is assessed by means of a comparison to a sexedprototype rather than a nonsexed prototype. Correlation analysesfurther strengthened this conclusion. The composite faces weremore often selected as the less distinctive face in the pair whentheir gender markers were more prominent. In this experiment, thegender markers corresponded to features that were averaged acrossat least eight faces. So there were probably few atypical sexedfeatures but more probably sexed features that corresponded toprototypical characteristics of each gender category. Thus, faces whose gender markers were more prominent probably tended to becloser to one of the two sexed prototypes. Consequently, thenegative covariation between distinctiveness ratings and the genderindicators further indicated that the faces were even lessdistinctive 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

	Number of faces		
Type of face	8	16	32
Sexed composite faces			
Mean selection rate (%)	39.5	39.5	32.8
SD	16.9	23.0	27.0
Nonsexed composite faces			
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 adeterminant of face distinctiveness. In Experiment 1, the covariationof the distinctiveness ratings and the gender indicators of realfaces correspond to the covariation observed when similarity to asexed rather than a nonsexed prototype was manipulated. In Experiment2, sexed composite faces were selected as more distinctiveless often than were nonsexed ones in a two-choice task, withthe selection frequency declining as the faces became closer to thesexed prototype. Considered together, these observations showthat distinctiveness ratings correspond to the distance between anindividual face and a sexed rather than nonsexed prototype.Our data do not allow us to decide between a norm-based and anexemplar-based model, but they have implications for both ofthese types of models. First, they show that gender is a dimensionof the face space, at least in an indirect way (i.e., without postulatingthe explicit or implicit categorization of gender). We haveseen that gender categories divide any face space that representsfaces of both genders (see Figures 1 and 2). This casts doubt on theassumption of a higher density on or around a central tendency andargues in favor of the existence of, at least, two central tendencies,one for each gender. The data from Experiments 1 and 2 show inaddition that distinctiveness ratings are based on these sexedcentral tendencies rather than on an overall reference used for bothgenders. Thus, participants referred to a sexed prototype ratherthan to a nonsexed face prototype to rate distinctiveness. It wouldbe risky to conclude that there is a perceptual effect of gendercategorization in face recognition, notably from distinctivenessratings. Some previous studies already demonstrated that a presumedperceptual effect in face processing can be accounted for bydecisional operators (see Wenger &Ingvalson, 2002, 2003). Nonetheless,the present study shows that the perceptual differencebetween gender categories is taken into account in such ratings, atleast at a decisional level. Further studies should be done toinvestigate the precise level (early or late) and nature (gendercategorization or physical properties) of this effect.

Further modeling of face recognition processes should thus takegender category into account. A norm-based model should firstsolve the problem of how many norms are extracted, with differentquestions arising from the various conceptions. For example, isthere a single norm for all types of faces or different norms fordifferent face categories? A single norm would imply that fewfaces are located close to the prototype and that some kinds ofdeviations—notably those resulting from gender—are more frequentthan 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 differenthigh density areas is less problematic. In this view, the distinctivenessof a face is evaluated by relying on the proximity in thespace of the closest neighbors. The existence of different higherdensity areas does not change anything except for the fact that theprobability of finding neighbors of the same gender is greater.

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

References

Baudouin, J.-Y., &Tiberghien, G. (2002). Gender is a dimension of facerecognition. *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 researchby exploiting emergent prototypes. In T. Valentine (Ed.), *Cognitive andcomputational 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 adistinctive face? *Quarterly Journal of Experimental Psychology: HumanExperimental Psychology, 47*(A), 119–141.

Bruce, V., Ellis, H. D., Gibling, F., & Young, A. W. (1987).Parallelprocessing of the sex and familiarity of faces.*Canadian Journal ofPsychology, 41,* 510–520.

Bruce, V., & Young, A. W. (1986).Understanding face recognition. *BritishJournal 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 affectperceptions of gender. *Perception, 28,* 489–504.

Chiroro, P., & Valentine, T. (1995).An investigation of the contacthypothesis of the own-race bias in face recognition.*Quarterly Journal ofExperimental Psychology: Human Experimental Psychology, 48*(A),879–894.

Cohen, M. E., & Carr, W. J. (1975).Facial recognition and the von Restorffeffect.*Bulletin of the Psychonomic Society, 6,* 383–384.

Ganel, T., & Goshen-Gottstein, Y. (2002). Perceptual integrity of sex andidentity of faces: Further evidence for a single-route hypothesis. *Journalof Experimental Psychology: Human Perception and Performance, 28,*854–867.

Going, M., & Read, J. D. (1974).Effect of uniqueness, sex of subject, andsex of photograph on facial recognitions.*Perceptual and Motor Skills,39,* 109–110.

Goldstein, A. G., & Chance, J. E. (1980).Memory for faces and schematheory.*Journal of Psychology, 105,* 47–59.

Goshen-Gottstein, Y., &Ganel, T. (2000). Repetition priming for familiarand unfamiliar faces in a sex-judgment task: Evidence for a commonroute for the processing of sex and identity. *Journal of ExperimentalPsychology: Learning, Memory, and Cognition, 26,* 1198–1214.

Johnston, R. A., Kanazawa, M., Kato, T., &Oda, M. (1997).Exploring thestructure 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). Dodistinctive faces come from outer space? An investigation of the statusof a multidimensional face-space.*Visual Cognition, 4,* 59–67.

Langlois, J. H., &Roggman, L. A. (1990). Attractive faces are onlyaverage. *Psychological Science, 1,* 115–121.

Levin, D. T. (1996). Classifying faces by race: The structure of facecategories. *Journal of Experimental Psychology: Learning, Memory,and Cognition, 22,* 1364–1382.

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

Morpheus Software. (1999 –2003). Morpheus (Version 1.85) [Computersoftware]. Santa Barbara, CA: Author.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings ofcaricatures: Implications for mental representations of faces. *CognitivePsychology, 19,* 473–497.

Rossion, B. (2002). Is sex categorization from faces really parallel to facerecognition? *Visual Cognition, 9,* 1003–1020.

Shepherd, J. W., Gibling, F., & Ellis, H. D. (1991). The effects ofdistinctiveness, presentation time and delay on face recognition. *EuropeanJournal 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 ExperimentalPsychology: 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 processperspectives on facial recognition: Contexts and challenges* (pp.83–113). Hillsdale, NJ: Erlbaum.

Valentine, T., & Bruce, V. (1986a). The effect of distinctiveness in recognizingand classifying faces. *Perception, 15,* 525–535.

Valentine, T., & Bruce, V. (1986b).Recognizing familiar faces: The roleof distinctiveness and familiarity.*Canadian Journal of Psychology, 40,*300–305.

Valentine, T., & Endo, M. (1992). Towards an exemplar model of faceprocessing: The effects of race and distinctiveness. *Quarterly Journal ofExperimental Psychology: Human Experimental Psychology, 44*(A),671–703.

Wenger, M. J., &Ingvalson, E. M. (2002).A decisional component ofholistic encoding.*Journal of Experimental Psychology: Learning, Memory,and Cognition, 28,* 872–892.

Wenger, M. J., &Ingvalson, E. M. (2003). Preserving informationalseparability and violating decisional separability in facial perception andrecognition. *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.