Journal of Experimental Child Psychology **77,** 269–291 (2000)
doi:10.1006/jecp.1999.2554, available online at http://www.idealibrary.com on IDEAL®

Recognition and Sex Categorization of Adults' and Children's Faces: Examining Performance in the Absence of Sex-Stereotyped Cues

Heather A. Wild

The University of Texas at Dallas

Susan E. Barrett

Lehigh University

and

Melanie J. Spence, Alice J. O'Toole, Yi D. Cheng, and Jessica Brooke

The University of Texas at Dallas

The ability of children and adults to classify the sex of children's and adults' faces using only the biologically based internal facial structure was investigated. Face images of 7- to 10-year-old children and of adults in their 20s were edited digitally to eliminate hairstyle and clothing cues to sex. Seven-year-olds, nine-year-olds, and adults classified a subset of these faces by sex and were asked, subsequently, to recognize the faces from among the entire set of faces. This recognition task was designed to assess the relationship between categorization and recognition accuracy. Participants categorized the adult faces by sex at levels of accuracy varying from just above chance (7-year-olds) to nearly perfect (adults). All participant groups performed less accurately for children's faces than for adults' faces. The 7-year-olds were unable to classify the children's faces by sex at levels above chance. Finally, the faces of children and adults were equally recognizable—a finding that has theoretical implications for understanding the relationship between categorizing and identifying faces. © 2000 Academic Press

Key Words: face categorization; face recognition; gender perception.

Discrimination of male and female faces is easily performed by adults, children, and infants (e.g., Burton, Bruce, & Dench, 1993; Intons-Peterson, 1988;

The authors gratefully acknowledge support from NIMH Grant MH5176501 awarded to A. J. O'Toole and a research initiative grant from the University of Texas at Dallas School of Human Development awarded to H. A. Wild.

Address correspondence and reprint requests to H. A. Wild, c/o A. J. O'Toole, School of Human Development, GR 4.1, The University of Texas at Dallas, Richardson, TX 75083-0688. E-mail: otoole@utdallas.edu.

0022-0965/00 $35.00

Fagot & Leinbach 1993, respectively). Studies examining children's discrimination of male and female faces, however, have used face stimuli that include social and cultural cues to sex, such as those available from hairstyle and clothing. When these cues are present, infants as young as 9 months of age categorize pictures of males and females (Fagot & Leinbach, 1993; cf. also Cornell, 1974), and children as young as 27 months of age correctly assign appropriate labels to pictures of men and women and of boys and girls (e.g., Hort, Leinbach, & Fagot, 1991; Fagot, Leinbach, & Hagan, 1986). Indeed, research on the development of gender constancy suggests that social cues for gender are of primary importance for children's gender assignments until about the age of 6 (Slaby & Frey, 1975). Preschoolers who have not yet achieved gender constancy believe that a boy who lets his hair grow long and wears a dress becomes a girl (Emmerich, Goldman, Kirsch, & Sharabany, 1977; Huston, 1983; Kohlberg, 1966; Marcus & Overton, 1978). The fundamental question these gender constancy studies address is whether children appreciate that an underlying reality, namely sex, remains unchanged across a sequence of transformations, a fact that frequently escapes the preoperational child.

In addition to culturally based cues, the structure of the face is a highly reliable cue to the sex of an individual (e.g., Bruce, Burton, Hanna, Healey, & Mason, 1993; Chronicle et al., 1995; Yamaguchi, Hirukawa, & Kanazawa, 1995). No studies have examined whether children can categorize on the basis of sex using only facial structure cues, although this information is certainly as relevant for sex classification as are cues such as hairstyle and clothing. In the present experiment, we tested the extent to which children and adults can determine the sex of individual faces based on internal facial structure in the absence of social and cultural cues.

Although it is an open question whether children are capable of using structurally based cues, there is a wealth of evidence indicating that adults can make accurate and efficient sex categorizations when only structural cues are present (e.g., Bruce et al., 1993; Chronicle et al., 1995; Yamaguchi et al., 1995). This is consistent with the anthropological literature, which points to a number of facial structures that differ for adult male and female faces (Enlow, 1982). For example, the shapes of male and female faces differ in that male faces tend to be longer and less round than female faces (Enlow, 1982). Additionally, males have a larger nose and a broader forehead than females. Several psychological studies have shown that the size of the nose can play a key role in determining the sex of an adult face (Chronicle et al., 1995; Yamaguchi et al., 1995).

Computational analyses of the information in faces can be used also to determine the quality and nature of the information useful for differentiating male from female faces. The global structural differences between Japanese male and female faces have been illustrated graphically by Yamaguchi et al. (1995) using morphing methodology. They created a prototypical male face and a prototypical female face by morphing together pairs of same-sex faces, and then repeatedly morphing these pairs of morphs together to converge on a single male prototype and a single female prototype. The resultant prototypes clearly showed global structural differences between the male and female faces. These global differences consist of large-scale shape differences that are difficult to describe concisely using verbal labels, but which are easily associated with male versus female faces.

The dichotomous global structures uncovered in the morphing demonstration of Yamaguchi et al. (1995) can be dissected further into individual components using a different computational strategy based on principal component analysis (PCA). PCA, a metric form of multidimensional scaling or factor analysis, is a method for quantifying the information in faces that has been applied both to two-dimensional images of adult faces (O'Toole, Abdi, Deffenbacher, & Valentin, 1993) and to three-dimensional adult head structures from laser scans (O'Toole, Vetter, Troje, & Bulthoff, 1997). In recent years, PCA has been proposed as a perceptual learning model of human face recognition and categorization (O'Toole, Abdi, Deffenbacher, & Valentin, 1995). PCA produces a set of orthogonal dimensions (eigenvectors), each of which explains a proportion of variance in a data set. Because PCA is applied directly to a physical representation of faces (e.g., the two-dimensional image or three-dimensional structure of a face), the axes or eigenvectors can be displayed (e.g., as two-dimensional images or three-dimensional structures). To interpret the meaning of these axes, individual eigenvectors can be added to or subtracted from the average face (or head structure if laser scans are used).

As is illustrated in Fig. 1, O'Toole et al. (1997) used this procedure to show that two eigenvectors were reliable predictors of the sex of the face. In the top row of this figure, the image on the left was created by adding the first eigenvector to the average head and the image on the right was created by subtracting the first eigenvector from the average head. As this procedure demonstrates, the first eigenvector contains relatively global, structural features useful for contrasting male and female faces. In O'Toole et al. (1997), the coordinates of faces on the first eigenvector of the PCA accurately predicted the sex of 77% of the faces. Not surprisingly, the model's performance improves when information from other eigenvectors are included in the analysis. Combining across several eigenvectors, the computational model accurately predicted the sex of 97.7% of the faces, a rate comparable to human performance. The effects of adding (left) and subtracting (right) the sixth eigenvector to the average head are shown in the second row of Fig. 1. In this row, the heads are rendered from the profile view to show that nose size, an important aspect of the difference between male and female heads, is captured by this eigenvector. As reported in O'Toole et al. (1993), single eigenvectors are also reliable predictors of the sex of two-dimensional images of faces.

The extent to which children's faces contain similar structural cues that enable reliable discrimination between boys and girls is less certain. In fact, in the

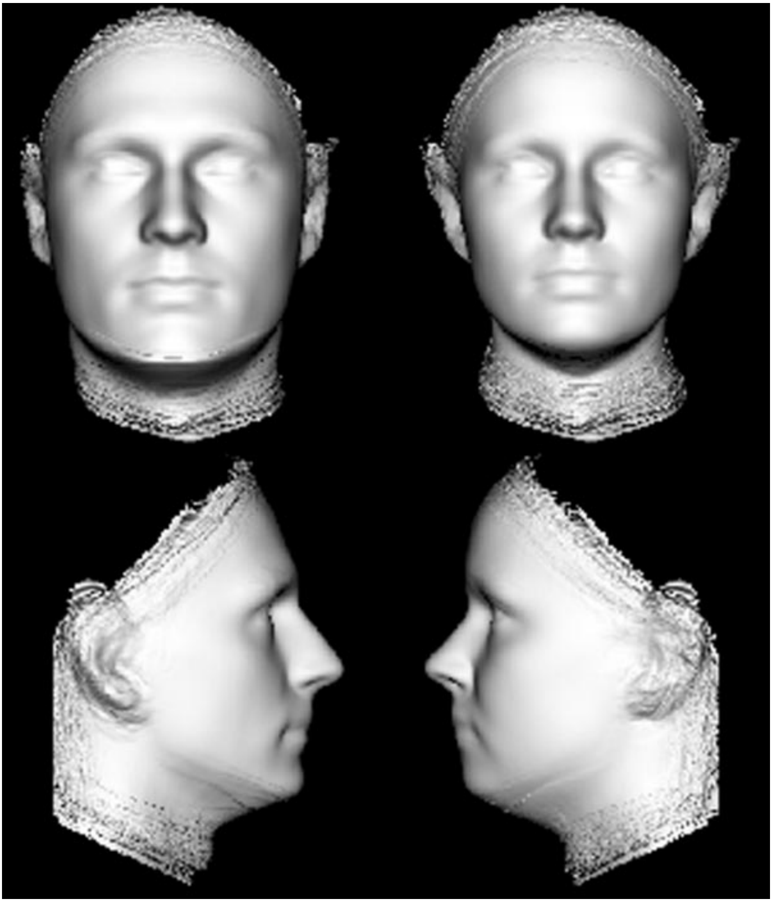

FIG. 1. The first eigenvector (top row) of the three-dimensional head structure added to the average head (left) and subtracted from the average head (right). This eigenvector contains relatively global, structural features useful for contrasting male and female faces. The second row of the figure shows the sixth eigenvector, which was also a statistically significant predictor the sex of the faces, added to (left) versus subtracted from (right) the average head. The heads are rendered from the profile view to show that nose size is an important aspect of the difference between male and female heads.

anthropological literature, Enlow (1982) claims, “The faces of prepubertal boys and girls are essentially comparable” (p. 9). One limitation of these anthropological studies, however, is that they consider only skeletal information. Additional components that contribute to the shape of a face include the muscle, fat, and overlying layers of tissue that can vary in thickness and texture. It is possible, though not certain, that these components provide information that observers can use to determine the sex of a face.

FIG. 2. The two prototypes produced by morphing: left, female prototype face; right, male prototype face.

Is there reliable structural information in children's faces for accurately determining the sex of the face? Before attempting to collect empirical data on sex classification of children's faces, we examined this question more formally by applying the computational strategy used by Yamaguchi et al. (1995). We digitally photographed the faces of 24 male and 24 female children between 7 and 10 years of age and then used a computer image editing program to crop the faces to eliminate hairstyle and clothing cues to the sex of the face. We then created a prototypical boy face and a prototypical girl face by applying the morphing methods used by Yamaguchi et al. (1995). The resultant prototypes appear in Fig. 2. These images seem clearly identifiable as boy and girl.

The first question we addressed in the present study was whether children and adults can reliably determine the sex of children's faces based on internal structural cues alone. To our knowledge, no previous studies have reported sex classification accuracy for children's faces when cultural cues are unavailable. The second question we addressed concerns children's performance on sex classification tasks. To our knowledge, no studies have reported children's accuracy in determining the sex of faces in the absence of social and cultural cues. In the present study first graders, third graders, and adults viewed faces of children and adults and were asked to indicate the sex of each face.

The third question we addressed was whether information about the sex of the face contributes to recognition accuracy. Here we explored whether accurate sex classification of a face would facilitate its later recognition. After observers completed the sex classification task, they viewed a larger set of faces and were asked to indicate which faces they had just classified. This recognition task is important in that it enables us to address a question that has been debated vigorously in the literature on adult face processing in recent years concerning the relationship between categorical and identity information in faces. Faces are unusual objects because they simultaneously support a number of categorical decisions (e.g., sex, race, and age), but must also be encoded and remembered in large numbers as individual exemplars. One might argue that these two types of information are logical complements. To categorize a face (e.g., as male or female), one must be able to extract and encode the information that it shares with all or most faces of the same sex. By contrast, to recognize a face as an individual, we must extract and encode the information that makes the face unique or different from all other faces in the world.

How we process these two types of information in faces has long been an issue for comprehensive models of face processing (cf. Bruce & Young, 1986; Ellis, 1986). Bruce & Young (1986) claim that there are several different types of information codes that are derived from faces. Specifically, they distinguish between a “visually derived semantic code” and “pictorial code.” They refer to the code specifying the sex of a face as a “visually derived semantic code,” a phrase they also apply to other categorical information about the face including race and age. The information in a picture of a face that can be used to make a recognition judgment is referred to as a “pictorial code.” According to their model, the information in faces used for categorization serves a purpose distinct from the information used for recognition. They further speculate that the processing of recognition information and the processing of categorical information proceed in an independent parallel fashion. Bruce (1986) offers some support for this claim by showing that familiarity with a face does not generally affect the time required to classify the sex of the face. However, this result should be interpreted with caution given the ease with which observers can identify the sex of most faces. In fact, Bruce found that familiarity did improve classification performance for a subset of the faces that were less clearly identifiable as “male” or “female.” Bruce speculates that when sex classification is difficult, observers may recruit other stored information about the known person instead of relying solely on the structural characteristics of the face.

A more direct approach to this question would be possible if one were able to selectively manipulate the categorical information in faces and to assess the effects of this manipulation on recognition accuracy. This has been attempted recently using the PCA approach described previously for image data. Specifically, Deffenbacher, Hendrickson, O’Toole, Huff, & Abdi (1998) “reversed” the contrast of eigenvectors that were useful for predicting the sex of a face and effectively “feminized” male faces and “masculinized” female faces. These altered faces were more difficult to classify by sex but were not less memorable than the original faces. Unfortunately, one drawback to this approach is that the synthesized images were rated as less natural looking than the original faces, thus introducing the possibility that the synthesized facial images were not treated in the same manner as the normal faces.

Children’s faces can serve as a “natural” alternative to these synthesized images for the purpose of evaluating the contribution of categorical information to recognition. Specifically, although we suspect that some subtle structural cues to sex are present in children’s faces, this information is likely to be far less salient and reliable than the structural cues to sex present in adult faces. In this sense, children’s faces are analogous to the somewhat androgynous-looking faces that Deffenbacher et al. (1998) created. In contrast to the synthesized faces, however, the children’s faces used in the present study are perfectly natural looking. If the sex of children’s faces is less accurately perceived than the sex of adult faces (as one might predict based on the anthropology), by comparing the recognizability of the children’s and adults’ faces, we can evaluate the extent to which the quality of categorical information in a face contributes to the accuracy with which it is recognized.

In the following experiment, first-graders, third-graders, and adults categorized the faces of children and adults by sex. The faces were cropped digitally to eliminate cultural and social cues to the sex of the faces. Following the categorization task, all observers were asked to recognize the faces they had just categorized by sex from among a larger set of faces.

METHOD

Participants

Sixteen first graders (8 boys and 8 girls) and 15 third graders (7 boys and 8 girls) from a suburban parochial school in eastern Pennsylvania participated in the study. The students were predominantly White and middle- to upper-middle-class. All children completed both the classification and the recognition tasks. The first graders ranged in age from 6 years 10 months to 7 years 11 months, with a mean age of 7 years 4 months. The third graders ranged in age from 8 years 10 months to 9 years 10 months, with a mean age of 9 years 2 months. Twenty undergraduates (10 men and 10 women) from the University of Texas at Dallas (UTD), ages 18 to 29, volunteered to participate in exchange for research credit in an introductory psychology course. The students were predominantly White and middle- to upper-middle-class.

Stimuli

The stimuli consisted of 20 digitized photographs of children’s faces and 20 digitized photographs of adults’ faces. All photographs were of Caucasian faces to avoid “the-other-race effect,” which affects both sex classification efficiency

(O'Toole, Peterson, & Deffenbacher, 1996) and recognition accuracy (e.g., O'Toole, Deffenbacher, Valentin, & Abdi, 1994).[1] The children were between 7 and 10 years of age and the adults were between 20 and 29 years of age. All the photographs were taken under similar conditions. Individuals were asked to remove their glasses and wipe off heavy makeup before the pictures were taken and were asked to display a neutral expression. Even with this instruction, a few faces still displayed some slight variations in expressions. In particular, some faces showed a slight turning up at the corners of the mouth. These faces, however, were equally distributed between males and females.

The digitized faces were 255- by 240- (w × h) pixel images, with full color resolution. The face images were edited using Adobe Photoshop. The pictures were cropped so that the outline of the face remained visible but the neck, shoulders, and some of the hair were eliminated. To eliminate as many of the hair cues as possible, each face was presented within a frame. Sample adult faces[2] are presented in Fig. 3. Although we think these faces are some of best controlled faces ever used in a sex classification experiment on the dimension of eliminating sex-stereotyped cues, we acknowledge that they are not perfect. For example, one remaining cue for some female faces was the grooming of eyebrows. We tried, however, to find faces in which the eyebrows appeared natural. For adult observers, the faces were displayed on a 35.6-cm computer monitor approximately 40 cm from the participant's face. For each picture, the window containing the face was approximately 4.7 cm high and 5.9 cm wide. The frame measured approximately 7.7 cm high and 7.9 cm wide. The faces were placed at approximately the same position in the picture so that a comparable amount of the face was visible for each picture.

A color printer with 360- × 360-dpi resolution was used to print copies of the faces that could be shown to the child observers. After each face was printed, a color photocopier was used to reduce the image so that the window containing the face was 4 cm high and 5 cm wide and the frame was 6.7 cm high and 7.1 cm wide. These images were then mounted on sheets of paper 27.9 cm high and 43.2 cm wide (11 in. × 17 in.). Ten faces appeared on each side of the page. The faces were arranged in three rows, with four faces spaced evenly across both the top and bottom rows. The middle row contained two faces, one centered between the first two faces in the top and bottom rows, and the second centered between the third and fourth faces in the top and bottom rows. After the faces were mounted, the pages were color photocopied and laminated. The arrangement of the males and females on the page was random.

---

[1] Although we have no particular reason to suspect that the results gathered from White participants with White faces would not generalize to other races, we would not wish to advance any claims without empirical data.

[2] We included only example photographs of the adult faces because permission was not requested from parents to publish the children's pictures. However, as noted, the pictures of the children and adults were processed digitally in exactly the same manner.

FIG. 3. Left, an adult male; right, an adult female. Both are in their 20s.

The frames that enclosed the faces that the children saw were very slightly different from the frames that were used with adults. For children, a small 2-cm × 2.8-cm white rectangle that contained a schematic image of a man and woman standing side by side was positioned along the top edge of the frame so that it extended down into the frame area. These male and female figures were included so that the children could indicate the sex of the face. This same frame appeared with all the pictures in both the classification and recognition tasks.

Procedure

Classification task. The children performed both the classification and the recognition tasks in a group setting within their classrooms. Each child was given one laminated sheet containing the faces and a page of 24 small yellow circular stickers. Ten adult faces (5 males and 5 females) appeared on one side of the page and 10 child faces (5 males and 5 females) appeared on the other side of the page. After the papers were handed out, the experimenter explained that the children were to do their best to decide if each person on the page was a boy or a girl. The children were instructed to mark each person as a boy or a girl by placing a sticker on the appropriate symbol in the top portion of the picture frame. Not surprisingly, all the children were familiar with these symbols. At this point, the children were asked to point to the symbol they would put a sticker on if they thought the face belonged to a boy. The children were then asked to point to the symbol they would put a sticker on if they thought the face belonged to a girl. None of the children had any trouble following these instructions. Children were told that they needed to put a sticker on all the pictures before they could turn the sheet over and do the other side. Each child worked at his or her own pace and the sheets were collected when all children had finished. Approximately half of the children judged the adult faces first. Two sets of 20 faces were generated from the original 40 faces. Approximately half of the children saw each set of faces. The order of the faces on each page varied across participants.

The adult participants read instructions indicating that they would be asked to classify the faces of adults and children as either male or female. For each trial, a single face appeared on the screen and remained visible until the participants responded by pressing the appropriately labeled key on the keyboard. Each observer saw 10 adult faces (5 males and 5 females) and 10 child faces (5 males and 5 females). Two different sets of adult and child faces were used. For each stimulus set, half of the observers saw adult faces first. Within the blocks of adult and child faces, the presentation order was individually randomized. All the experimental tasks that were administered to the adult participants were programmed on a Macintosh computer using PsyScope (Cohen, McWhinney, Flatt, & Provost, 1993).

Recognition task. The sheets for the recognition task and another set of stickers were handed out to the children after the sheets for the classification task had been collected. Each child was given two laminated sheets, one containing the complete set of 20 child faces and the other containing the complete set of 20 adult faces. On each side of the page, half of the faces had appeared in the classification materials the child had been given. Children were told that some of the faces on these new pages were the same as the ones they had just seen and that some of the faces were new ones they had not seen before. The children were instructed to place a sticker on all the faces that they remembered seeing on the earlier sheet. No information about the number of previously seen faces on the recognition sheets was provided to the children.

The adult participants completed the recognition task immediately after the gender classification task for each face age group, according to the counterbalance scheme described previously. Each face appeared individually on the screen and the participant was asked to indicate whether the faces had appeared previously by pressing the appropriately labeled key to indicate whether the face was “old” or “new.” The faces in both the child and the adult recognition tasks were presented in a different random order for each participant.

RESULTS

Classification

Sex classification accuracy was measured with A', a response-bias-free measure of discrimination accuracy based on signal detection theory (Green & Swets, 1966). In the context of a sex classification experiment, response bias refers to the tendency of participants to guess “male” more than “female” when they are unsure. We used A' to control for this kind of response bias, which can confound other kinds of measures such as percentage correct. Intons-Peterson (1988) used A' in her sex classification experiments also, and indeed found that children were biased to respond “male” with a variety of face and body stimuli. As we shall see shortly, in the present study we found a similar bias for both children and adults responding “male” substantially more often than “female.” An A' is computed based on “hits,” defined here as the response “female” to female faces, and “false alarms,” defined here as the response “female” to male faces. In this way, it yields a measure of sex classification accuracy that controls for biased guessing strategies.[3]

A' was computed for each observer in each face age condition. These data were submitted to a two-factor analysis of variance (ANOVA) with the age of the participant as a between-subjects variable and the age of the face as a within-subjects variable. As expected, we found a significant effect of the age of the participant, $F(2, 48) = 9.8$, $p < .001$, with performance generally improving with age (see Fig. 4). The results also indicated a significant main effect of face age, $F(1, 48) = 18.3$, $p < .001$, with accuracy better for adults’ versus children’s faces. No interaction between the age of the participant and the age of the face was found, $F < 1$.

It should be noted that chance performance for A' is equal to a score of .50, with a maximum score equal to 1. Thus, it can be seen, using the standard error bars in Fig. 4, that performance was above chance for all conditions, with the exception of the first graders judging the sex of the children’s faces. At the opposite extreme, performance was at ceiling only for the adult observers judging the sex of the adult faces.

Following Intons-Peterson (1988), we further applied signal detection theory methodology to examine possible biases for guessing male versus female. We did this by computing a criterion or C score[4] for each observer on both the children’s and the adults’ faces. It should be noted that positive numbers indicate a bias to respond “male,” negative numbers indicate a bias to respond “female,” and zero indicates no bias. The average criterion for each condition is displayed in Fig. 5. As can be seen, in all cases indicating a bias, the direction of the bias revealed

---

[3] A' is a nonparametric version of the better known classic signal detection measure of d'. We chose A' due to the nonlinear behavior of d' as performance reaches ceiling, which occurs here for adult sex classification of the adult faces. A' is defined as follows:

if hits (H) $>$ false alarms (F), then
$A' = 0.5 + (H-F)\,(1+H-F)/[4H\,(1-F)]$

if hits (H) $<$ false alarms (F), then
$\mathrm{A}' = 0.5 + (F-H)\,(1+F-H)/[4F\,(1-H)]$

[4] C is defined as $-.5(z_{\text{hit rate}} + z_{\text{false alarm rate}})$. We used C, rather than the more classic signal detection theory measure of d', for reasons explained in Macmillan & Creelman (1991).

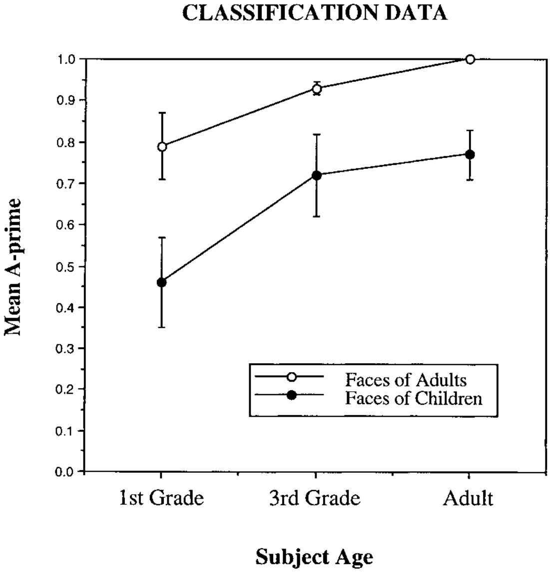

FIG. 4. Mean *A′* performance for classifying children's faces versus adults' faces by sex for all participant age groups. Vertical bars indicate standard error of the mean.

a tendency to respond "male" more frequently than "female." This finding was remarkably consistent. Across all age groups, all but one participant showed either a male bias or no bias at all. Typically, these later "no-bias" cases were adults classifying adult faces, for which performance was perfect (i.e., no bias possible). The standard error bars also indicate that the tendency to guess "male/boy" is statistically greater than zero for all conditions, with the exception of the adult participants classifying the adult faces.

More formally, the criterion data were submitted to a two-factor ANOVA with the age of the face as a within-subjects variable and the age of the participant as a between-subjects variable. The results indicated a significant main effect of the age of the face, $F(1, 48) = 39.5$, $p < .001$, indicating that the children's faces elicited stronger response biases than did the adults' faces (see Fig. 5). We also found an effect of the participant's age, $F(2, 48) = 15.91$, $p < .001$, with younger observers displaying stronger biases than adults (see Fig. 5). This latter difference, however, might be due to the fact that adults performed close to ceiling in the adult face classification task, which leaves little opportunity for the criterion to vary substantially. The age of the participant and the age of the face did not interact, $F < 1$. There was no indication of an effect of participant sex on the size of this bias, $F < 1$. We tested this effect only for the children's faces, because of the near perfect performance of the adults on the adult faces. Further, we were able to test this effect only for the adult participants, as we did not retain the sex of the individual children with the data files.

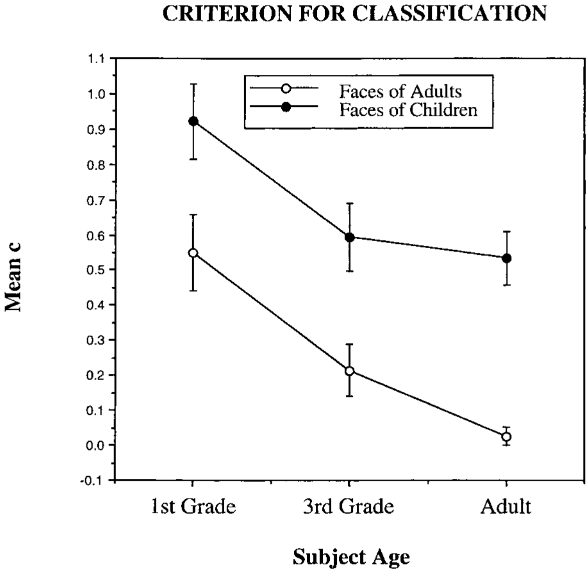

FIG. 5. Criterion for performance on classification tasks for children's and adults' faces for all participant age groups. Vertical bars indicate standard error of the mean.

In summary, not surprisingly, the adult faces were classified more accurately by sex than were the children's faces. All participants categorized the adult faces accurately at performance levels varying from just above chance (first graders) to nearly perfect (adults). The children's faces were classified accurately by the adults and the third graders, but not by the first graders, whose performance on these faces was at chance. In no case was the performance on classifying children's faces at ceiling. This indicates that although children's faces are more difficult to classify by sex than adult faces, they are nonetheless classifiable at levels above chance, even in the absence of social and cultural cues to sex.

Recognition

Recognition accuracy was also measured by computing A'. In the context of a face recognition experiment, A' is a response-bias-free measure for discriminating between "studied" and "novel" faces. A' is again defined as a function of hits and false alarms, as described previously. In this case, however, a hit is defined as an "old" response to a studied face and a false alarm is defined as an "old" response to a novel face. The average A' data appear in Fig. 6 and indicate that recognition performance is comparable for children's and adults' faces. More formally, the A' data were submitted to a two-factor ANOVA with the age of the face as a within-subjects variable and the age of the participant as a

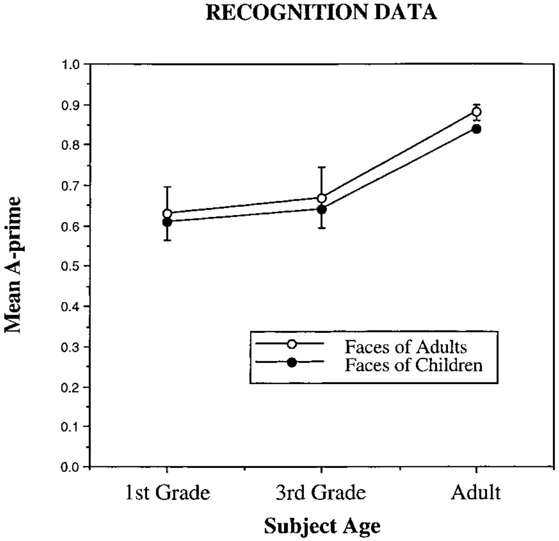

FIG. 6. Mean A' performance for recognizing children's faces versus adults' faces for all participant age groups. Vertical bars indicate standard error of the mean.

between-subjects variable. A significant effect of participant age was found, $F(2, 48) = 8.46$, $p < .001$, with adult participants more accurate than the children. No effect of age of face was found, $F < 1$, and there was no interaction between age of face and age of participant, $F < 1$.

Although no bias was expected, for consistency with the sex classification experiment we analyzed the *C* scores as well. In the context of a recognition experiment, response bias refers to the tendency to allocate responses to the categories of "old" versus "new." We submitted the *C* data to a two-factor ANOVA and found no significant effects, though we noted a nonsignificant trend for the participants to be more conservative, that is, respond "new" more frequently than "old," with the children's faces, $F(1, 48) = 3.6$, $p = .07$.

In summary, the results of the recognition task indicated that performance was comparable for children's and adults' faces across all participant groups. Tentatively, in light of the results of the classification experiment, the recognition results suggest that categorical information relevant for sex classification does not contribute to recognition accuracy. At a more detailed level, because the theoretical issues are usually framed at the level of individual faces, it is also informative to examine the relationship between classification and recognition at this level. Specifically, does the quality of the categorical information in a particular face affect its recognizability? Data from the adult subjects for the individual faces were easily available in the computer files to address this question. We thus computed A's for the sex classification and recognition tasks for each picture individually. Consistent with the subject-based analysis, the correlation between sex classification and recognition accuracy for the children's faces was low ($r = .09$, $p = .67$). We did not carry out similar analyses for the adult faces because adult performance on the sex classification task was at ceiling.

Before we continue to examine the implications of the combined findings of the classification and recognition tasks, we must explore two alternative explanations of our results, which we addressed in two control experiments.

Control Experiment 1

An alternative possibility to the conclusions just advanced concerning the relationship between categorical and identity information in faces might be argued as follows. As noted previously, because we were unsure how well children would perform on these tasks, we allowed the children unlimited time to respond. Rather than vary exposure time with age, we also opted to give adults unlimited time to respond. Because the children's faces were more difficult to classify by sex than were the adult faces, it is possible that our observers may have "studied" the children's faces longer than the adult faces. Thus, it is conceivable that the adult faces were more recognizable than the children's faces, but that the extra study time for the children's faces might have improved recognition performance to levels comparable with the adult faces. In fact, the computer program controlling the adult experiment recorded response times and so we were able to confirm, post hoc, that there was indeed a difference in response time for judging the sex of children's versus adults' faces. The adult participants spent somewhat longer looking at children's faces (mean = 1889 ms) than at adults' faces (mean = 1515 ms). This difference approached statistical significance, $F(1, 20) = 3.33$, $p = .08$. Due to the way the task was administered to children, it was not possible to measure response times for the children.

To control for the possible confounding effect of longer study time on recognition performance for the children's faces, we replicated the experiment again with adult participants. However, we limited the face presentation time in the sex classification experiment to 1500 ms. We chose this time because it approximates the mean amount of time adult participants spent looking at the adults' faces in the first sex classification experiment. Further, it is worth noting that this new exposure time is nearly 400 ms shorter than the mean amount of time adults spent looking at the children's faces in the previous experiment. Therefore, if the exposure time differences that were present in the first experiment had contributed to equalizing recognition performance for the children's and adults' faces, limiting exposure time here should result in a decrease in performance on the recognition task for children's faces.

For this experiment, our participants consisted of 24 UTD students ages 20–29, recruited as described for the first experiment. The control experiment was counterbalanced as described previously and the procedure was identical to the previous one in all respects other than the exposure time limitation. The results of this control experiment yielded mean recognition performance scores for the children's versus adults' faces that were virtually identical to those found previously. For the adult faces, $A' = .88$ ($SE = .016$) and for the children's faces, $A' = .84$ ($SE = .040$). For the initial experiment these scores were $A' = .87$ ($SE = .018$) for the adults' faces and $A' = .84$ ($SE = .047$) for the children's faces. More formally, recognition accuracy did not differ as a function of face age when the exposure time at study was controlled, $F < 1$. The classification data replicated the previous finding. Children's faces were classified less accurately than adult faces, $F(1, 46) = 18.5$, $p < .001$.

Thus, our initial conclusions hold. For the children's faces, only adults and third graders were able to accurately perceive the differences in boys' and girls' faces. Further, it is clear that children's faces were more difficult to classify by sex, but were not more difficult to recognize.

Control Experiment 2

Another concern was that there were two methodological differences between the procedures used with adults and children. The first was the difference between the use of laminated sheets with the children and the use of computer presentation of the faces with adults. This enabled the children, but not the adults, to make simultaneous comparisons among the faces. The second difference was the delay between learning and testing for the children versus adults. The delay was slightly longer for the children than for the adults, due to differences in the way we blocked the subtasks for the adults and children. Thus, we devised a second control experiment in which the adults performed the laminated sheet tasks used previously with the children. Sixteen undergraduate students (2 males and 14 females) from UTD volunteered to participate in this experiment. The participants classified and recognized the faces in the same manner as the children in the initial experiment (i.e., one laminated sheet of faces for classification and two laminated sheets of faces for recognition). This experiment was also counterbalanced as described previously. The only difference between the procedures used with children and adults was that the adults were tested individually, whereas the children were tested in a classroom situation.

The results of this control experiment replicated those found in the initial experiment for the adult subjects. For the sex classification task, the participants were more accurate for the adult faces ($A' = .99$) than for the children's faces ($A' = .77$). This compares closely to the performance obtained in the initial experiment ($A' = 1.0$ for the adult faces and $A' = .77$ for the children faces). For the recognition task, there was no difference in recognition performance for the adult ($A' = .89$) versus children's ($A' = .85$) faces. This performance is comparable to the performance seen in the initial experiment ($A' = .87$ for the adult's faces and $A' = .84$ for the children faces). Thus, our initial findings cannot be due to the small differences between the tasks and stimuli for the adults and children in Experiment 1.

Combined Categorization and Recognition Results Summary

The first question we addressed was whether children and adults can reliably determine the sex of children's faces based on internal structural cues alone. Our results clearly indicate that both adults and third graders were able to discriminate the sex of children's faces. First graders, however, could not perform this task accurately. Additionally, consistent with previous work (Intons-Peterson, 1988), we found a response bias to guess "male" for all conditions, with the exception of the adult participants judging the sex of the adult faces.

The second question we addressed concerns children's accuracy in determining the sex of faces in the absence of social and cultural cues. We demonstrated that by about the first grade, children can accurately discriminate the sex of adult faces without sex-stereotyped cues. We showed also that by about third grade, children can categorize both adults' and children's faces on the basis of sex.

The third question we addressed was whether success at classifying the sex of the face contributes to recognition accuracy. Our results indicate that it does not. Performance on sex categorization was more accurate for adults' versus children's faces. Performance on recognition, however, did not differ as a function of the age of the face. Thus, even though the sex of adult faces is more salient than the sex of children's faces, the salience of this category information in faces did not seem to improve recognition.

DISCUSSION

The findings of the present study have several implications for understanding broader issues in the development of expertise on perceptual categorization and recognition tasks. Faces provide a highly meaningful and challenging stimulus with which to examine children's abilities. As a basic-level category of objects (Rosch & Mervis, 1975), faces consist of a set of features (eyes, nose, mouth, etc.) arranged in a particular, universally recognizable configuration. Subcategories of faces, such as those belonging to males and females, can be distinguished further as specific, gross variations on this configural theme. Still further, individual faces can be distinguished as variations about their local category prototypes. To categorize faces by sex we must learn to attend to the differences between male and female faces. To recognize an individual face, we must learn to encode specific information about the face that helps us to distinguish it from all other faces we know, and what is perhaps more difficult, from the enormous set of faces we do not know.

The present study provides much-needed empirical data concerning children's skills on the complex perceptual discrimination task of classifying faces by sex. Without social or cultural cues to sex, it is very difficult to list a set of clear-cut distinguishing features for male and female faces. And yet, by the time we are adults, we can perform this task effortlessly and with remarkable speed (with less than 75-ms exposure times; O'Toole, Peterson & Deffenbacher, 1996). It is harder still to generate a set of clear-cut feature labels for distinguishing the faces of boys and girls. The few cues that both are relevant for classifying adult faces by sex and are accessible to linguistic labeling (e.g., nose size, brow protrusion) are simply less useful for classifying children's faces by sex. Nonetheless, the present data indicate that adults and third graders are able to perform this task.

The use of faces that are devoid of cultural and social cues enables us to focus on the acquisition of information that is not easily conveyed in nonperceptual terms. This kind of information contrasts with information that is easy to convey through a single featural label, like "Ladies have *long hair*." The elusiveness of labels for describing the structural differences between male and female faces, and boy and girl faces, is typical of other stimuli for which perceptual learning is invoked as the underlying acquisition mechanism (e.g., categorizing music by composer, Mozart versus Bach, soft/round versus hard/pointed; wine by grape, pinot noir versus gamay, woody versus fruity; see O'Toole et al., 1995, for a discussion of the applicability of perceptual learning to face processing). Thus, it is not surprising that perceptual learning has been proposed as a possible mechanism underlying the acquisition of expertise for faces (O'Toole et al., 1995). Further evidence for the relevance of perceptual learning as a general theoretical construct for understanding human expertise with faces comes from a recent study by Gauthier & Tarr (1997). They used a computer algorithm to generate "greebles," a synthetic class of objects that parallel faces in their complexity, structural uniformity, and statistical diversity. Gauthier and Tarr found that many of the effects typical of face processing can be obtained with greebles, if observers are allowed to accumulate sufficient expertise with them.

The role of perceptual learning in face processing relates also to the question of the relative importance of feature versus configural information for classifying and recognizing faces. Empirical data gathered in the past two decades converge on the widely accepted tenet that configural information is used more predominantly for recognizing faces than for recognizing other objects (see Farah, Wilson, Drain, & Tanaka, 1998, for an excellent review). Carey & Diamond (1977) argue that a marker characteristic of human expertise with faces is the transition from a piecemeal to a configural representation of faces, which they suggest occurs between 6 and 10 years of age. Carey & Diamond make this claim based on differences between the way children and adults process upright versus inverted faces. Face inversion is thought to selectively disrupt the processing of the configural information in faces (Bartlett & Searcy, 1993). Carey & Diamond found that although face recognition performance for adults is worse for inverted versus upright faces (see also Yin, 1969), 6-year-old children performed equally well for inverted and upright faces. Performance on upright faces gradually increased until the age of 10, whereas performance on inverted faces leveled off. Carey & Diamond (1977) suggest that young children below age 10 attend to the featural ("piecemeal") information in faces as opposed to the configural information. Thus, it may be possible that perceptual learning is required as a mechanism for refining our ability to attend to and encode the configural information relevant for the different tasks.

An important aspect of the problem we consider in this paper concerns the nature of the information useful for classifying the sex of children's versus adults' faces. Both the psychological and anthropological literatures support the idea that the sex of adult faces can be determined reliably, although not perfectly, from single or small subsets of features (e.g., Burton et al., 1993; Chronicle et al., 1995; Enlow, 1982). This feature-based information complements the information available at the configural level. The computational analysis with PCA is consistent with the psychological and anthropological evidence. As we noted in Fig. 1, of the two eigenvectors that were most predictive of the sex of faces, one captured relatively global structural differences between male and female faces, whereas the other eigenvector seemed to capture relatively local differences in the size of the nose.

Prepubescent faces are almost certainly not differentiable by sex on the basis of single localized features (Enlow, 1982). The fact that adults and 9-year-olds are able to reliably classify children's faces by sex indicates that they are probably making use of configural differences in boys' and girls' faces. This links our data to the recognition data of Carey & Diamond (1977) because it suggests that our 7-year-old participants might have been unable to classify the children's faces by sex because of the limited availability of feature-based sex cues for the children's faces. Given that these young children were able to classify the adult faces reliably by sex, it is interesting to wonder if they relied more on the local feature cues than did the adults and 9-year-olds. This might be determined in future studies by comparing sex classification performance for children and adults with inverted and upright faces. Here we might suppose that the 7-year-olds should perform more comparably on the upright and inverted faces than the adults or 9-year-olds.

The present data speak also to the relationship between the categorical and identity information in faces. Although this issue has been considered primarily in the context of fully developed face processing skills (i.e., adult perceivers), it is interesting also to consider the development of these skills. Children must learn how to make accurate categorical judgments about faces (i.e., male versus female) and must also learn how to accurately discriminate "known" and "unknown" individuals (i.e., Have I seen this person before or is he/she a "stranger"?). As noted previously, performing these tasks accurately requires an ability to "pick and choose" among the relevant information for a given task from the rich perceptual information available. The question for comprehensive theories of face processing concerns the extent to which categorical and identity information are processed independently. By looking at performance on a combined sex categorization and face recognition task for children's and adults' faces (wherein the quality of the sex categorization information differs naturally), we were able to show that recognition accuracy is not related to the accuracy of categorizing faces by sex. The ability of human observers to process the sex categorization and identity information independently is consistent with models such as PCA, which separate the categorical and identity information naturally into eigenvectors explaining larger versus smaller amounts of variance, respectively (O'Toole et al., 1993).

As noted previously, in the real world, diverse and redundant cues provide information about the sex of an individual and many of these cues are culturally based. Moreover, previous studies suggest that young children do use culturally based cues, especially hairstyle, when deciding someone's sex (Intons-Peterson, 1988). In the gender constancy task, when a salient gender cue such as hair or clothing is changed, many preoperational children believe that the individual's sex has also changed. Different explanations have been offered to account for this misconception. Researchers such as Kohlberg (1966) and Marcus & Overton (1978) note parallels between gender constancy and conservation; both hinge on understanding that an underlying reality has been preserved despite salient perceptual changes. For Piagetians, success on the gender constancy task depends on operational structures that allow one to grasp the reversibility of the transformations. Others have countered that, like many Piagetian tasks, the gender constancy task places superfluous demands on the child and may also be problematic linguistically (Johnson & Ames, 1994).

To varying degrees, both Carey (1985) and Bem (1989) attribute failure on the gender constancy task to inadequacies in the child's domain-specific knowledge base. For Bem, children's lack of genital knowledge is the chief culprit. In her study, only those 3- to 5-year-olds who passed her genital knowledge test evidenced gender constancy. Carey makes a similar argument, although she does not focus specifically on genital knowledge. In her view, preschoolers either lack or have only an impoverished naïve theory of biology. Consequently, they view gender as primarily a social phenomenon and it is this reliance on a social theory of gender that leads to failure on gender constancy tasks.

The fact that the 6-year-olds in our study were sensitive to structural differences in adult male and adult female faces raises the possibility that perceptual factors may facilitate the development of gender constancy. Changes in hair and clothing do not alter the structure of the face. Thus, although young children are likely to rely on the cultural cues when they are available, in the absence of these cues children may rely on biologically based facial cues to gender. Whether the use of biologically based facial cues to gender is a by-product of gender constancy or whether it precedes, and possibly facilitates, the development of gender constancy is an open question that is worth pursuing.

Additionally, although it is not central to the issues we have addressed here, we were nonetheless intrigued by the highly consistent bias to guess "male" in the sex classification task. Recall that this bias was evident in all but one case—the adult participants classifying the sex of the adult faces. Our results indicated an especially strong bias on the part of young children, who may be using an explicit guessing strategy when unsure. The results of the classification task further indicate relatively larger criterion effects when sex classification performance is poor. This bias would seem to be unrelated to the presence/absence of cultural cues to sex given that it has been found also by Intons-Peterson (1988) in experiments using stimuli that contain such cues. We might speculate from a social perspective, therefore, that there may be a "higher price" attached to mistaking a male for a female than for making the inverse error. As suggested by Intons-Peterson (1988) in a study with child participants, it is possible also that faces without distinctively female hairstyles may be called "boys" by default. Our replication of the male response bias with adult participants judging children's faces, however, weakens this interpretation. Adults clearly understand that the digitally imposed absence of the hair has no bearing on the sex of the child pictured, and yet they are also subject to male response bias. Nonetheless, it may be possible that even in understanding this principle, the lack of hair may be difficult to ignore perceptually. The reason for this bias is probably quite complex and may have a social component. In any case, this question will require further study before a satisfactory explanation can be offered.

In summary, the sex of children's faces was perceived less accurately than the sex of adult faces for all participants. The fact that adults perceived the sex of children's faces with reasonable accuracy indicates that there are useful facial structure differences for differentiating boys and girls. Young children, however, were unable to perceive these differences, although they were able to perceive facial structure differences between male and female adult faces. Our data suggest that young children are still learning how to detect and make use of the subtle facial structure differences between boys and girls. In light of the gender constancy literature, our data suggest also that children may master the use of sex-stereotyped social and cultural cues to gender earlier than they master the use of facial structure cues to sex. Finally, we found that the ability to categorize faces by sex had no evident effect on the accuracy with which the faces were recognized. Children's faces were categorized much less accurately than adults' faces, but were recognized equally accurately. This suggests a degree of independence between the tasks of categorizing and recognizing faces.

REFERENCES

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology,* **25,** 281–290.

Bem, S. L. (1989). Genital knowledge and gender constancy in preschool children. *Child Development,* **60,** 649–662.

Bruce, V. (1986). Influences of familiarity on the processing of faces. *Perception,* **15,** 387–397.

Bruce, V., Burton, A. M., Hanna, E., Healey, P., & Mason, O. (1993). Sex discrimination: How do we tell the difference between male and female faces? *Perception,* **22,** 131–152.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology,* **77,** 305–327.

Burton, A. M., Bruce, V., & Dench, N. (1993). What's the difference between men and women?: Evidence from facial measurement. *Perception,* **22,** 153–176.

Carey, S. (1985). *Conceptual change in childhood.* Cambrige, MA: MIT Press.

Chronicle, E. P., Chan, M., Hawkings, C., Mason, K., Smethurst, K., Stallybrass, K., Westrope, K., & Wright, K. (1995). You can tell by the nose—Judging sex from an isolated facial feature. *Perception,* **24,** 969–973.

Cohen, J. D., McWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments & Computers,* **25,** 257–271.

Cornell, E. H. (1974). Infants' discrimination of photographs of faces following redundant presentations. *Journal of Experimental Child Psychology,* **18,** 98–106.

Deffenbacher, K. A., Hendrickson, C., O'Toole, A. J., Huff, D. P., & Abdi, H. (1998). Manipulating face sex: Effects on categorization and recognition judgments. *Journal of Biological Systems,* **6,** 219–239.

Ellis, H. D. (1986a). Disorders of face recognition. In K. Poeck, H. J. Freund, & H. Ganshirt (Eds.), *Neurology.* Berlin: Springer.

Ellis, H. D. (1986b). Processes underlying face recognition. In R. Bruyer (Ed.), *Neuropsychology of face perception.* Hillsdale, NJ: Erlbaum.

Emmerich, W., Goldman, K. S., Kirsch, B., & Sharabany, R. (1977). Evidence for a transitional phase in the development of gender constancy. *Child Development,* **60,** 930–936.

Enlow, D. H. (1982). *Handbook of facial growth.* Philadelphia: Saunders.

Fagan, J. H., III (1976). Infant's recognition of invariant features of faces. *Child Development,* **47,** 627–638.

Fagot, B. I. (1995). Psychosocial and cognitive determinants of early sex-role development. *Annual Review of Sex Research,* **6,** 1–31.

Fagot, B. I., & Leinbach, M. D. (1993). Sex-role development in young children: From discrimination to labeling. *Developmental Review,* **13,** 205–224.

Fagot, B. I., Leinbach, M. D., & Hagan, R. (1986). Sex labeling and the adoption of sex-typed behaviors. *Developmental Psychology,* **22,** 440–443.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review,* **105,** 482–498.

Gauthier, I., & Tarr, M. J. (1997). Becoming a greeble expert: Exploring mechanisms for face recognition. *Vision Research,* **37,** 1673–1682.

Green, D. N., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Gross, T. F. (1997). Children's perception of faces of varied immaturity. *Journal of Experimental Child Psychology,* **66,** 42–63.

Hort, B. E., Leinbach, M. D., & Fagot, B. I. (1991). Is there coherence among the cognitive components of sex acquisition? *Sex Roles,* **24,** 195–207.

Huston, A. C. (1983). Sex typing. In P. H. Mussen & E. M. Hetherington (Eds.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social behavior.* New York: Wiley.

Intons-Peterson, M. J. (1988). *Children's concepts of gender.* Norwood, NJ: Ablex.

Johnson, A., & Ames, E. (1994). The influence of gender labelling on preschoolers' gender constancy judgements. *British Journal of Developmental Psychology,* **12,** 241–249.

Kohlberg, L. (1966). A cognitive-developmental analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences.* Stanford, CA: Stanford Univ. Press.

Leinbach, M. D., & Fagot, B. I. (1986). Acquisition of sex labels: A test for toddlers. *Sex Roles,* **15,** 655–666.

Leinbach, M. D., & Fagot, B. I. (1993). Categorical habituation to male and female faces: Sex schematic processing in infancy. *Infant Behavior and Development,* **16,** 317–332.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, UK: Cambridge Univ. Press.

Marcus, D. E., & Overton, W. F. (1978). The development of cognitive gender constancy and sex role preferences. *Child Development,* **49,** 434–444.

Miklashevskaya, N. N. (1962). Sex differences in growth of the head and face in children and adolescents. *Human Biology,* **41,** 250–262.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., & Valentin, D. (1993). Low dimensional representation of faces in high dimensions of the space. *Journal of the Optical Society of America A,* **10,** 405–410.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., & Valentin, D. (1995). A perceptual learning theory of the information in faces. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition.* London: Routledge.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of faces and the other-race effect. *Memory and Cognition,* **2,** 208–224.

O'Toole, A. J., Peterson, J., & Deffenbacher, K. A. (1996). An "other-race effect" for classifying faces by gender. *Perception,* **25,** 669–676.

O'Toole, A. J., Vetter, T., Troje, N. F., & Bulthoff, H. H. (1997). Sex classification is better with three-dimensional head structure than with image intensity information. *Perception,* **26,** 75–84.

O'Toole, A. J., Vetter, T., Volz, H., & Salter, E. M. (1997). Three-dimensional caricatures of human heads: Distinctiveness and the perception of facial age. *Perception,* **26,** 719–732.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structures of categories. *Cognitive Psychology,* **8,** 382–439.

Slaby, R. G., & Frey, K. S. (1975). Development of gender constancy and selective attention to same-sex models. *Child Development,* **46,** 849–856.

Strauss, M. S. (1979). Abstraction of prototypical information by adults and 10-month-old infants. *Journal of Experimental Psychology: Human Learning and Memory,* **5,** 618–632.

Weinraub, M., Clemens, L. P., Sockloff, A., Etheridge, T., Gracely, E., & Myers, B. (1984). The development of sex role stereotypes in the third year: Relationships to sex labeling, sex identity, sex-typed toy preference, and family characteristics. *Child Development,* **55,** 1493–1503.

Yamaguchi, M. K., Hirukawa, T., & Kanazawa, S. (1995). Judgment of sex through facial parts. *Perception,* **24,** 563–575.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology,* **81,** 141–145.

Received February 5, 1999; revised November 2, 1999