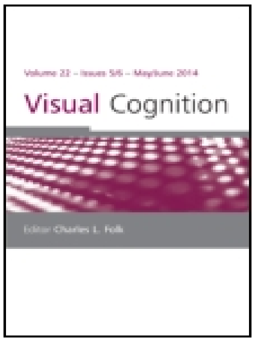

Visual Cognition

ISSN: 1350-6285 (Print) 1464-0716 (Online) Journal homepage: https://www.tandfonline.com/loi/pvis20

Categorical perception of sex occurs in familiar but not unfamiliar faces

Isabelle Bülthoff & Fiona Newell

To cite this article: Isabelle Bülthoff & Fiona Newell (2004) Categorical perception of sex occurs in familiar but not unfamiliar faces, Visual Cognition, 11:7, 823-855, DOI: 10.1080/13506280444000012

To link to this article: https://doi.org/10.1080/13506280444000012

Published online: 02 Aug 2010.

VISUAL COGNITION, 2004, *11* (7), 823–855

Categorical perception of sex occurs in familiar but not unfamiliar faces

Isabelle Bülthoff

Max-Planck Institute for Biological Cybernetics, Germany

Fiona N. Newell

Department of Psychology, Trinity College, Dublin, Ireland

We investigated whether male and female faces are discrete categories at the perceptual level and whether familiarization plays a role in the categorical perception of sex. We created artificial sex continua between male and female faces using a 3-D morphing algorithm and used classical categorization and discrimination tasks to investigate categorical perception of sex. In Experiments 1 and 2, 3-D morphs were computed between individual male and female faces. In Experiments 3 and 4, we used face continua in which only the sex of the facial features changed, while the identity characteristics of the facial features remained constant. When the faces were unfamiliar (Experiments 1 and 3), we failed to find evidence for categorical perception of sex. In Experiments 2 and 4, we familiarized participants with the individual face images by instructing participants to learn the names of the individuals in the endpoint face images (Experiment 2) or to classify face images along a continuum as male or female using a feedback procedure (Experiment 4). In both these experiments we found evidence for a categorical effect for sex after familiarization. Our findings suggest that despite the importance of face perception in our everyday world, sex information present in faces is not naturally perceived categorically. Categorical perception of sex was only found after training with the face stimulus set. Our findings have implications for functional models of face processing which suggest two independent processing routes, one for facial expression and one for identity: We propose that sex perception is closely linked with the processing of facial identity.

Please address all correspondence to Isabelle Bülthoff, Max-Planck-Institut für biologische Kybernetik, Spemannstrasse 38, D 72076 Tübingen, Germany.
Email: isabelle.buelthoff@tuebingen.mpg.de

We are grateful to Volker Blanz and Thomas Vetter for their help with the face stimuli and the morphing technique and to Ian M. Thornton and Heinrich H. Bülthoff for insightful comments on earlier drafts of this manuscript. FN was supported by the Max Planck Society and Trinity College Dublin Research Fund.

http://www.tandf.co.uk/journals/pp/13506285.html DOI: 10.1080/13506280444000012

Usually we have little trouble in determining the sex of a person that we meet for the first time. Nevertheless, most of us have experienced the complexity of this seemingly easy task when confronted by a person in baggy clothing, with short hair and no makeup, leading us to scrutinize his/her face, voice, body shape, and body language for any telltale signs of his/her sex. Here we were concerned with how we perform this task when the only information available is static visual information of an adult face. More specifically, we examined if unfamiliar faces are readily perceived as two discrete categories, one female and the other male, or whether familiarization of a face is required for sex categorization.

In our visual world we are confronted with a multitude of object shapes, often from novel viewpoints and sometimes moving and changing. Given this variation, a perceptual system needs to solve the problem of object constancy. That is, variations in object images should be recognized as being from the same object source, whereas images of another similar object should be recognized as unique. One strategy to simplify the mental representation of the external world is to organize all stimuli into discrete categories (e.g., Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). While some categories are physically very different from each other (e.g., cars and insects), others are very similar (e.g., happy and sad faces). Many studies have found that the visual system has a clever way of discriminating between items from highly similar categories in our world by exaggerating small perceptual differences between similar items belonging to different categories, thus creating clear boundaries between groups of items. This phenomenon is called categorical perception (Harnad, 1987): Objects within a category are perceived as more similar to each other than to objects belonging to another category even if the physical differences between them are equal.

Many instances of categorical perception (CP) have been described in the literature. One well-known example of categorical perception is our perception of a rainbow. Although a rainbow constitutes a monotonic increase of wavelength of visible light, we perceive a series of discrete colour bands. Furthermore, a wavelength in the yellow band of the rainbow is perceived as qualitatively more similar to another yellow wavelength than to a wavelength of the same physical distance but belonging to the orange band of the rainbow.

Initially, CP effects were thought to be a result of innate or overlearned perceptual categories. For example, CP was first observed with auditory stimuli such as phonemes (Eimas, Miller, & Jusczyk, 1987; Liberman, Harris, Hoffman, & Griffith, 1957) or musical sounds (Burns & Ward, 1978) but was not found for novel auditory stimuli (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). CP effects were also found for colour (Bornstein, 1987; Bornstein & Korda, 1984; Boynton, 1979; de Valois & de Valois, 1975). These early studies suggested that CP occurs in innate, unidimensional perceptual categories. However, later studies using more complex visual stimuli have also suggested that categorical perception occurs for overlearned categories such as facial expressions (Calder, Young, Perrett, Etcoff, & Rowland, 1996; de Gelder, Teunisse, & Benson, 1997; Young, Rowland, Calder, Etcoff, Seth, & Perrett, 1997) and familiar artificial objects (Newell & Bülthoff, 2002). Furthermore, CP effects emerge for facial identity when the faces are highly familiar but not when they are unfamiliar (Beale & Keil, 1995).

Yet a number of other studies have suggested that effects of CP are not limited to innate or highly overlearned categories from the real world, and that CP effects can emerge due to short-term learning. For example, Medin and Barsalou (1987) argued that CP occurs from acquired equivalence between members within a category and acquired distinctiveness of cues across a category. This equivalence between category members and distinctiveness across categories occurs as a result of familiarization with the items in the categories (Goldstone & Steyvers, 2001; Lawrence, 1949, 1950). Accordingly, the argument here is that CP effects for unfamiliar or novel stimuli can emerge as a result of short-term learning, or familiarization of category items. Many recent studies have suggested that this is indeed the case. For example, effects of CP have been found for artificial perceptual categories learned during the course of an experiment (Goldstone, 1994; Livingston, Andrews, & Harnad, 1998), for learned unfamiliar biological categories in the form of histological slides (Adamson & Sowden, 2000, 2001) and for unfamiliar faces learned in the course of an experiment (Levin & Beale, 2000).

In our experiments we investigated whether an effect of CP naturally occurs for the sex of faces. Male and female faces are well-known categories and it is argued that our ability to recognize the sex of an unfamiliar face, in the absence of social and biological cues (e.g., makeup or beard), is extremely good (Bruce et al., 1993). Moreover, both adults and older children can determine with a high degree of accuracy the sex of an unfamiliar adult or child by viewing an image of the face only (Wild, Barrett, Spence, O'Toole, Cheng, & Brooke, 2000). Indeed the well-known Bruce and Young (1986) model of face processing proposes that sex perception in faces, like expression recognition, is independent of the processes involved in recognizing a face, stating that sex information can be derived irrespective or whether the face is familiar or not (Bruce, Ellis, Gibling, & Young, 1987). Many studies have supported the claim that expression identification is a process independent of both facial identity (e.g., Campbell, Brooks, deHaan, & Roberts, 1996; Humphreys, Donnelly, & Riddoch, 1993; Young, Newcombe, deHaan, Small, & Hay, 1993; but see Schweinberger & Soukup, 1998, for further qualification) and sex perception (le Gal & Bruce, 2002).

Other studies have, however, challenged the notion that sex perception in faces is a process unrelated to the familiarity of the face: Whereas facial familiarity has been shown to be unrelated to expression identification, recent studies have found that sex identification is indeed related to face familiarity (Baudoin & Tiberghien, 2002; Bruce, 1986; Ganel & Goshen-Gottstein, 2002;

Goshen-Gottstein & Ganel, 2000; Rossion, 2002). For example, Bruce (1986) and Rossion (2002) found that sex discrimination was easier in familiar than unfamiliar faces. Furthermore, Goshen-Gottstein and Ganel (2000) conducted a repetition priming study where participants first conducted a sex judgement task on a set of familiar and unfamiliar faces. They found evidence for repetition priming from a single exposure to faces in the sex judgement task. Similarly, in a later study these authors reported that sex information interferes with face familiarity in a speeded classification task, suggesting again that facial identity and sex are nonindependent processes (Ganel & Goshen-Gottstein, 2002). Interestingly, in both of these studies the authors found that when hairstyle information was included in the images of faces, effects of priming and interference disappeared. The authors warned that when sex is determined from hairstyle, processes typically related to face perception are no longer involved and ''false''evidence of sex and identity separability can occur.

If sex is related to facial identity, unlike facial expression, what prediction can we make for the categorical perception of the sex of faces? Given that expressions are overlearned and universal categories (Ekman, 1994) and are unrelated to facial identity we would expect effects of CP to emerge for facial expression in unfamiliar faces. A number of recent studies have found evidence for the CP of facial expression in unfamiliar faces (Calder et al., 1996; Etcoff & Magee, 1992). On the other hand, CP effects of face identity should be confined to familiar faces only and not to unfamiliar faces. Indeed Beale and Keil (1995) reported effects of CP for famous (e.g., Clinton/Kennedy) but not for unfamiliar faces (e.g., Burns/Harris). Later, Levin and Beale (2000) argued that mere familiarization of faces, even during the course of an experiment, is sufficient for effects of CP on facial identity to emerge. With regard to sex perception, a recent study has found that effects of CP do exist for sex of faces (Campanella, Chrysochoos, & Bruyer, 2001). Campanella et al. tested one group of participants on categorization of a set of unfamiliar faces and found heightened discrimination performance in a second group of participants to faces straddling the category boundary. In their study, however, facial identity was confounded with the sex of the face. They tried to address this issue by demonstrating that effects of CP did not emerge for pairs of two individual faces of the same sex, but it did emerge for two faces of opposite sex. Nevertheless, CP effects in their study may still be related to facial identity or familiarization within the course of the experiment and not to sex *per se*.

In our study here we directly tested the role of face familiarity on the emergence of CP for the sex of faces. If sex perception is related to facial identity then we would expect that CP effects emerge for familiar faces only. If, on the other hand, the sex of a face is information unrelated to its identity (Bruce & Young, 1986), we would expect effects of CP to emerge for all faces, irrespective of familiarity, as suggested by the findings of Campanella et al. (2001).

To study categorical perception of, for example, facial expressions, one typically creates morph sequences between two images of the same face showing different facial expressions (e.g., happiness and anger). The resulting face stimuli from the morph sequence are then used for categorization and discrimination tasks to test whether the stimuli are grouped perceptually in two different categories. After creating morph sequences between pairs of male and female faces, we followed the classical procedure for testing categorical perception (Liberman et al., 1957, 1967). Participants were first required to discriminate between pairs of morph images in a discrimination task. Thereafter participants had to label each image as male or female in a categorization task. A CP effect occurs when (1) all faces are perceived as male or female in the categorization task, with a sharp change at the subjective category boundary even though all faces presented are evenly distributed along the artificial sex continuum and (2) pairs of faces are discriminated more accurately when they straddle the subjective category boundary than when both faces belong to the same category. In other words, in CP the category boundary defined by the categorization response function predicts the peak in performance in the discrimination task.

The face stimuli in our Experiments 1 and 2 were derived from face continua created between three-dimensional computer-reconstructed faces of real male and female heads. With new media technology and computational methods (Blanz & Vetter, 1999), however, it is possible to create another type of face continua based on the face of one person only, thus creating face images ranging along a sex continuum alone. We created and used such stimuli in experiments 3 and 4 to investigate how sex in faces is perceived.

STIMULI

We used faces chosen from a database of 100 female and 100 male faces collected by Vetter, Troje, and coworkers (O'Toole, Vetter, Troje, & Bülthoff, 1997; Troje & Bülthoff, 1996) to generate our stimuli. These faces were derived from three-dimensional scans of male and female participants (18–45 years old) obtained using a laser scanner (Cyberware™ 3030PS). All faces were devoid of secondary cues that could connote the sex of the face, such as makeup, accessories, hairstyle, or the presence of chin stubble or a beard. The faces in the database had previously been rated for distinctiveness (O'Toole, Edelman, & Bülthoff, 1998). We picked six male and six female faces with similar, low distinctiveness ratings that were most suitable for the morphing procedures. In all experiments we used face stimuli derived from some or all of these 12 faces.

Prior to being scanned, each participant was required to wear a swimming cap to cover his or her hair. The scanner created a profile of his/her face by shining a low-intensity laser beam in the shape of a vertical stripe onto the head. The laser beam moved around the head in 15 s while a video sensor captured the profile at a rate of about 30 times per second, sampling the shape of the head on a regular cylindrical grid of 512×512 points with a resolution of 0.8° horizontally and 0.615° vertically. Simultaneously, a second video sensor acquired colour information in the same spatial resolution. Consequently each scan is represented by two sets of data of 512×512 points each. One set describes the 3-D shape of the head (geometric data); the other one the RGB-colour values of each pixel of the head image (textural data). Face images were subsequently cropped at the hairline such that the resulting faces were devoid of hair or scalp, but included the ears and ended at the neck. After processing, each face was represented by approximately 7×10^4 vertices and the same number of colour values.

We used a correspondence method developed by Vetter, Blanz, and colleagues to generate our face continua between male and female faces (for more details see Blanz, 2000; Blanz & Vetter, 1999; Vetter & Poggio, 1997). Each continuum consisted of a set of 11 face images. In each set there were nine morphs evenly distributed in 10% increments between the endpoint faces, depending on how much ''female'' contributed to the face image. For example, the female endpoint face image was 100% female (and 0% male); the next face morph was 90% female (and 10% male), etc., with the final endpoint face image being 0% female (and 100% male). The different types of sex continua will be described in more detail within the appropriate experimental sections.

EXPERIMENT 1

In this experiment, our purpose was to investigate whether sex continua created between unfamiliar male and female faces were perceived categorically.

Method

Participants. There were 16 participants in the following experiment. Eight participants were female. All observers (age range 18–35 years) were paid volunteers and naive as to the purpose of the experiments. Most of the participants were undergraduate students from the Eberhard-Karls University of Tübingen, Germany, and others were members of the Max-Planck Institute, Tübingen. All participants had normal or corrected-to-normal vision.

Stimuli. Sex continua were created between each of all possible pairs of 6 male and 6 female laser-scanned faces, i.e., for 36 face combinations. Each continuum consisted of the original male and female endpoint faces and nine equidistant morph images between these endpoints. Both the shape and texture[1] of the face stimuli were informative about the sex of the face (Bruce et al., 1993; Hill, Bruce, & Akamatsu, 1995). All faces and morph images were presented

[1] Also called pigmentation or skin colour of the face.

from a full-face perspective. For examples of the face stimuli used in our experiment see Figure 1.

Apparatus. The digital face images (256 $\times$ 256 pixels in size and 8 bits per colour channel) subtended approximately 6° $\times$ 6° of visual angle and the average viewing distance was 57 cm. All experiments were conducted in a darkened room. The stimuli were presented using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993) on a Power Macintosh 9500/132 linked to a standard colour monitor. A button box was used for collecting responses.

Design. The experiment was based on a within subject design with face continua (36) and face images (11) as factors. The experiment itself was divided into two separate tasks: A discrimination task and a categorization task. The discrimination task was based on an XAB match-to-sample design used by Newell and Bülthoff (2002; see also Calder et al., 1996; Etcoff & Magee, 1992; Young et al., 1997) in which an image of a face (stimulus X) was presented initially in a trial followed by two face images (stimuli A and B) presented simultaneously, left and right of fixation. Stimuli A and B (test pair) were always physically different from each other; stimulus X was identical to either stimulus A or B. Stimuli A and B differed by 20% or two morph steps (e.g., face images 90% and 70%) along the sex continuum. All stimuli in one trial belonged to the same sex continuum. Each test pair was shown in 2 trials; X was identical to A in one trial and to B in the other, giving a total of 648 trials (36 face

Figure 1. Illustration of (top row) one of the 36 sex continua used in Experiment 1 and (lower rows) two of the 6 sex continua used in Experiment 2. Alternate face images from each pictured continua is shown. The endpoint face images on the very left and very right correspond to the original faces. All faces in all experiments were shown in colour.

continua, 9 image pairs, repeated twice). The order of the trials was randomized across participants. In the categorization task all face images from all continua were presented in a random order to each participant.

Procedure. Participants performed the discrimination task followed by the categorization task. This order ensured that performance on the discrimination task was not affected by prior knowledge of the faces in the categorization task.

In the discrimination task, a fixation cross preceded the first face image for 250 ms in each trial. The first face image (X) was shown for 750 ms in the centre of the screen followed by a blank screen for 1000 ms. The next pair of stimuli (A and B) remained on the screen until the participant pressed a response button. Each of the A and B stimuli were displayed 3 cm to the left and right of the centre point of the screen. An intertrial interval of 500 ms followed the participant's response. In order to acquaint participants with the XAB procedure, the experiment began with a random selection of eight practice trials. Three experimental blocks followed the training block and participants received a self-timed break between blocks. Participants were instructed to respond as fast and as accurately as possible, indicating which face image of the AB pair was identical to the preceding face image X. Participants were instructed to press the left (or right) response button of the button box to indicate that the left image (or right image) was identical to the first image. There were 648 trials (36 sex continua, 9 image pairs, repeated twice) in this task.

We used a forced-choice paradigm in the categorization task. Each image in each of the continua was used in the categorization task. A trial consisted of a single face image. Participants were shown all face images from all continua one at a time in random order. Participants were asked to classify each face image as male or female, as fast and as accurately as possible. Each face image was preceded by a 500 ms fixation cross. The face image then appeared and remained on the screen until the participant responded. An intertrial interval of 500 ms followed a participant's response. The right response button of the button box was assigned as the ''female'' response key, and the left was assigned the ''male'' response key. This allocation of left/right buttons to female/male respectively remained fixed throughout the experiment. The face images in the categorization task were always the same as those shown in the discrimination task. 336 trials were shown in this task (36 sex continua, 9 morphs images shown once, and 12 endpoint images shown once). Participants received self-timed breaks during the experiment. It took approximately 90 min to complete both tasks of the experiment.

Results

The mean frequencies with which participants categorized each image as female in the sex categorization task and the mean number of correct responses made to the XAB task are shown in Figure 2(a) and (b) respectively.

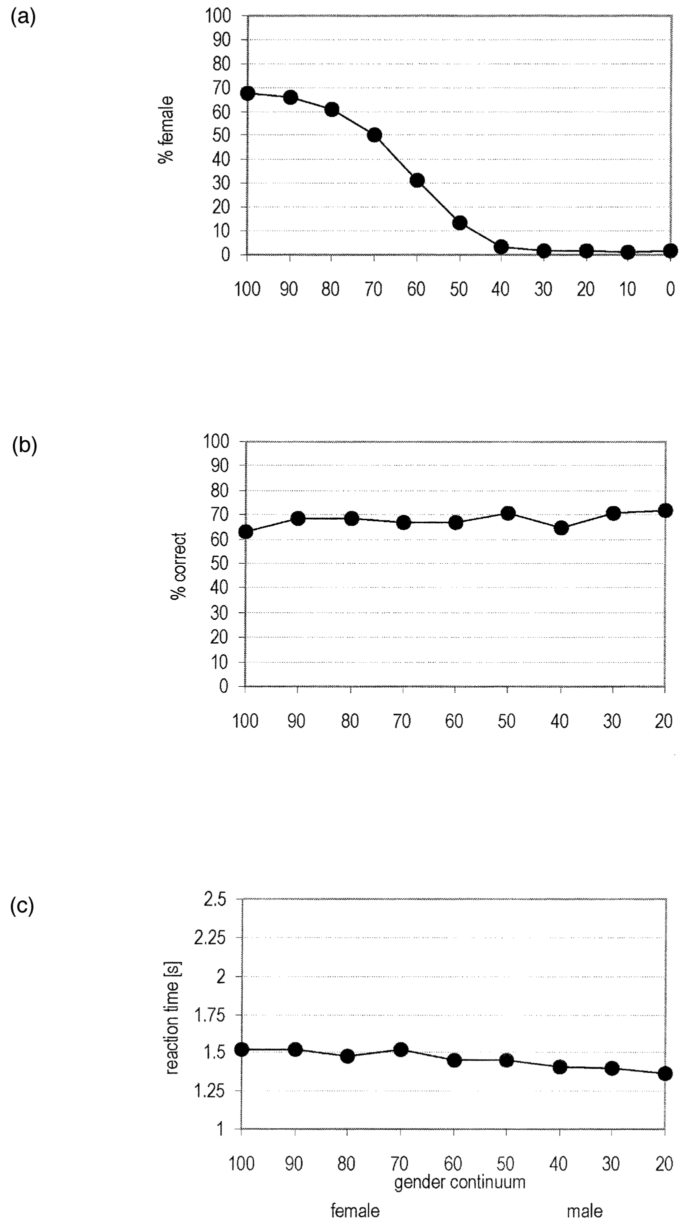

Figure 2. Plots showing data from Experiment 1: (a) Mean categorization data, (b) mean discrimination data, and (c) mean reaction times in the discrimination task. The categorization plot represents the mean percentage female responses to each face stimulus along the continuum. The discrimination plot represents the mean percentage of correct responses to each AB pair. In the upper plot, the sex continuum is represented by individual faces, in the middle and lower plot, by face pairs. In the lower plots, only the most female image of the pair is written on the X axis (i.e., 100 corresponds to the pair 100%–80%, etc.).

Categorization task

The subjective category boundary was determined as the point at which the categorization function crosses the 50% sex response level for all face combinations. The category boundary was calculated across all participants. Figure 2(a) shows that the point of subjective category boundary lies close to face image 70%.

Discrimination task

Performance was calculated across all participants and all face combinations (see Figure 2b). On average, discrimination performance was 68.2% correct. We conducted a one-way ANOVA across the mean percentage correct discrimination responses to each of the face pairs and found a significant effect of face pairs for subjects, $F(8, 120) = 2.10$, $p < .05$ and for items $F(8, 280) = 4.60$, $p < .001$. A post hoc Newman-Keuls test revealed that a greater number of correct responses were made to the 20%–0% face pair than the 100%–80% face pair. We then conducted post hoc planned comparisons between the discrimination performance to pairs of face images that lay at either end of the sex continua (i.e., the average performance to face image pairs 100%–80% and 20%–0%) and the discrimination performance at each image pairs that lay along the continuum. For evidence of CP the discrimination performance of images pairs that straddle the category boundary (i.e., faces images 80%–60%) should be greater than performance to the endpoints. We failed to find any significant differences between performance to the endpoint stimuli and any other image pairs. Moreover, we found no difference between performance to the endpoints and images straddling the category boundary, $F(1, 15) < 1$.

We investigated results for each face combination and each participant separately to ensure that the averaging process did not mask discrimination performance for some face combinations. Scrutiny of the observers' performance for all face pairs separately shows that, even for face combinations for which the sex categorization task was easier (i.e., a steeper step), discrimination was not better or more categorical than for other faces.

Reaction times. The reaction time across all trials in the discrimination task was 1.45 s. The average reaction times for each face pair are plotted in Figure 2(c). We conducted a one-way ANOVA on the average reaction times to the correct trials in the discrimination task. We found a significant effect of subjects, $F(8, 120) = 4.39$, $p < .001$, and of items, $F(8, 280) = 4.24$, $p < .001$. A post hoc Newman-Keuls analysis revealed that the reaction times to the image pair 100%–80% were significantly slower than reaction times to image pairs 40%–20%, 30%–10%, and 20%–0%. We also found that reaction times to the 90%–70% image pairs were longer than to the 30%–10% and 20%–0% pairs and the reaction times to 70%–50% were longer than to the 20%–0% pairs ($p < .05$). We failed to find, therefore, that reaction times to the pair of images straddling the category boundary (80%–60%) were significantly faster than to pairs of images lying within each category.

Discussion

The categorization plot indicates that the subjective sex boundary lies close to faces 70%, although in physical image-based terms the boundary lies at face 50%, which is equally male and female. We therefore found a male bias in the categorization judgement of the participants. Likewise, categorization performance for the endpoint faces shows that on average 32.2% of the female endpoint faces (image 100%) were misjudged as male faces while only 1.5% of the male endpoint faces (image 0%) were misjudged as female. Bruce et al. (1993) in their study with laser scans of faces have also reported that the sex of female faces is more often misjudged than that of male faces. Other researchers who used images (not laser scans) of adults' or children's faces have also found a male bias in sex judgement (Cheng, O'Toole, & Abdi, 2001; Intons-Peterson, 1988; Wild et al., 2000).

Potential reasons for a male bias are numerous. There could be social as well as perceptual reasons for this male bias. For example, the judgement of female faces might be affected by the absence of hair. Although adults know that the sex of a person does not strictly correlate with the presence and/or length of hair and makeup (and so on), it might be hard for them to ignore the absence of hair in our face stimuli. Another, perhaps more significant, factor appears to be that there are no positive markers for female adult faces, while there are some for male faces such as the presence of facial hair (or stubble), bushy eyebrows, strong jaws, etc. Female faces in contrast keep many features of a child's face, for example small, round faces, big eyes, and smooth skin (e.g., Abdi, Valentin, Edelman, & O'Toole, 1995; Clemente, 1985; Zebrowitz, 1997). In our experiments, many participants reported categorizing female faces as male because they judged these faces to be young male faces. Thus some female faces might have been mistaken for male because of some ambiguity about the age as well as the sex of the faces. Finally, another factor contributing to the male bias might be perceived attractiveness. It is often commented upon that most laser-scanned faces in our database look less attractive than the original faces of known individuals. We speculate that some observers might have categorized female faces as male because of their expectancy for female faces to be more attractive. Despite this male response bias, however, participants perceived two sex categories; an obvious prerequisite for the potential presence of a categorical effect. Whether this perception is categorical or not is determined by the results of the discrimination task.

A comparison of the within-category versus between-category discrimination performance failed to reveal any categorical effects. The results suggest that we do not perceive the sex of unfamiliar faces categorically. An analysis of the reaction times show a similar pattern to the accuracy data, thus we found no evidence of a speed–accuracy tradeoff; i.e., participants were not less accurate when they answered quickly.

Despite the fact that performance was over chance level, we were concerned that many of our participants complained about the difficulty of the discrimination task. To ensure that the absence of CP in Experiment 1 was not due to the discrimination task being too difficult for the observers, in a series of control experiments we repeated the same test with image pairs that were three steps apart. Furthermore, we repeated the experiment using face stimuli presented from a 20° viewing angle, in order to better reveal any geometric information pertaining to the sex of the face (Burton, Bruce, & Dench, 1993). First, we found that by increasing the image step size the overall correct response rate increased in the discrimination task, although we again failed to find evidence for categorical perception. We found 14.0% errors in total, 11.7% errors were to the endpoint faces and 12.8% errors to the image pairs across the boundary. These findings suggest that the difficulty of the task in the present experiment did not obscure effects of CP that may otherwise have been present. Second, when we rotated the faces by 20°, participants were indeed around 10% better at correctly labelling female faces while male endpoint faces were still correctly labelled as male nearly 100% of the time, and the categorization curve was consequently more step-like. Nevertheless, the results of the discrimination task did not show any categorical effect (Bülthoff, Newell, Vetter, & Bülthoff, 1998).

EXPERIMENT 2

It has been suggested that familiarity with the endpoint faces on a continuum is a prerequisite for CP of aspects of face perception such as facial identity or expression (Beale & Keil, 1995; Etcoff & Magee, 1992). Effects of CP on facial identity also emerge after short term learning of previously unfamiliar faces (Levin & Beale, 2000). In the following experiment, we tested whether familiarization with the endpoint faces would facilitate CP for face sex. Familiarity was manipulated by exposing the participants to the endpoint face stimuli prior to the experiment. Furthermore, we optimized conditions for learning the identity of the face endpoints by adopting a methodology used by Levin and Beale of blocking the trials per face continua. Levin and Beale have shown that CP for the identity of newly learned faces exists after short-term learning of the endpoint faces. In the first experiment of their study participants viewed pairs of faces and were asked which one was more similar to one of two faces they saw at the beginning of the testing block. Therefore, face combinations were blocked in their experiment. In our experiment we used a procedure somewhat similar to that of Levin and Beale, in that we tested each face combination separately in different blocks. Our procedure differs in that the observers were not familiarized with the endpoint faces prior to each experimental block but instead they were familiarized with all the faces grouped by sex before the whole experimental session. Furthermore, we did not use their ''better-likeness'' task, but used the same XAB task as used in Experiment 1. In this experiment we predicted that if the sex of a face is linked to facial identity then we would expect effects of CP to emerge for the sex categorization of familiar faces.

Method

Participants. Twenty participants took part in the following experiment. Ten of the participants were female. All observers (age range 18–29 years) were paid volunteers and naive as to the purpose of the experiments. The participants were undergraduate students from the Eberhard-Karls University of Tübingen, Germany, or were members of the Max-Planck Institute, Tübingen. All participants had normal or corrected-to-normal vision. None of these participants had taken part in the previous experiment.

Stimuli. The same six male and six female faces used in Experiment 1 were used in this experiment. However, here each face was paired only once, giving a total of six sex continua. As already mentioned, we found in other experiments not reported here that by presenting the face stimuli at a slight angle of 20° we increased overall performance relative to the full-face view, although CP effects did not emerge to unfamiliar faces (Bülthoff et al., 1998). In order to optimize sex perception therefore, we decided to present our face stimuli at 20° away from the full-face view (Bruce et al., 1993; Burton et al., 1993). As in the previous experiments, 11 equally distant image steps were extracted from each continuum.

Design. The experiment was based on a within subject design with face continua (six) and face images (eleven) as factors. In this experiment, the images from each of the six continua were tested in separate blocks in both the categorization and discrimination tasks. The order of the blocks was counterbalanced across participants. The XAB discrimination task was similar to that in Experiment 1 and each test pair was shown in six trials (X was identical to A in half of the trials and to B in the other half), giving a total of 324 trials (six face continua, nine image pairs, repeated six times). As in the previous experiment, one practice block of eight randomly chosen trials preceded the six experimental blocks in the discrimination task. In the categorization task, each image was shown twice, giving a total of 132 trials (six face continua, eleven images, repeated twice). The presentation order of the trials in each block was randomized across participants.

Procedure. Before performing the experiment, all participants were familiarized with the six male and six female original faces for 10 min. The images of all faces shown from a 20° view were presented as printed images and displayed together on two pages (one for each sex). We labelled each face with a short first name, and each name clearly (for our German participants) connoted the sex of the face, e.g., Kurt and Heidi. Participants were told to view carefully the face images and to learn their names because they would be asked to recognize those particular faces later. In the subsequent recognition test the participant had to name each randomly presented face image. We set a performance criterion of 11 out of a possible 12 correct responses in the recognition task before the participant could continue with the experiment. Most of the participants reached this criterion after one learning block. The learning block was repeated for those who did not reach criterion. After successful completion of the recognition task, participants were told that the names would never be used in the subsequent experiments and proceeded to perform a discrimination task and a categorization task following the procedure described in Experiment 1. Participants needed about 50 min to complete this experiment.

Results

The mean percentage of female responses to the sex categorization task and the mean number of correct responses made to the XAB task is shown in Figure 3(a) and (b) respectively.

Categorization task

The subjective category boundary was determined by the categorization performance as described in the previous experiments. In this experiment, the category boundary lay close to face image 60%.

Discrimination task

The average correct performance over all face pairs was 65%. We conducted a one-way ANOVA across performance to the image pairs. We found a significant effect of image pairs for the subjects analysis, $F1(8, 152) = 2.36$, $p < .05$, and for items analysis, $F2(8, 40) = 2.53$, $p < .05$. Post hoc Newman-Keuls analysis revealed that a greater number of correct responses were made to the 70%–50%% image pair than the 100%–80% image pair ($p < .05$). No other differences were found. We then conducted planned comparisons between the mean performance to the endpoint image pairs and each of the other image pairs along the continuum. We found a significant difference between the mean performance to the endpoints and performance to the image pairs 70%–50%, $F(1, 19) = 4.60$, $p < .05$. We found no other differences between the image pairs and the endpoints. The 70%–50% pair of face images straddles the sex boundary as determined by the categorization task; therefore, in this experiment we have found evidence for the categorical perception of sex in familiarized faces.

(a)

(b)

(c)

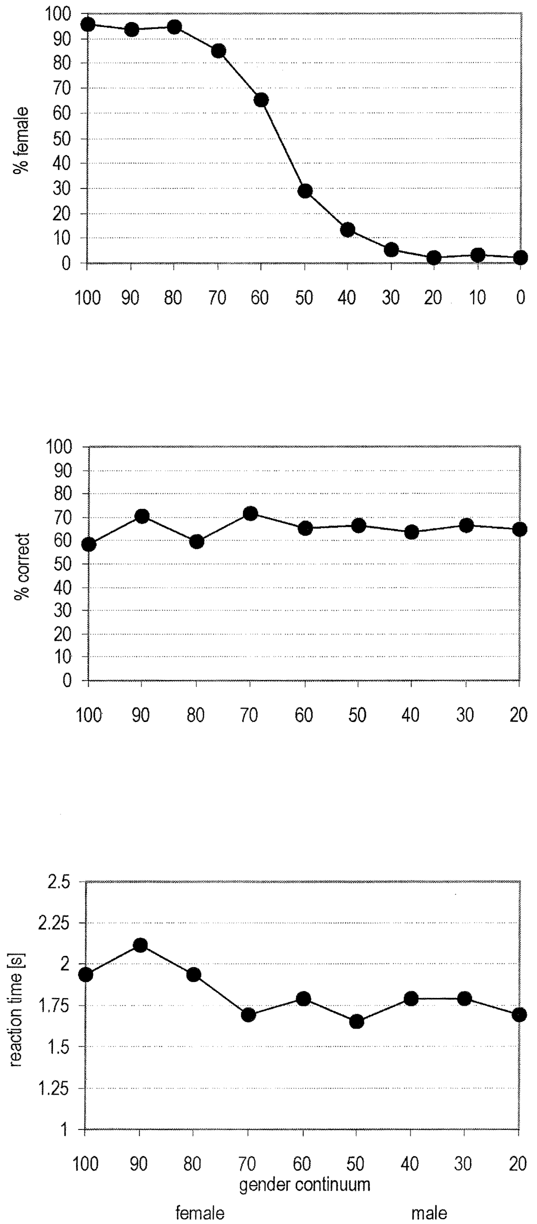

Figure 3. Plots showing data from Experiment 2: (a) Mean categorization data, (b) mean discrimination data, and (c) mean reaction times in the discrimination task. In the lower plots, only the most female image of the pair is written on the X axis (i.e., 100 corresponds to the pair 100%–80%, etc.).

Many studies on categorical perception have used a second measure as evidence for the presence of categorical perception. Typically, the discrimination data is correlated with predicted performance calculated on the results of the categorization task.[2] The formula we used for deriving the predicted performance is adapted from that used by Calder et al. (1996) and Liberman et al. (1957). We conducted a Pearson correlation between the observed and predicted discrimination performance. The correlation was not significant, $r = .33$, n.s.

Reaction times. The reaction time across all trials was 1.82 s. We conducted a one-way ANOVA on the mean reaction times to the correct trials in the discrimination task (see Figure 3c). We found a significant effect of subjects, $F(8, 152) = 2.20$, $p < .05$, and of items, $F(8, 40) = 2.70$, $p < .05$. Post hoc Newman-Keuls analyses revealed that the reaction times to the image pairs 100%–80% were significantly slower than reaction times to image pairs 60%–40% and 30%–10% ($p < .05$). Thus the variations in reaction times were not related to the CP effect found in the discrimination responses.

Discussion

The shape of the categorization response function in Figure 3(a) indicates that categorization performance was improved for familiar faces at the female end of the continua; the endpoint female faces as well as morphs at the female end of the continua were more often categorized as female than in Experiment 1 (see categorization performance plotted in Figure 2a). The same endpoint female faces were judged as female in 95.8% of the trials compared to 67.8% in Experiment 1. At the male end of the continua, endpoint male faces were judged to be male in 97.8% of the trials compared to 100% in Experiment 1. The subjective sex boundary was between faces 60% and 50% for familiar faces, which is closer to the physical image-based sex boundary (i.e., face 50%) than for unfamiliar faces (the sex boundary was close to face image 70% in Experiment 1). Thus, the male bias was reduced, along with the better categorization for faces at the female end of the continua.

The results of the discrimination task show a categorical effect after familiarization that was not present in Experiment 1. Face pairs straddling the sex boundary were significantly better discriminated than end pairs of face images. However, we failed to find a significant correlation between the predicted discrimination performance (derived from the categorization data) and the observed performance, as is often reported in studies of CP (Calder et al., 1996; Newell & Bülthoff, 2002). One reason why we failed to find a correlation may be due to

[2] In summary, the function used to calculate the predicted discrimination performance for each face pair was the sum of the mean discrimination performance to the pairs of face images on either end of the sex continuum and 0.3 of the identification difference for each pair of images tested. See Calder et al. (1996), Liberman et al. (1957), and Newell and Bülthoff (2002) for further details.

the categorization task and the discrimination task measuring two different percepts. For example, the absence of a significant correlation might indicate that participants performed the categorization task by using facial information pertaining to the identity of the faces, whereas the design of the discrimination task obliged participants to perform this task using facial information pertaining to the sex of the faces only. Thus, we speculate that the lack of a significant correlation may be due to different information underlying different tasks.

Our face stimuli, like face stimuli used by other authors (Campanella et al., 2001), confounds information relevant to sex with information related to identity. For example, by morphing a male face with a female face we also morphed the identity of the male face (e.g., Kurt) with the female identity (e.g., Heidi). Therefore, effects of CP may emerge due to idiosyncratic differences between the faces of, e.g., Kurt and Heidi and not between male and female characteristics *per se*. We were concerned whether the categorical effect measured in this experiment was a result of the familiarization with the identity not with the sex of the endpoint faces. Furthermore, the blocked presentation might have favoured the use of identity related information. We controlled for possible effects of identity in Experiment 3.

EXPERIMENT 3

In the following experiment we investigated the effect of CP in male and female faces when identity information was controlled. Using the algorithm of Blanz and Vetter (1999) we created new sex continua in which the endpoint faces were of different sex but shared the same facial features. With these face stimuli we could investigate whether face sex was perceived categorically when there was no change of characteristic facial features related to identity. All face stimuli were unfamiliar to the participants in this experiment; the effect of familiarization was investigated in Experiment 4.

Method

Participants. Eighteen participants (five males and thirteen females) took part in this experiment. All observers (age range 18–38 years) were paid volunteers and naive as to the purpose of the experiments. The participants were undergraduate students from the Eberhard-Karls University of Tübingen, Germany, or were members of the Max-Planck Institute, Tübingen. All participants had normal or corrected-to-normal vision. None of these participants had taken part in any of the previous experiments.

Stimuli. For this experiment we used what we term ''one-identity'' continua. The facial variations in our morph sequences were solely determined by the sex differences between the endpoint faces and not by any differences in identities. (Identity means here the unique characteristic features of a face.) Each of these continua was based on a single individual face. In a morphing procedure, all features of a female face were transformed (masculinized) using the sex vector described in the Appendix. We computed the corresponding male faces of six female faces. Six one-identity face combinations; each based on a different female face, yielded a total of 66 stimuli. Figure 4 shows the six original female faces, their computed equivalent male faces, and four morphed faces between the original female and derived male face. The faces were presented from a 20° angle. In all other ways the stimulus images were prepared as described in Experiment 1.

Apparatus. We used Eprime software (Psychological Software Tools, Inc.) on a personal computer running Microsoft Windows 98, instead of PsyScope running on a Macintosh as in the previous experiments. A button box was again used to collect response data.

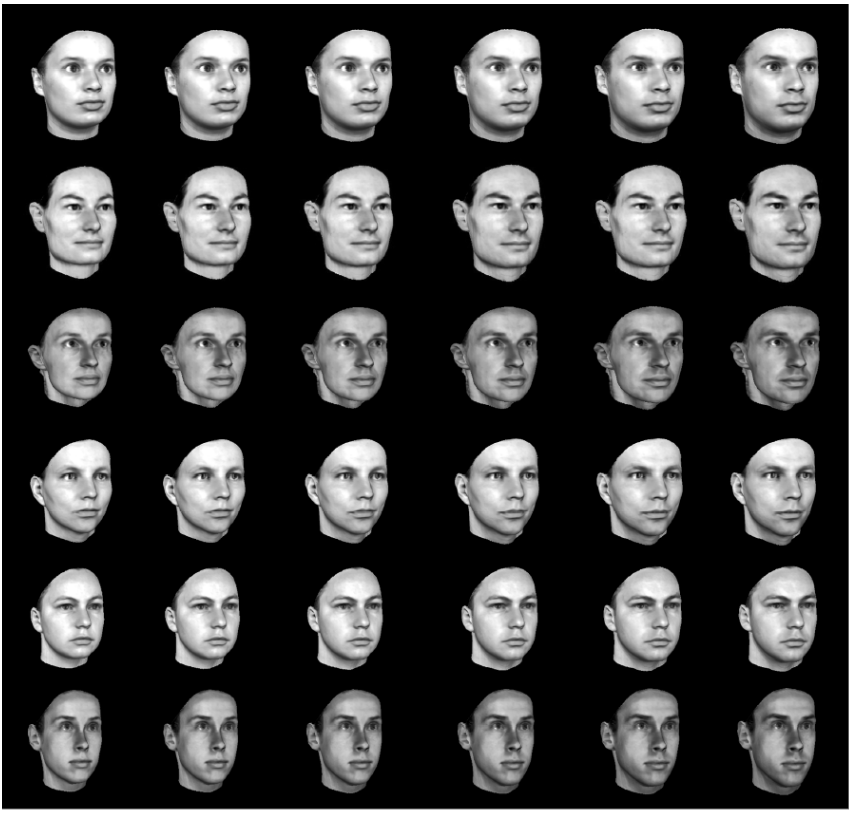

Figure 4. Illustration of the six one-identity continua used in Experiment 3. In all rows, the six original female faces are on the left. Sex continua in the second and third row only were used in Experiment 4. For more explanation, see text.

Design. Experiment 3 was based on a within-subjects design with ''one-identity'' continua (six) and face images (eleven) as factors. Pilot studies with ''one-identity'' sex continua had revealed that participants were performing at around chance level even when the AB images were three steps apart in the XAB match-to-sample paradigm. Therefore, in this experiment we chose a more accessible discrimination task. Here participants performed a simultaneous same/different task (see Calder et al., 1996). In the different trials the two face images were three steps apart. There were eight ''different'' stimuli pairs (100%–70%, 90%–60%, ... 30%–0%) and eleven ''same'' pairs (100%–100%, 90%–90%, ... 0%–0%). Trials were blocked by face combination. There were 65 trials per face combination with 33 same trials and 32 different trials. (Each same trial was repeated three times and each different trial was repeated four times.) Thus there were 390 trials per session (6 sex continua, and 65 trials per face combination).

Procedure. A fixation cross preceded the pair of face stimuli for 250 ms in each trial. The pair of stimuli were presented simultaneously about 3 cm left and right of the fixation point and remained on the screen until the participant pressed a response button of the button box. An intertrial interval of 500 ms followed the participant's response. Participants had to indicate whether both images were the same or different by pressing the left or right response button respectively. Face images remained on the screen until the participant made a response. There was always a short practice block of eight trials before starting the test blocks. In the categorization task each face image was presented four times, giving a total of 264 (6 continua $\times$ 11 images repeated 4 times) trials in one session. As in Experiment 2, the stimuli were blocked by face combination in order to promote any effects of CP that might be present. In all other ways the procedure was the same as in Experiment 1. Participants needed about 55 min to complete this experiment.

Results

The mean percentage of female decisions to the sex categorization task is shown in Figure 5(a). The discrimination scores on the same/different task were converted to d′ scores (a sensitivity measure). Figure 5(b) shows the mean d′ scores across all participants and all continua.

Categorization task

As before, the point of subjective category boundary was determined by the results of the categorization task. In this experiment, the subjective category boundary lay between face image 60% and face image 70%, although the categorization curve does not appear as a step-like function across the stimulus continuum.

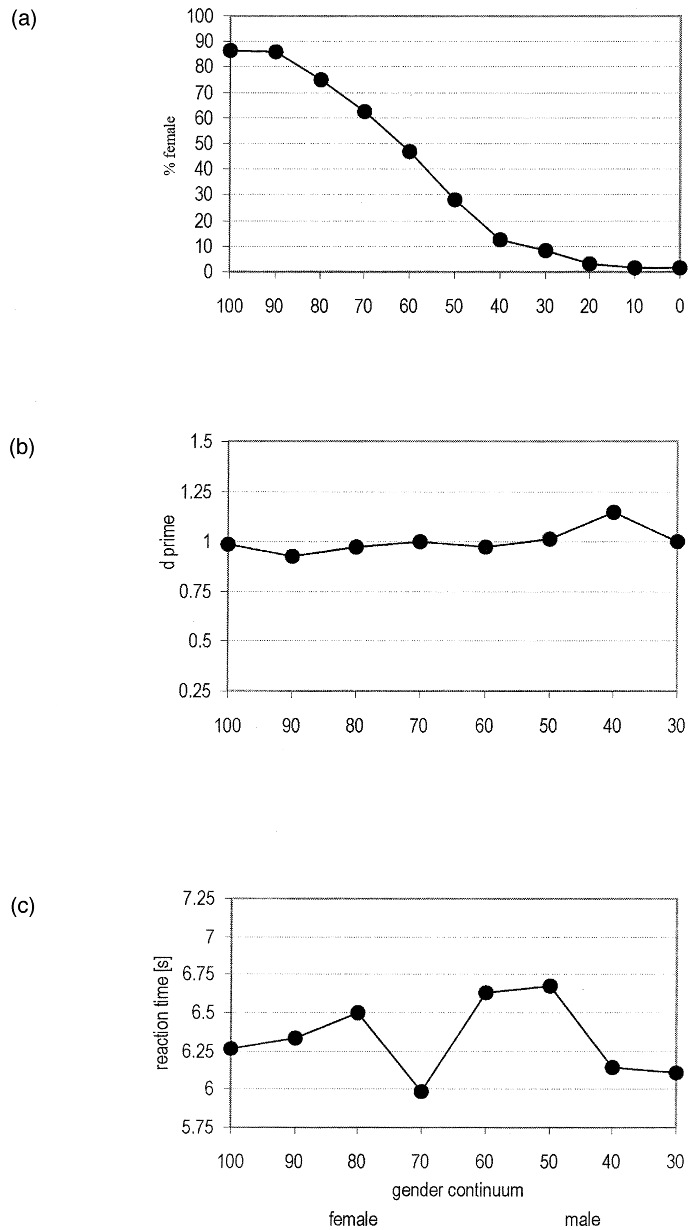

Figure 5. Plots showing data from Experiment 3: (a) Mean categorization data, (b) mean d′ discrimination data, and (c) mean reaction times in the discrimination task. In the lower plots, only the most female image of the pair is written on the X axis (i.e., 100 corresponds to the pair 100%–70%, etc.).

Discrimination task

The mean percentage performance was 68.8 across all ''different'' face trials and 64.5 across all ''same'' face trials. We conducted a one-way ANOVA on the d' scores across all face pairs. We found no significant difference of performance across the image pairs, $F(7, 199) = 0.43$, n.s. We then conducted planned comparisons between the average d′ performance to the endpoint image pairs and the d′ performance to each of other image pairs along the continuum. We found no significant differences and failed to find a difference between the endpoints and the 80%–50% pair straddling the category boundary, $F<1$.

Reaction times. Participants reported that they had great difficulties in performing the task. This difficulty is reflected in their reaction times. The average reaction times for the correct trials to all face pairs are plotted in Figure 5(c). The mean reaction time to the ''same'' trials was 6.8 s and to the ''different'' trials was 6.3 s. We conducted separate one-way ANOVAs on the mean reaction times to the ''same'' image pairs and the ''different'' face image pairs in the discrimination task. We failed to find a significant effect of face pair on response times either for the ''same'' trials, $F(10, 170) = 1.22$, n.s., or the ''different'' trials, $F(7, 119) = 0.93$, n.s.

Discussion

In this experiment only the faces at the female endpoint of each continuum corresponded to the features of a real person, and all other faces were artificially created. Although the shape of the categorization function is not step-like, it suggests that participants could identify the proportion of both sexes in the images and could classify the faces by their sex mixture accordingly. Note that a male bias in categorization responses is again present to the same degree as was found in Experiment 1 (see Figure 2a).

We were unable to find that the face pairs straddling the sex boundary were significantly easier to discriminate than within-category pairs. Moreover, because of the difficulty of the task, the reaction times were extremely slow. In fact, we were surprised to realize how difficult it was for the participants to discriminate between faces from the ''one-identity'' continua. Although we are expert at deciphering slight changes in facial expression, we are obviously not trained to discriminate between faces differing in sex quality only.

Our findings suggest that sex-related changes in unfamiliar facial features are less obvious to the observers than identity-related changes.[3] Here we needed to employ a same/different paradigm instead of the XAB paradigm used in the

[3] As an interesting aside, we used the same method to masculinize the head of one of the authors (F. N.). The resulting face was clearly identified by other family members as the face of one of her brothers.

previous studies because participants performed at chance level on the XAB task using ‘‘one-identity’’ face continua. This difficulty in discrimination is especially clear when we compare the performance of participants with the one-identity continua used in this experiment to the results obtained by Beale and Keil (1995) with face continua in which the sex remained the same but the identity varied. Participants in the first experiment of the Beale and Keil study could perform a XAB match-to-sample task with AB pairs that were two steps apart, while participants in our experiment were at chance performance even when the one-identity face pairs were three steps apart. The results of Beale and Keil that we refer to were obtained for well-known faces, but unfortunately they used a different paradigm when they tested unfamiliar faces so it is difficult to compare our studies directly.

What is important for our investigation here is that participants did not perform the discrimination task in a categorical way. This result confirms the findings in Experiment 1, i.e., a lack of CP for sex of unfamiliar faces, and demonstrates that it was not facial information related to the change of identity that hindered CP for sex in Experiment 1. The results we obtained in Experiment 3 suggest again that sex generally is not perceived categorically in faces.

Due to the difficulty of the discrimination task for the face stimuli used here and the general difficulty to obtain categorical perception of sex throughout this study, we wondered whether pure sex information in faces was at all available for categorical perception. We addressed this question in the next experiment by familiarizing participants with the faces from the ‘‘one-identity’’ continua.

EXPERIMENT 4

Many studies have shown that CP can be induced for unfamiliar categories after proper training (see Adamson & Sowden, 2000, 2001; Goldstone, 1994; Goldstone, Lippa, & Shiffrin, 2001). In Experiment 3, participants were not familiar with the face stimuli. In the present experiment we treated sex categories as unfamiliar categories that participants had to learn. Consequently, we trained participants to categorize the face stimuli by sex (see Goldstone, 1994; Goldstone et al., 2001, for similar procedures) prior to testing and investigated whether familiarization could induce categorical perception of sex as we found in Experiment 2.

Method

Participants. Eighteen participants took part in the following experiment. Thirteen of the participants were female. All observers (age range 19–27 years) were paid volunteers and naive as to the purpose of the experiments. The participants were undergraduate students from the Eberhard-Karls University of

Tübingen, Germany, or were members of the Max-Planck Institute, Tübingen. All participants had normal or corrected-to-normal vision. None of these participants had taken part in any of the previous experiments.

Stimuli and apparatus. In this experiment we only used two of the six face continua used in Experiment 3. We wanted to get enough measurements on a few face pairs whilst not overburdening our participants with this difficult task. We used the same apparatus as described in Experiment 3.

Design. Participants went through a training session before performing the discrimination and categorization tasks similar to those described in Experiment 3. The training phase consisted of the same forced-choice sex categorization task as in the test, except that participants were given appropriate visual and auditory feedback on the accuracy of their response on each trial. A response was labelled as correct every time a participant categorized any of the five images on the female (or male) side of the continuum as female (or male), otherwise it was labelled as incorrect. Ten faces of each continuum were shown 10 times in the training phase; faces 100–60 (40–0) had to be labelled ''female'' (''male'') by the participant for a correct response. Note that face image 50 was not presented because it lay on the sex boundary. Thus there were 100 trials per training session. The order of the trials within each session was randomized across participants. There were two training blocks, one for each one-identity continuum, and participants conducted the experimental session after each training block. The order of presentation of the two training blocks was counterbalanced across participants. In the experimental session, participants were presented with the same stimuli as in the training phase. The design of the discrimination and the categorization tasks was the same as in Experiment 3 except that image 50% was not used in the categorization task because this face was not explicitly trained. There were 65 trials in each discrimination task and 30 trials in each subsequent categorization task. Participants received a self-timed break between each block.

Procedure. In the training session each training trial consisted of the following sequence of events: A fixation cross was shown for 250 ms, followed by a training face image that remained on the screen to which the participant responded ''male'' or ''female'' by pressing the appropriate button. As soon as the participant responded, feedback in the form of a text display stating the correctness of the participants' decision was given. This feedback was displayed for 750 ms. Visual feedback was coupled with auditory feedback. Depending on the correctness of the response, one of two different beeps was presented. An intertrial interval of 500 ms followed the feedback display. We set a response criterion of 90% correct in the training block for participants to be allowed to proceed to the experimental sessions.

The procedure of the experimental session followed that of Experiment 3. Participants needed around 40 min to perform the experiment.

Results

Training phase

Three of our participants (one male and two female) failed to reach criterion in the training session. The average performance of the remaining 15 participants, calculated over both training periods, was 92.2% correctly classified trials.

Test phase

The mean percentage of female decisions to the sex categorization task and the mean discrimination performance for the same/different task are shown in Figure 6(a) and (b) respectively.

Categorization task

The perceived sex boundary was very close to face image 50%, and therefore close to the physical sex boundary. Participants classified 94% of all faces belonging to the female category (face 100% to face 60%) as female and 94% of all male faces (face 40% to face 0%) as male. Almost all categorization errors occurred to the two faces closest to the sex boundary in each sex category. For comparison, participants without training in Experiment 3 correctly classified 81% of the female faces and 91% of the male faces from the same continua.

Discrimination task

The mean performance was 58% correct to the different-face pairs and 66% correct to the same-face pairs. The discrimination scores on the same/different task were converted to d′ scores. Some of the participants made more false alarm responses than hits to some face pairs and d′ could not be calculated. This missing data constituted 2% of the entire data set. We conducted a one-way ANOVA on d′ performance across all face pairs and found no significant effect of face pair, $F(7, 105) = 1.98$, n.s. We then conducted planned comparisons between the average d′ performance to the endpoint images and each face pair along the continuum. We found a significant difference between performance to the endpoints and to the face pair 60%–30%, $F(1, 15) = 6.25$, $p < .05$. No other differences were found. The face pair 60%–30% straddles the sex boundary at 50%. Therefore, our findings suggest evidence for categorical perception of sex after familiarization.

We then used the same method as in Experiment 2 to derive predicted performance on the results of the categorization task. Note that we have no

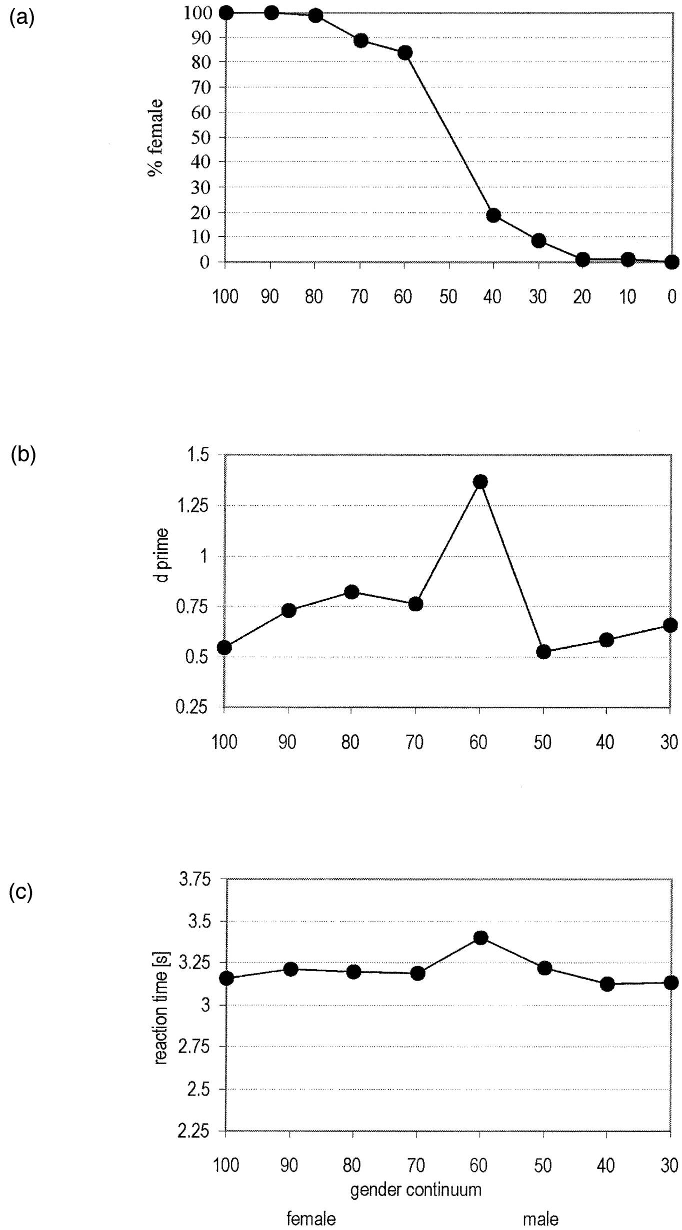

Figure 6. Plots showing data from Experiment 4: (a) Mean categorization data, (b) mean d′ discrimination data, and (c) mean reaction times in the discrimination task. In the lower plots, only the most female image of the pair is written on the X axis (i.e., 100 corresponds to the pair 100–70%, etc.).

categorization data for face image 5. We conducted a Pearson correlation between the observed and predicted discrimination performance. The correlation was close to being significant, $r = .75$, $p = .054$.

Reaction times. Participants in this experiment found the task less difficult than participants in Experiment 3. This is reflected in their reaction times, i.e., here participants needed about half the time to respond than participants in Experiment 3. The average reaction times for all face pairs are plotted in Figure 6(c). The mean reaction time to the same-face trials was 3.57 s and to the different-face trials was 3.20 s. We conducted separate one-way ANOVAs on the mean reaction times to the same image pairs and the different face image pairs in the discrimination task. We found no significant effect of reaction times for the same trials, $F(10, 130) = 1.52$, n.s., or for the different trials, $F(7, 98) = 0.27$, n.s.

Discussion

The important result here was that CP for sex was present after sex category training on face stimuli that differed on sex characteristics only.

In comparison to the results of Experiment 3 where the participants were not familiarized with the stimuli prior testing, classification of faces by sex was improved with training and this categorization improvement was still present after participants performed the discrimination task (see Figures 5a and 6a). We also noted that the response times of the trained participants were twice as fast compared to the response times of participants without training. Nevertheless, the response times were relatively slow, at about 3 seconds, indicating that even with training this task was a difficult one for our participants to perform. We also suggest that these long response times may have obscured any effects of speeding up in discriminating faces straddling the category boundary than within-category face pairs.

We found that the correlation between predicted and actual performance was very close to significant. We think the only reason it is not significant is because of the missing data point in the predicted function. At the very least, however, it is interesting to note that the correlation between predicted and actual performance is much better than that found in Experiment 2. Thus we suggest that the finding from the current experiment reflect that the same underlying processes were involved in both the categorization and discrimination tasks, unlike in Experiment 2 where the categorization task may have been based on face identity alone.

GENERAL DISCUSSION

In Experiments 1 and 3 we failed to find evidence that unfamiliar male and female faces were perceived as discrete categories. In contrast, when participants were trained prior to the experiment on the identity of the face (Experiment 2) or on categorizing faces based on sex differences only (Experiment 4) then effects of CP emerged.

In Experiment 1 we tested for effects of categorical perception in unfamiliar faces, under presentation conditions that minimized learning of the endpoint stimuli. We failed to find evidence for CP. In Experiment 2, we familiarized participants with the endpoint faces in each continuum by asking them to learn the names of these faces and subsequently testing recognition performance prior to the experiment. Here we found that discrimination performance to face pairs straddling the category boundary was better than face pairs within the female or male categories. This is the hallmark of CP. However, the correlation between the predicted and observed discrimination performance was not significant. We suggest that this is because the categorization and discrimination tasks engaged different types of face representation, facial identity for the categorization, and sex for the discrimination tasks. Unfortunately the only study reporting CP of sex in faces did not include this analysis due to the design of the experiment (Campanella et al., 2001); therefore it is not certain whether the changes in facial identity affected the correlation. Nevertheless, we were concerned that changes in identity within our face continua (i.e., from Kurt to Heidi) were having an effect on the CP of the sex of faces. We addressed this issue in Experiments 3 and 4.

In Experiments 1 and 2, identity and sex changed in the face stimuli; therefore the possibility remained that participants were extracting identity information in order to perform the tasks. For Experiments 3 and 4 we created a set of stimuli where only the sex information changed along each continuum. In Experiment 3 the participants were unfamiliar with the face stimuli used. Here again we failed to find evidence of CP for unfamiliar faces. However, in Experiment 4, we trained participants to classify the faces of sex-only continua using a feedback procedure. We found evidence for effect of CP of sex in familiarized faces with an almost significant correlation between the predicted and actual discrimination performance.

At first glance, our findings seem to disagree with those of Campanella et al. (2001), who reported effects of CP for the sex of unfamiliar faces. However, differences in face familiarity and identity changes may account for the difference in findings between the two studies. In terms of face familiarity, for example, Campanella et al. used fewer original faces in their study, relative to our experiments. In their Experiment 2, three female and three male original faces were used to create the morphing continua. Although these originals were not used in the experiments, very similar versions were (e.g., 8% and 92% morphs). Furthermore, each of these versions was repeated 12 times during the course of the experiment. Fewer numbers of morph images were used in these experiments, consequently about 30% of the trials contained repeats of the near-original faces. We would argue that the combination of a small number of face originals (i.e., six) together with repetitions of images near the originals (i.e., twelve repeats) amongst a small number of morph images allowed for effects of familiarization to emerge during the course of their study. In contrast, in our Experiment 1 we used 12 original faces each shown 12 times within a total of over 600 trials in the experiment, all randomly presented. These experimental conditions may have discouraged any learning of the faces during the course of the experiment, and effects of CP were therefore not found. In terms of facial information, Campanella et al. argued that their effects of CP were due to sex and not identity changes; because CP was found only between opposite sex faces and not between two different same sex faces (see their Experiment 3). However, two faces of opposite sex are more easily discriminable than two of the same sex; therefore, effects of CP here may have been due to more rapid acquisition of the distinguishing features between the different-sex than same-sex endpoint faces. Campanella, Hanoteau, Seron, Joassin, and Bruyer (2003) reported that CP effects emerge between two unfamiliar faces belonging to different identities but not between two unfamiliar faces originating from the same identity, suggesting that face-pair similarity may play a role in CP. In any case, we would argue that the findings of Campanella et al. (2001) may be due to familiarization, or rapid learning, of the identity of the faces during the experiment, and consequently the role of familiarity in CP effects of sex is equivocal in this study.

In our study we found that the sex of unfamiliar faces is not perceived categorically but with training, however, CP effects emerge. The absence of a categorical effect of sex in unfamiliar faces is surprising given that other important information about faces (e.g., expressions) has been proved repeatedly to be perceived categorically (Young et al., 1997). We believe that there is a simple explanation for the differences between the perception of facial expressions and perception of sex in faces, namely that the sex of a person is related to his/her identity but that expression is unrelated to identity. In our study we have shown that familiarity and sex perception are related which is a finding consistent with other recent studies of Baudouin and Tiberghien (2002), Goshen-Gottstein and Ganel (2000) and Rossion (2002). Our findings, together with the findings from these recent studies, suggest that the claim presented in the Bruce and Young (1986) model that sex and identity are unrelated, needs revision.

We noted above that effects of CP to changes in facial identity and sex together might emerge more readily than to changes in sex only, since different individuals are less similar than different sexes of the same face. This raises the issue of the role of item similarity in CP. In a previous study with familiar objects, we found that effects of CP were correlated with the degree of similarity between objects (Newell & Bülthoff, 2002). We reasoned that CP emerges for highly similar but subtly different objects because it is precisely these object types that the visual system learns to discriminate. For example, we may need to learn to discriminate different types of bottles, but the differences between a bottle and a lamp are readily perceivable. When this learning can occur, effects of CP emerge. Faces, however, might be a different story: Given that all faces represent a class of objects that are remarkably similar, it might not be surprising that effects of CP emerge in some face tasks because distinguishing between individual faces requires learning subtle differences between the features. For unfamiliar face at least, some faces, are more discriminable from each other and their distinguishing features may be more rapidly learned than the features of other more similar faces. Interestingly, a recent study reported that the similarity between faces is indeed related to effects of CP: Angeli, Davidoff, and Valentine (2001) reported that categorical perception of identity was not present when both endpoint faces were typical faces but effects of CP did emerge for distinctive faces. Since typical faces are more similar to each other and distinguishing features are more subtle (Valentine & Bruce, 1986), it follows that effects of CP would emerge more rapidly for distinctive than typical faces. Accordingly, effects of CP should emerge to typical faces after training or familiarization. It would be interesting to investigate how facial distinctiveness interacts with face familiarity in the emergence of effects of CP.

Finally, we noticed that effects of CP coincided with improved categorization performance for the female faces in our experiments. For example, in Experiment 1, endpoint female faces were correctly categorized as female 70% of the time, whereas after training categorization performance is close to 100% (Experiment 2). A similar but smaller improvement (a little over 10% difference) is found between Experiments 3 and 4. We may ask, therefore, if effects of CP emerge as a result of good representations of sex at the endpoints, and not to familiarization *per se*. To investigate this point, we conducted an experiment (not reported here) using sex continua with endpoint faces which were the average male and average female face derived from our database of 100 male and 100 female faces. Here we found that categorization responses to the average female and male faces were both close to 100%. Therefore, the endpoint faces were good representative faces of each sex even though these faces were unfamiliar to our participants. Nevertheless, we failed to find an effect of CP, suggesting that familiarity with the faces, and good sex discrimination, are the keys to CP effects.

In sum, our findings suggest that effects of CP for the sex of unfamiliar faces do not occur unless the observer has learned these faces (see also Bülthoff et al., 1998; Newell, Bülthoff, Vetter, & Bülthoff, 1998, for similar findings with face stimuli varying in texture, shape, or orientation). Once the faces are familiar, however, effects of CP for the sex of the faces are found. These results contribute to a growing body of research suggesting that facial identity or familiarity and the sex of the face are linked (Ganel & Goshen-Gottstein, 2002; Goshen-Gottstein & Ganel, 2000; Rossion, 2002). The classic functional model of face processing proposed by Bruce and Young (1986) needs revising in order to account for these findings.

REFERENCES

Abdi, H., Valentin, D., Edelman, B., & O'Toole, A. J. (1995). More about the difference between men and women: Evidence from linear neural networks and the principal component approach. *Perception*, *24*, 539–562.

Adamson, S. E., & Sowden, P. T. (2000). Induced categorical perception of novel, real-world, complex stimuli. *Perception*, *29*(Suppl.), 91c.

Adamson, S. E., & Sowden, P. T. (2001). Does induced categorical perception of novel, real-world, complex stimuli result from perceptual learning? *Perception*, *30*(Suppl.), 102c.

Angeli, A., Davidoff, J., & Valentine, T. (2001). Distinctiveness induces categorical perception of unfamiliar faces, *Perception*, *30*(Suppl.), 58.

Baudoin, J. Y., & Tiberghien, G. (2002). Sex is a dimension of face recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 362–365.

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

Blanz, V. (2000). *Automatische Rekonstruktion der dreidimensionalen Form von Gesichtern aus einem Einzelbild*. Unpublished doctoral dissertation, Eberhard-Karls-Universität Tübingen, Germany.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *SIGGRAPH 99 Conference Proceedings* (pp. 187–194). New York: ACM Press/Addison-Wesley.

Bornstein, M. H. (1987). Perceptual categories in vision and audition. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition*. Cambridge, UK: Cambridge University Press.

Bornstein, M. H., & Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research*, *46*, 207–222.

Boynton, R. M. (1979). *Human colour vision*. New York: Holt, Rinehart, & Winston.

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., Coombe, A., Fright, R., & Linney, A. (1993). Sex discrimination: How do we tell the difference between male and female faces? *Perception*, *22*, 131–152.

Bruce, V., Ellis, H., Gibling, F., & Young, A. (1987). Parallel processing of the sex and familiarity of faces. *Canadian Journal of Psychology*, *41*, 510–520.

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

Bülthoff, I., Newell, F. N., Vetter, T., & Bülthoff, H. H. (1998). Is the gender of a face categorically perceived? *Investigative Ophthalmology and Visual Sciences*, *39*(suppl.), 171.

Burns, E. M., & Ward, W. D. (1978). Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, *63*, 456–68.

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

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

Campanella, S., Chrysochoos, A., & Bruyer, R. (2001). Categorical perception of facial sex information: Behavioural evidence and the face–space metaphor. *Visual Cognition*, *8*, 237–262.

Campanella, S., Hanoteau, C., Seron, X., Joassin, F., & Bruyer, R. (2003). Categorical perception of unfamiliar facial identities, the face–space metaphor and the morphing technique. *Visual Cognition*, *10*, 129–156.

Campbell, R., Brooks, B., deHaan, E., & Roberts, T. (1996). Dissociating face processing skills: Decisions about lip-read speech, expression and facial identity. *Quarterly Journal of Experimental Psychology*, *49A*, 295–314.

Cheng, Y., O'Toole, A. J., & Abdi, H. (2001). Classifying adults' and children's faces by sex: Computational investigations of subcategorical feature encoding. *Cognitive Science*, *25,* 819–838.

Clemente, C. D. (1985). *Gray's anatomy of the human body* (30th ed.). Philadelphia: Lea & Febiger.

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

De Gelder, B., Teunisse, J., & Benson, P. J. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition and Emotion*, *11*, 1–23.

De Valois, R. L., & de Valois, K. K. (1975). Neural coding of colour. In E. C. Carterette & M. P. Friedman (Eds.), *Seeing: Handbook of perception* (Vol. 5, pp. 117–166). New York: Academic Press.

Eimas, P. D., Miller, J. L., & Jusczyk, P. W. (1987). On infant speech perception and the acquisition of language. In S. R. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 161–195). Cambridge, UK: Cambridge University Press.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin*, *115*, 268–287.

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

Ganel, T., & Goshen-Gottstein, Y. (2002). Perceptual integrality of sex and identity of faces: Further evidence for the single-route hypthosesis. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 854–867.

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

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*, 27–43.

Goldstone, R. L., & Steyvers, M. (2001). The sensitisation and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*, 116–139.

Goshen-Gottstein, Y., & Ganel, T. (2000). Repetition priming for familiar and unfamiliar faces in a sex-judgement task: Evidence for a common route for the processing of sex and identity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1198–1214.

Harnad, S. R. (Ed.). (1987). *Categorical perception: The groundwork of cognition*. Cambridge, UK: Cambridge University Press.

Hill, H., Bruce, V., & Akamatsu, S. (1995). Perceiving the sex and race of faces: The role of shape and colour. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *261*, 367–373.

Humphreys, G. W., Donnelly, N., & Riddoch, M. J. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces: Neurological evidence. *Neuropsychologia*, *31*, 173–181.

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

Lawrence, D. H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology*, *39*, 770–784.

Lawrence, D. H. (1950). Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *Journal of Experimental Psychology*, *40*, 175–188.

Le Gal, P. M., & Bruce, V. (2002). Evaluating the independence of sex and expression in judgements of faces. *Perception and Psychophysics*, *64*, 230–243.

Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, cross-race faces, and inverted faces. *Perception and Psychophysics*, *62*, 386–401.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.

Livingston, K., Andrews, J., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 732–753.

Medin, D. L., & Barsalou, L. W. (1987). Categorization processes and categorical perception. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 455–490). Cambridge, UK: Cambridge University Press.

Newell, F. N., & Bülthoff, H. H. (2002). Categorical perception of familiar objects. *Cognition*, *85*, 113–143.

Newell, F. N., Bülthoff, I., Vetter, T., & Bülthoff, H. H. (1998). Effects of shape and texture on the perceptual categorization of gender in faces. *Investigative Ophthalmology and Visual Sciences, 39*(Suppl.), *S*173.

O'Toole, A. J., Edelman, S., & Bülthoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research*, *38*, 2351–2363.

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

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–349.

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

Schweinberger, S. R., & Soukup, G. R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1748–1765.

Troje, N. F., & Bülthoff, H. H. (1996). How is bilateral symmetry of human faces used for recognition of novel views? *Vision Research*, *38*, 79–89.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology*, *43A*, 161–204.

Valentine, T., & Bruce, V. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception*, *15*(5), 525–535.

Vetter, T., & Poggio T. (1997). Linear object classes and image synthesis from a single example image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*, 733–742.

Wild, H. H., Barrett, S. E., Spence, M. J., O'Toole, A. J., Cheng, Y. D., & Brooke, J. (2000). Recognition and sex categorization of adults' and children's faces in the absence of sex stereotyped cues. *Journal of Experimental Child Psychology*, *77*, 261–299.

Young, A. W., Newcombe, F., deHaan, E. H. F., Small, M., & Hay, D. C. (1993). Face perception after brain injury: Selective impairments affecting identity and expression. *Brain*, *116*, 941–959.

Young, A. W., Rowland, D. A., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett D. I. (1997). Facial expression megamix: Test of dimensionality and category accounts of emotion recognition. *Cognition*, *63*(3), 271–313.

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?* Boulder, CO: Westview Press.

Original manuscript received January 2003
Revised manuscript received February 2004

APPENDIX

If each 3-D face of the database is described by a vector containing the geometrical and textural data (see Blanz & Vetter, 1999, for details), each face will have a location vector in a multidimensional space called face space (Valentine, 1991). Each dimension of this face space represents a different facial characteristic. Within this continuous space, it is possible to traverse a path from any face to any other face, morphing through locally similar faces along that path. In Figure A we show a very simplified 2-D representation of that face space with faces segregated by sex on each side of the sex boundary. The average of all faces is the neutral or average face that is at the centre of the face space subtended by the faces of the database. The average male face and average female face are placed correspondingly. The mean sex differences between male and female faces are represented by the *sex vector* joining the average female face to the average male face. By manipulating the position of a face along, what we term, its sex trajectory given by the sex vector, one can systematically vary this face's sex alone. Manipulations of six original female faces along the sex vector were done for Experiment 3. The sex of any individual face (e.g., Heidi) can be manipulated along the face's sex trajectory to the corresponding masculinized face. Note that the characteristic features (identity) of this face remains constant along this vector. Individual faces were manipulated along the sex vector in Experiment 3 and Experiment 4. By contrast a *sex and identity trajectory* is shown between two faces of different identities (here Heidi and Kurt). Manipulations of individual faces along a sex and identity vector together were done for Experiment 1 and Experiment 2.

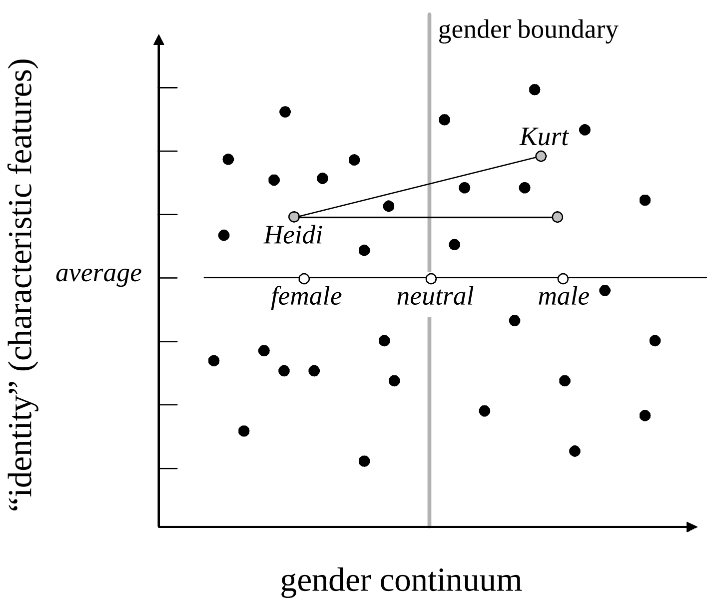

Figure A. An illustration of face space and manipulations of face sex.