Perception, 1993, volume 22, pages 131–152

Sex discrimination: how do we tell the difference between male and female faces?

Vicki Bruce¶, A Mike Burton¶, Elias Hanna, Pat Healey, Oli Mason
Department of Psychology, University of Nottingham, Nottingham NG7 2RD, UK
Anne Coombes, Rick Fright, Alf Linney
Department of Medical Physics and Bioengineering, University College London, London WC1E 6BT
Received 8 October 1991, in revised form 13 April 1992

Abstract. People are remarkably accurate (approaching ceiling) at deciding whether faces are male or female, even when cues from hairstyle, makeup, and facial hair are minimised. Experiments designed to explore the perceptual basis of our ability to categorise the sex of faces are reported. Subjects were considerably less accurate when asked to judge the sex of three-dimensional (3-D) representations of faces obtained by laser-scanning, compared with a condition where photographs were taken with hair concealed and eyes closed. This suggests that cues from features such as eyebrows, and skin texture, play an important role in decision-making. Performance with the laser-scanned heads remained quite high with 3/4-view faces, where the 3-D shape of the face should be easiest to see, suggesting that the 3-D structure of the face is a further source of information contributing to the classification of its sex. Performance at judging the sex from photographs (with hair concealed) was disrupted if the photographs were inverted, which implies that the superficial cues contributing to the decision are not processed in a purely 'local' way. Performance was also disrupted if the faces were shown in photographic negatives, which is consistent with the use of 3-D information, since negation probably operates by disrupting the computation of shape from shading. In 3-D, the 'average' male face differs from the 'average' female face by having a more protuberant nose/brow and more prominent chin/jaw. The effects of manipulating the shapes of the noses and chins of the laser-scanned heads were assessed and significant effects of such manipulations on the apparent masculinity or femininity of the heads were revealed. It appears that our ability to make this most basic of facial categorisations may be multiply determined by a combination of 2-D, 3-D, and textural cues and their interrelationships.

1 Introduction

In recent years psychologists have made significant progress at understanding, in broad functional terms, the relationships between different aspects of face processing (such as expression analysis versus recognition of identity) and between different stages of person identification [eg see Bruce and Young (1986) and Young and Bruce (1991) for reviews]. However, we are still some way from understanding the nature of the visual information which is encoded from faces to form the basis of the subtle discriminations we are able to make from these complex patterns.

Although the relative salience of different parts of the face which allows such discriminations to be made has been investigated [eg see Shepherd et al (1981) for a review], these studies have been limited in a number of ways. First, much research on feature salience was devoted to matching unfamiliar faces, yet it has been shown that the processing of familiar faces involves representations weighted more towards internal and away from external features compared with matching unfamiliar faces (Ellis et al 1979). Second, in much research on feature salience full-face views have been used and the effects of concealing or displacing face features have been studied. The general finding is that, of the internal features, eyes are more important than mouths, which are in turn more important than noses. However, the shape of the nose is not well depicted by a full-face image so perhaps this result is not surprising.

¶ Now at University of Stirling, Stirling, FK9 4LA, UK.

Third, and most important, the pattern of effects may vary from face to face (Haig 1986) and from task to task (Roberts and Bruce 1988) in ways that perhaps render general conclusions about feature salience somewhat suspect.

In this paper, we illustrate the complexity of the information that may be derived from (caucasian) faces to allow (predominantly caucasian) people to make a single, straightforward classification into 'male' or 'female'. How do we arrive at such decisions from the information provided by the face? In our culture, as in many others, there are obviously cosmetic and other fashion artefacts, over and above clothing, that contribute to this decision. Hair length and hair style are the most obvious ones, facial jewellery and cosmetics also vary, and additionally even clean-shaven males may have visible beard 'shadow' (indeed this is currently a fashionable attribute). It is possible that it is these 'superficial' cues alone that are used by viewers when deciding the sex of the face, but a simple demonstration conducted in our laboratory suggests that this cannot be the case.(1)

Photographs of 88 female and 91 male adults (aged 18 to 30 years) were taken in full face, with a neutral expression, and with each wearing a swimming cap to conceal hair. No makeup was worn by the females and the males were clean-shaven (and had shaved recently). Prints of these pictures were made and the faces cut out from the background to remove all clothing and any stray wisps of hair that might otherwise have been visible. Masking fluid was painted onto the earlobes of all faces to conceal all earrings (some photos had been taken with studs left in inadvertently) and ear-piercing holes. The photographs were given to 13 subjects unfamiliar with the faces shown in the photographs and the subjects were asked to sort, without time constraints, the photographs into two piles according to sex.

Performance was highly accurate (96.0% correct). The majority (3/4) of the errors were made in judging female faces as male: overall accuracy for female faces was 93.8% compared with 98.2% with the male. Clearly, then, the removal of major superficial cues leaves additional information that is sufficient for highly accurate discrimination.

What might the residual information be? There are two related ways of tackling this question. The first way is to embark upon a detailed analysis of the actual physical dimensions along which male faces differ from female faces, in order to examine the sources of information that the human visual system could use. This approach is taken in a companion article (Burton et al 1993). The second method is to investigate further the effects of removing or reducing information sources that might be used by human vision to make these judgements, and draw conclusions about what information sources are used on the basis of the impact of such manipulations. It is this second approach which forms the basis for the studies reported here.

We explored the uses made by the visual system of a number of different potential sources of information. We have already alluded to a distinction between superficial cues (hair, fashion accessories), and structural cues, which arise from the actual physical shape of the face.(2) We can also distinguish local from configural cues. A local cue might be the thickness of the eyebrow, or the width of the mouth, while a configural cue might be the length of the nose in relation to the width of the face.

---

(1) The task is described here, and in the companion article (Burton et al 1993) as discriminating sex, rather than gender, since the focus is on the classification of the physical pattern of the face into one of two biologically determined categories. We reserve the word gender for studies in which it is the psychological, rather than the physical, characteristics of masculinity and femininity that are of interest.

(2) We are not drawing any distinction beween cues that may arise primarily from secondary sexual characteristics (eg hairy eyebrows) and those which are primarily dictated by fashion (eg hair style), since there are in any case complex interactions between these (eg some women pluck their eyebrows; some men are bald).

Finally, there is a further potential distinction to be drawn between two-dimensional (2-D) and three-dimensional (3-D) cues. The length of the nose measured in the picture plane of a full-face view of a face is an example of a potential 2-D cue. The protuberance of the nose (how much it juts out at the bridge) is an example of a potential 3-D cue, though this is not to deny that such 3-D properties might be recovered from 2-D cues from shading in a full-face view, or from the angle of the bridge of the nose measured from a profile view. In order to investigate the roles played by these different potential sources of information, we explored the effects of removing or reducing different potential cues on the accuracy with which subjects could perform the task of sex discrimination.

Roberts and Bruce (1988) took this approach in investigating the features that were salient for judging the sex compared with those important to determine the familiarity of faces. Roberts and Bruce photographed clean-shaven male and female faces from media sources, and trimmed the photographs to remove the more obvious clues to sex available from hairstyle. Roberts and Bruce then assessed the effects of masking the eyes, the noses, or the mouths on the speed and accuracy of different types of decision made. In a task where subjects had to decide whether or not each face was familiar (half the faces were famous), they found that masks which concealed the eyes slowed familiarity decisions the most. When the task required subjects to decide whether each face was male or female, masks which concealed the noses slowed sex decisions the most. They related their results to Enlow's (1982) suggestion that females have a more concave muzzle compared with males, whose muzzle is more protuberant (the muzzle is the shape of the nose and brow combined). By showing, in a second study, that noses in isolation could not be categorised with above-chance accuracy, Roberts and Bruce suggested that it is the relationship between the size and shape of the nose and the other facial features which provides information useful for determining the sex of the face.

Roberts and Bruce's analysis suggests that potentially important roles are played by the overall configuration of facial features and by their 3-D shape in the perception of the sex of faces. In the first experiments reported here, we investigated more directly the role played by these and other factors in the judgement of the sex of faces. We achieved this by combining a further assessment of the effects of concealing different parts of the face with masks (cf Roberts and Bruce) with a novel investigation of the role played by superficial facial attributes compared with basic facial structure.

Although photographs of the type produced for experiment 1 remove many, they do not remove all, superficial cues which might be used to judge the sex of faces. There may be remaining superficial cues of skin texture (due to beard hair follicles; men may have coarser skins than women) and/or eyebrows (women often pluck eyebrows to remove hair between or beneath the eyebrows). In order to investigate the influence of these factors, we compared decisions concerning heads wearing swimming caps shown in photographs (ie as in our preliminary demonstration) with decisions concerning the same people's heads shown as accurate 3-D models devoid of their normal surface markings and texture.

This was achieved by using a laser-scanning device, developed at the Department of Medical Physics, at the University College London, to measure the surface of the human face and head [eg see Arridge et al (1985), Moss et al (1989), and Linney (1991) for reviews; and Bruce et al (1989) and Bruce et al (1991) for other perceptual research using laser-scanned data]. The major variable of interest in our second experiment was therefore whether heads were photographed directly, wearing swimming caps and with eyes closed (for comparability with the laser scans) or whether photographs were taken of the reconstituted 3-D surfaces displayed with the use of a computer-aided-design package following scanning by laser (see figure 1). In this paper, we label the first condition 'natural' (since the heads retain their natural superficial cues, albeit with caps and eyes closed), and the second condition 'laser'.

An attempt was also made in experiment 1 to identify which areas of the face might be more helpful for determining its sex, by systematically concealing different portions with masks designed to conceal the eyes, the nose, or the chin of each face. We were interested in learning whether any of these masks selectively disrupted decisons made about the sex of each face, and particularly whether the masks had differential effects depending on whether the faces shown preserved their normal superficial features (natural) or not (laser).

Because it is difficult to conceal exactly the same parts of the face when angled views are shown as when full-face views are shown, two separate experiments were conducted. In experiment 1 we used only 3/4 and profile views and compared the effects of concealing the eyes, nose, or chins. In experiment 2 we used full-face images and compared different types of mask applied to the area of the eyes and nose.

A final, but subsidiary, factor which was investigated in experiment 1 was the role played by the absolute size of the head in these judgements. Men, and hence their heads, are generally larger than women so that size could be used as a cue to sex. Although we did not think that such size factors were crucial, since people can easily judge the sex of individual faces in isolation, we nevertheless felt it was important to investigate this factor. Therefore, in experiment 1 we compared decisions made with regard to heads which were all scaled to present the same image size with those made with regard to heads whose sizes were left unscaled.

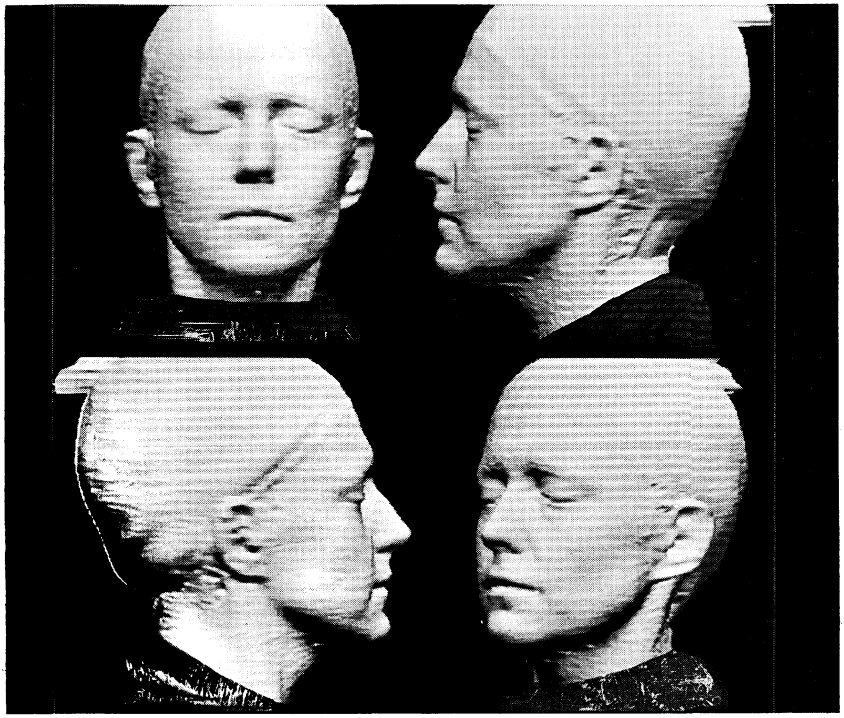

Figure 1. Examples of the representations of the shape of the head produced from measurement by a laser scanner and display of the multi-faceted surface with the use of Gouraud lighting. Four different views are shown of the head of an adult female, scanned while wearing a swimming cap.

2 Experiment 1

2.1 *Method*

2.1.1 *Materials, design, and subjects.* The faces used in this experiment were those of a set of 16 individuals (8 males and 8 females) aged between 21 and 35 years, all of whom had visited the UCL department and had their heads measured with the laser scanner. These people were also photographed wearing swimming caps and with their eyes closed (ie in the manner in which the scans were taken), with the use of top-lighting from a height and angle roughly comparable to the notional light source in the computer package which displayed the results of laser scanning. Details of the generation and display of these laser heads are given in Bruce et al (1989; 1991). Briefly, each head was represented by a set of some 20 000 coordinate points in three dimensions, which were joined to form a multifaceted surface which could be viewed at different angles. This set of facets was displayed as a smooth surface with the use of Gouraud lighting angled at 45° above the horizon line. Resulting images were copied onto slides.

For each head 32 different slides were prepared, produced from all combinations of two formats (natural or laser), two sizes (either 'normal' or 'scaled' to a standard size for all heads), two viewpoints (3/4 and profile), and four levels of mask. There was

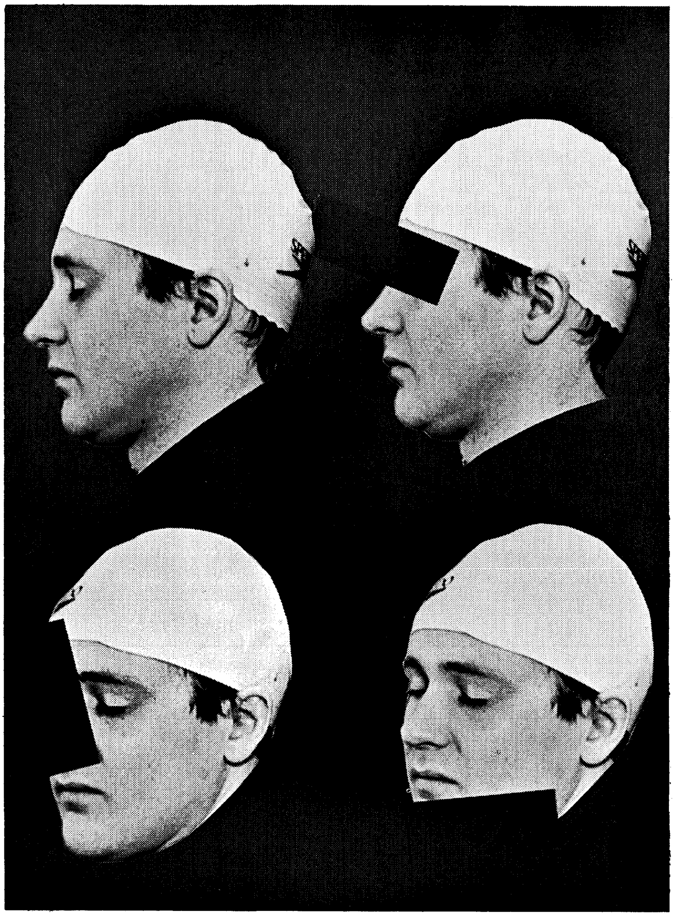

Figure 2. Examples of the 'natural' condition used for comparison with the laser condition in experiment 1, showing the eyes, nose, and chin masks used.

either no mask (ie the whole face was shown) or the same rectangular-shaped mask was overlaid on the face to conceal areas which included the eyes, the nose, or the chin (see figure 2).

The experiment was run as a mixed design with size and viewpoint varied between four groups of 20 undergraduate subjects, all of whom were unfamiliar with the faces shown. Format was varied within subjects, but an additional control factor was varied between groups so that each subject saw each head in only one format, but different heads were tested in different formats between subgroups of 10 subjects. Masking was varied within-subjects. Thus each subject viewed 64 slides, comprising 8 individuals (half of whom were males) shown in laser format and 8 (half of whom were males) shown in natural format with each individual appearing under four different mask conditions.

2.1.2 *Apparatus and procedure.* The experimental trials were preceded by four warm-up trials with faces which were not used in the main part of the experiment. Each face was preceded by a warning tone 500 ms before the slide appeared. The faces were then exposed for a maximum of 3000 ms or the slide was changed 1000 ms after the subject's response. Slides were presented with a Kodak Carousel projector controlled by an AIM microcomputer which also collected the reaction times (RTs) for each subject. Subjects were instructed to press one button if the face was male and another if the face was female, as quickly and accurately as possible.

2.2 *Results and discussion*

The mean correct response latencies (RTs) and associated error rates are shown for each of the major conditions of interest in tables 1 and 2. Table 1 summarises the overall effects of viewpoint and format; table 2 presents the detailed effects of

Table 1. Overall mean reaction times (in milliseconds) and errors (percentage, shown in parentheses) in experiment 1.

View and format	Male faces	Female faces
3/4 view		
laser	1 176 (16.6)	1 178 (15.6)
natural	971 (5.0)	935 (6.7)
Profile		
laser	1 203 (17.7)	1 354 (38.9)
natural	1 067 (12.3)	1 001 (13.6)

Table 2. Mean reaction times (in milliseconds) and errors (percentage, shown in parentheses) when different types of mask were used in experiment 1.

View and format	Male faces: mask type				Female faces: mask type			
	none	eyes	nose	chin	none	eyes	nose	chin
3/4 view								
laser	1 170	1 191	1 222	1 124	1 139	1 267	1 121	1 185
	(10.6)	(26.9)	(10.6)	(18.1)	(14.4)	(14.4)	(16.9)	(16.9)
natural	929	1 011	996	946	876	1 042	911	912
	(4.4)	(5.0)	(5.6)	(5.0)	(5.0)	(11.3)	(6.3)	(4.4)
Profile								
laser	1 219	1 187	1 213	1 194	1 299	1 463	1 398	1 255
	(11.9)	(21.3)	(17.5)	(20.0)	(33.8)	(45.6)	(40.0)	(36.3)
natural	1 017	1 152	1 041	1 058	962	1 098	962	982
	(10.6)	(16.9)	(12.5)	(9.4)	(10.0)	(14.3)	(15.6)	(14.4)

masking at each level of viewpoint and format. Since the scaling factor did not affect performance systematically (see below) this is not shown in these tables. Separate analyses were conducted on the latency and accuracy of responses. In our discussion we will concentrate on those effects which were found in both of these analyses.

2.2.1 *Latencies.* A 2 (scaling) × 2 (viewpoint) × 2 (format) × 4 (mask type) × 2 (sex of head) mixed design ANOVA was performed on the subject mean correct response latencies. This revealed a main effect of format ($F_{1,76} = 87.6$, $p < 0.001$), with laser heads responded to more slowly than natural heads. This main effect was qualified by an interaction between format and the sex of the heads ($F_{1,76} = 13.78$, $p < 0.001$), and between format, sex of heads, and viewpoint ($F_{1,76} = 6.9$, $p < 0.01$). As table 1 indicates, these interactions reflect the disproportionately slow RTs with respect to females shown in laser rather than natural format when the female heads are shown in profile. The only other significant effect was a main effect of masks ($F_{3,228} = 11.1$, $p < 0.001$), which reflects a slowing of responses when eyes, and to a lesser extent when noses, are concealed. The average RTs in the four mask conditions were 1076 ms (no mask), 1082 ms (chin masked), 1108 (nose masked) and 1176 (eyes masked). The interaction between masks and sex of head was inflated but not significant ($p > 0.05$) (see table 2) reflecting a tendency for nose masks to affect responses to male more than to female faces, and eye masks to produce the opposite effect. It is worth stressing that there was no significant effect of scaling of the heads, nor did this enter into any interaction terms. It appears that the normal variation in head size between the two sexes is not a useful cue in this task, as we anticipated.

2.2.2 *Accuracies.* As in the analysis of latencies, the initial 5-way ANOVA showed no significant effect of scaling, and indeed the non-significant trend went in the opposite direction to that predicted (ie scaled heads gave rise to somewhat more accurate performance). For this reason we pooled data from the scaled and non-scaled between-subjects conditions in order to simplify subsequent analyses and interpretations. The resulting 2 (viewpoints) × 2 (formats) × 4 (masks) × 2 (sex of head) ANOVA gave rise to a number of significant effects including the four-way interaction of all these factors ($F_{3,234} = 5.3$, $p < 0.001$). Because of the presence of this four-way interaction, the data were then analysed separately for 3/4 views and for profile views (ie separate ANOVAs, were conducted at each of the two levels of the second between-subjects factor of viewpoint, after having collapsed across levels of the first between-subjects factor of scaling).

In the analysis of data from the profiles alone, there was a significant main effect of format ($F_{1,39} = 46.4$, $p < 0.001$), with laser heads identified less accurately overall than natural heads. There was a significant main effect of sex of heads ($F_{1,39} = 30.6$, $p < 0.001$), with females responded to less accurately overall than males, and there was a significant interaction between format and sex ($F_{1,39} = 14.4$, $p < 0.001$). As seen in table 1, there is no apparent difference between the accuracy in responses to male and female faces in natural format, but a huge difference between them when they are shown in laser scans. Laser profiles of females gave rise to high error rates. The only other significant effect was a main effect of masks ($F_{3,117} = 6.53$, $p < 0.001$). The mean accuracy in each condition was 83.4% (no mask), 80.0% (chin masked), 79.0% (nose masked), and 75.0% (eyes masked). Thus the trend is for eye masks to reduce accuracy the most, consistent with the latency data. The effect of masks did not interact with other factors.

In the analysis of the data from 3/4 views alone there was again a significant effect of format ($F_{1,39} = 34.01$, $p < 0.001$), with laser scans producing less accurate performance, but there was no significant effect of sex of heads, nor any interaction.

Thus the disproportionate difficulty observed with the female heads in profile view does not occur with 3/4 views (consistent with our analysis of latencies). There was again a significant effect of masks ($F_{3,117} = 6.8$, $p < 0.001$), apparently entirely attributable to lower performance when the eyes were concealed. The only other effect was a significant three-way interaction between format, sex of head, and masks ($F_{3,117} = 6.59$, $p < 0.001$). This appears to be due to the particularly low accuracy associated with 3/4-view laser scans of male heads with eyes concealed (see table 2), which were correctly classified on only 73% of occasions. None of the other mask conditions varied so dramatically with format and sex of head, and we have no ready explanation for this particular dip in performance. We also note that the overall high error rate for the laser scan of male faces with eyes masked is not reflected in the response latency for this condition. It is possible that in this condition we have a speed-accuracy trade-off which also makes it unwise to overinterpret the observation.

Overall, however, the separate analysis of the profile and 3/4-view data for accuracy, combined with the analysis of response latencies, appears to give a reasonably clear picture. Laser heads are responded to less accurately than are natural heads, suggesting that the superficial characteristics preserved in the natural, compared with the laser, format do make some contribution to the information used to tell the sexes apart. However, this superficial information cannot be entirely responsible for performance, because performance with the laser heads in 3/4 view is generally fairly accurate (around 85%). Discriminating the sex of profile heads is more difficult than judging 3/4-view heads, and this is particularly evident in the laser heads. Since performance with natural as well as with laser heads declines in profile, it appears that some of the information needed to classify the sex of heads is less visible in this viewpoint. Profile views reveal less information about the eyes, eyebrows, and mouth than 3/4 views, though more about the shape of the nose. Information about the shapes of cheeks and other surfaces is less well shown in patterns of shading in profile views. Any or all of these factors could contribute to the decline in performance with profile views.

It seems then that the information that we use to discriminate between the sexes is conveyed in part by the overall structure of the head as preserved by the laser scans, and in part by information preserved in the natural photographs but not in the laser scans. The effect of masking out different parts of the images has generally not been dramatic in this experiment, suggesting that the discrimination of the sex of the head is determined by multiple and/or nonlocalised sources of information. The effects of masking which were observed suggest that, overall, it is information in the area that is concealed by the eye masks that is most informative, followed by that in the area concealed by the nose masks. Apart from the curious effect of eye masks on accuracy of judging 3/4-view male heads, the effect of masking appears to be relatively constant across format and sex of heads, suggesting that the main effect of the masking, where observed, is to conceal structural cues rather than 'superficial' cues present only in the natural format condition (eg if the eye masks concealed important information about the bushiness of the eyebrows, we would expect the effect to be more evident in natural format than in the laser condition, since information about facial hair is severely reduced in the laser format—see figure 1).

The greater effect of masking eyes compared with masking noses appears inconsistent with the effects reported by Roberts and Bruce (1988), who found that masking eyes had a much smaller effect on sex judgements than masking noses. However, Roberts and Bruce masked the eyes but not the brows with the eye masks they used, whereas their nose masks extended to the top of the nose (and hence occluded some information from eyebrows). In the experiment reported above, the nose masks did not conceal any portion of the brow region, although the eye masks did.

Thus the two studies are not really comparable. In the next experiment we made a further investigation of the effects of concealing different parts of the faces shown in natural or laser format, by using full-face images (as used by Roberts and Bruce), and by comparing the effects of eye masks and nose masks that either concealed or revealed the part of the face between the brows. In addition, the further (and major) aim of experiment 2 was to compare the efficacy of performance with laser and natural formats when faces are shown in full-face views.

Table 3. Overall reaction times (in milliseconds) and errors (percentage, shown in parentheses) in experiment 2.

Format	Condition	Head	
		male	female
Laser	no mask	1 007 (2.1)	1 317 (47.9)
	mask	1 131 (7.0)	1 346 (52.6)
Natural	no mask	977 (5.3)	1 028 (4.2)
	mask	1 150 (12.5)	1 104 (9.4)

Table 4. Effect of mask type on reaction times (in milliseconds) and errors (percentage, shown in parentheses) in experiment 2.

Format	Mask type	Head	
		male	female
Laser	eyes1	1 085 (3.1)	1 356 (58.3)
	eyes2	1 186 (8.3)	1 373 (55.2)
	nose1	1 124 (9.4)	1 338 (42.7)
	nose2	1 127 (7.3)	1 318 (54.2)
Natural	eyes1	1 161 (12.5)	1 169 (11.5)
	eyes2	1 127 (17.7)	1 184 (12.5)
	nose1	1 174 (11.5)	1 014 (4.2)
	nose2	1 139 (8.3)	1 047 (9.4)

3 Experiment 2

3.1 *Method*

The same 16 faces were used, this time in full-face scaled images throughout. Each face was produced in ten different versions. Half were in laser format and half were in natural format, and within each format there were five conditions of masking employed: no mask; a horizontal mask which covered the eyes and brows including the central brow at the top of the nose ('eyes1'); a horizontal mask of the same size but split into two halves so that the gap revealed the central brow at the top of the nose ('eyes2'); a vertical mask of the same size which concealed the nose and top of brow ('nose1'); and the vertical mask split into two so that the top of the brow was revealed ('nose2') (see figure 3 for examples). Subjects saw each face in each condition of mask, but which faces were shown in laser and which in natural format was counterbalanced between two groups of 12 subjects. The procedure was the same as for experiment 1.

3.2 *Results and discussion*

The subject mean correct latencies and associated error rates are reported in tables 3 and 4. Table 3 shows the overall effects of format and masking on this task, and table 4 gives a breakdown of the effects of each mask condition employed. As can be seen, the extraordinarily high error rates in recognising female faces in laser scans limit the conclusions that can be drawn from the analysis of latencies. We report here only the analysis of accuracies, and note that an analysis of latencies confirms the conclusions drawn.

A 2 (format) × 2 (sex of head) × 5 (mask) within-subjects ANOVA revealed main effects of format ($F_{1,23} = 129.6$, $p < 0.001$) with laser scans giving less accurate performance overall. The effect of sex was also highly significant ($F_{1,23} = 89.3$, $p < 0.001$), as was the interaction between format and sex ($F_{1,23} = 60.2$, $p < 0.001$). This interaction reflects the extreme inaccuracy with which subjects respond to the laser scans of the female heads in this experiment. There was also a main effect of mask condition ($F_{4,92} = 5.74$, $p < 0.001$) and an interaction between mask and sex of head ($F_{4,92} = 4.12$, $p < 0.01$). As can be seen from table 4, there is a great deal of variability in the effects of the different masks, but, on average, masking the eye region seems to be more detrimental than masking the nose. However, as noted as a trend in experiment 1, masking the nose region seems to have a greater effect on the identification of male than of female heads, while masking the eyes seems to have a greater effect on the identification of female than of male heads.

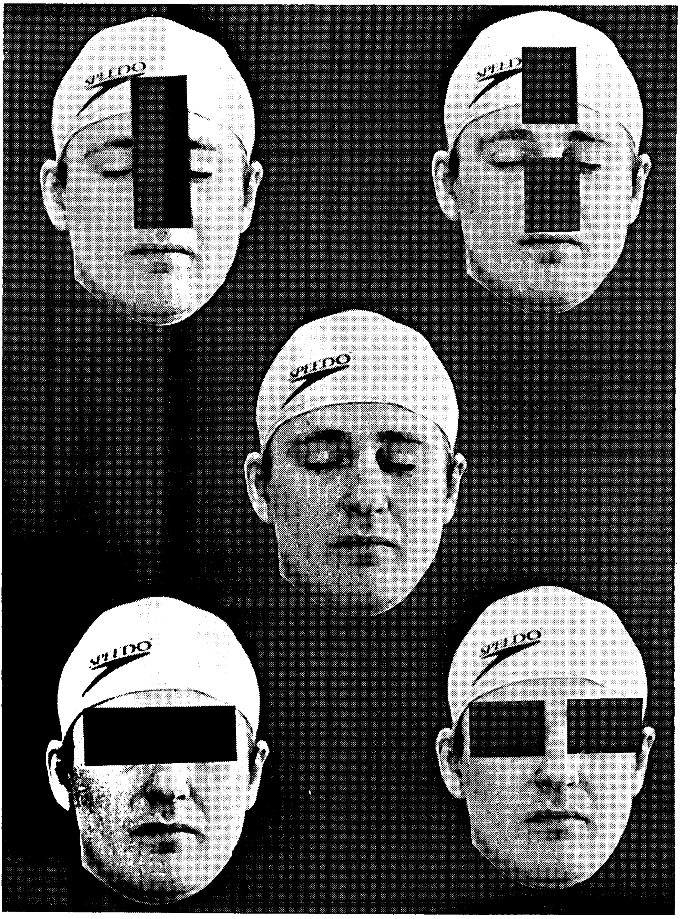

Figure 3. The four kinds of masks used in experiment 2. Top left: nose1; top right: nose2; bottom left: eyes1; bottom right: eyes2.

Rather than overinterpret what seem to be rather inconsistent effects of masks (since the accuracy data and latency data do not always show the same pattern) we conducted an analysis with items rather than subjects as random factor. This confirmed the main effects of sex ($F_{1,14} = 26.3$, $p < 0.001$), format ($F_{1,14} = 17.6$, $p < 0.001$) and their interaction ($F_{1,14} = 26.0$, $p < 0.001$), plus a main effect of type of mask ($F_{4,56} = 3.58$, $p < 0.05$), but no interaction between the effect of mask and other factors. The overall accuracy in responses to unmasked faces was 85.4%. Overall accuracy was affected very little by masks to the nose area, whether or not these concealed the brow region (84.7% and 82.6% respectively for nose1 and nose2 conditions), but was affected rather more by masking the eye region. Performance dropped to 79.4% for the eyes1 condition (where the mask concealed the brow region) and 76.6% for the eyes2 condition (where the mask revealed the brow region).

The most dramatic effect observed in this experiment was the extreme inaccuracy in judging the sex of the full-face females shown in laser scans, despite the accuracy with which the sex of full-face natural images can be determined. This suggests that superficial information—from eyebrows, skin texture and so forth—plays a considerable role in images where the 3-D shape will be hardest to see. (In a full-face view, information about the relative protuberance of regions such as nose and chin must be obtained from variations in the patterns of shading, whereas in angle views, the angles and lengths of visible contours can contribute.) It seems that the full-face laser heads look male, and this bias disadvantages identification of the female faces while marginally benefitting identification of the male heads (performance on the male heads was slightly more accurate with laser than with natural format). However, perhaps this bias is itself the result of superficial cues present in the laser scans but not the natural heads? For example, the laser heads look bald, whereas in the natural format the reason for this baldness (the swimming cap) is readily visible. Furthermore, the reconstruction of the images from laser-scan data introduces some wrinkling which could be interpreted as the rougher skin texture of a male rather than a female head. If responses to the laser heads are determined more by the addition of artefactual cues than the reduction of natural ones, then we will only be able to draw limited conclusions from these studies. However, it is important to note that such artefacts cannot account for the variation in performance across viewpoint. The 3/4-view laser scans look just as bald and wrinkly as the full-face and profile views, yet performance in the 3/4-view condition of experiment 1 was reasonably accurate (around 85%) and equally accurate for the female and male heads. There must therefore be additional sources of information to specify sex of the laser heads which are revealed best in 3/4 views and least in full-face views. It is only when these sources and natural superficial cues are reduced (as they are in full-face laser scans) that the 'male' bias becomes evident.

The effect of portraying the female heads in laser scans is much more dramatic than the effects of any of the masking conditions used in the experiment. In some ways this is surprising, since if superficial cues such as the shape or extent of the eyebrows were useful, we would expect their concealment to affect performance. Although the effects of the different masks were measurable, they were not large in size. This suggests that individual local cues to sex are perhaps not very useful compared with relationships between a multiplicity of local cues. Moreover, the dramatic effects of viewpoint observed in these experiments suggest that there may be a contribution from 3-D, as well as 2-D, configural relationships. The contributions of these sources is explored further in experiment 3.

The importance of relationships between different features may help explain why the visibility of the central brow region seemed to make little difference to performance in the task. We know from measurements made on faces that the gap separating the eyebrows is one of the better discriminators of facial sex (Burton et al 1993) so that might have led us to expect that natural format images preserving this information (eyes2, nose2) should be perceived more accurately than those in which this information is concealed (eyes1, nose1). Clearly there is no evidence for the use of this information in the performance observed in this experiment. Indeed the effects of masking in this experiment appear to be at variance with those reported by Roberts and Bruce (1988) who used full-face images where concealing the nose/brow region produced a dramatic time penalty in performance, though the latter remained fairly accurate. It is difficult to account adequately for this discrepancy. The experiment of Roberts and Bruce tapped fairly accurate performance (all faces had visible hair, though lengths were trimmed, and eyes were open) with effects of all mask conditions showing a similar relatively small disruption to the accuracy of sex judgement. The dramatic effect of masking the nose showed up only in latency differences. It is possible that the use of the laser-scan heads and virtual absence of cues from hair or eyes in any of the images used here may have promoted somewhat different strategies. Furthermore, the time penalties in the study of Roberts and Bruce may have resulted as much from disruption of a particular looking strategy as from the inaccessibility of useful information. The fact that Roberts and Bruce used media faces may also have made cues from the brows particularly informative in that task, since actors and actresses might be expected to exaggerate features associated with masculinity and femininity (eg actresses may pluck their eyebrows more than the sample of females whose heads were used in the current experiments).

Whatever the cause of this discrepancy, however, the effects of concealing parts of the faces in experiments 1 and 2 have been relatively small, suggesting that information resides as much in the relationship between different attributes of the face as in the attributes themselves. A further conclusion to be drawn from experiments 1 and 2 taken together is that there appears to be some role played by the 3-D structure of the face in judging its sex. Performance with the laser heads was very much better in 3/4 view, where the 3-D structure would be most visible, since the 3/4 view provides strong shading patterns around the brows and cheeks plus visible angled contours of nose and chin which may specify 3-D shape. In the absence of superficial cues, it seems that an angle view may be needed to reveal the 3-D structural differences which can be used to discriminate male from female heads.

Finally, we note that the disadvantages of presenting faces as laser scans show up particularly with female heads. A similar phenomenon was demonstrated by Bruce et al (1991) in the recognition of identity from laser scans, where from laser scans females were much harder to identify than males, but there was no difference between the two sexes shown in natural format (wearing swimming caps, as in the present experiments). The effect on identification was not in itself mediated by the misperception of the sex of the female heads, since the female laser heads were much more poorly identified than the male heads even when the task was to choose the identity from a set of possible names of the appropriate sex. Note that the poor judgement of female heads in the tasks described here and in Bruce et al (1991) cannot be directly attributed to the absence of hair in the laser scans, since it is also absent in the natural comparison condition. Bruce et al (1991) discuss possible reasons for the catastrophic decline in performance with female faces when superficial cues in addition to hairstyle are removed.

4 Experiment 3

4.1 *Introduction*

The next experiment pursues the idea that the determination of the sex of faces relies in part on relationships between different local cues or features, and in part on information about the 3-D structure of the face. If relationships between different aspects of the face are important, then we would expect sex discrimination to be impaired by inverting faces (presenting them upside down). The results of a large number of studies now suggest that inversion selectively disrupts configural processing, hence face processing is particularly disrupted by inversion [see Valentine (1988) for a review]. For example, Young et al (1987) made composite faces from the top and bottom halves of faces of two different celebrities. When presented upright, subjects found it very difficult to name the top halves when the composites were made so that the two face-halves were closely aligned. It was as though a 'new' identity emerged from the configuration provided by the two separate halves. However, when the composite was inverted, subjects' ability to name the separate halves improved, an effect attributed to the disruption of the configural processing which created the new identity in the upright condition. In contrast, other workers have shown that inversion appears not to affect the perception of isolated face features (eg Endo 1986). If subjects in the current sex discrimination experiments were making use of simple local features alone (eg 'hairiness' of eyebrow, or 'stubbliness' of cheek) then we might expect rather little effect of inversion, which might make subjects a bit slower to localise, and hence use, the information but should not necessarily affect their accuracy. If relationships between different features are important, then we would expect inversion to disrupt performance.

In experiment 3 we investigated the effects of inverting faces alongside effects of negation of the grey levels in the images. Faces shown in photographic negatives are extremely difficult to recognise (Phillips 1972) and recent research has revealed that there is a selective effect of negation on information conveyed by low spatial frequencies (Hayes et al 1986; Hayes 1988). The recognition of high spatial frequencies (Hayes et al 1986), or simple line-drawn images of faces (Hanna and Bruce 1992), is not impaired by negation, suggesting that negation does not impede the computation of simple measures of features or their interrelationships (which would be preserved in a line drawing). Rather, it seems that it is the reversal of the contrasts in an image containing varying grey levels which impairs performance. There are at least two sources of information which might be disrupted by negation. First, negation reverses pigmentation values, so that black hair appears white, for example. While such reversal of pigmentation would be expected to affect identification (a blonde becomes a brunette, and vice versa), it is difficult to see why it should affect the perception of whether a face is male or female. Second, negation may affect the computation of shape from shading by reversing the patterns of lighting (Phillips 1972). If 3-D shape contributes to our ability to distinguish male from female faces, and if this shape is determined to any degree by computing shape from patterns of shading, then presenting the faces in photographic negative should impair performance. To the extent that routines disrupted by negation (eg shape-from-shading) are independent of those disrupted by inversion (configural processing), we might expect the effects of these two variables to be additive. In a recent study of sensitivity to feature displacement in a face matching task, Kemp et al (1990) did indeed observe additivity of the two effects, consistent with independent processes. Experiment 3 investigated whether such additivity would also be observed in a sex classification task.

4.2 *Method*

4.2.1 *Subjects*. The subjects were two groups of 32 high-school students and teachers attending one-day university courses in Psychology. The majority of subjects in each group were female, and the majority were aged 16–18 years. There were approximately equal numbers of males in the two groups (8 in one group and 5 in the other) and equal numbers of subjects aged 20–45 years (8 in each group).

4.2.2 *Materials and design.* The test series comprised the faces of 6 males and 6 females wearing swimming caps, with eyes closed (the faces chosen were a subset of those used in experiments 1 and 2, to exclude the faces of any people who were teaching these subjects during the day's course). Each face was shown in photographic positive and negative, in each of three views—profile, 3/4, and full face—to form a test series of 72 slides in total. Thus face type (positive/negative) and viewpoint were varied as within-subjects factors. The slides were all presented upright to the first group of subjects and inverted to the second group. Thus face orientation was varied between subjects. The slides were arranged into a pseudorandom series such that no two successive slides showed the same person, and so that different instances of the same person were distributed throughout the series. The same slide order was then used for all subjects.

4.2.3 *Procedure.* Subjects were tested in two groups in the same lecture theatre, seated at different distances from the projected images. Subjects were shown three examples of faces in different viewpoints which included examples of each sex and each type (positive and negative); the faces shown were of people different from those appearing in the main series. The subjects were then asked to view each of the experimental faces and to write down whether they thought each was male or female, guessing if unsure. Each face was then shown for approximately 4 s and subjects wrote their decisions for each image.

4.3 *Results*

The group mean error rates are summarised in table 5, which shows the overall accuracy in the conditions of interest and these data broken down by viewpoint and sex. The accuracy with which upright positive images were identified agrees well with that reported elsewhere in this paper and, as in the previous studies, accuracy was higher for male than for female faces. Because of the very high accuracy with upright positive male faces (99.1% accurate overall), which meant there was little variance in these cells of the design, in this experiment the data from male and female faces were pooled before the analysis of variance, so that each subject's score was the accuracy over all 12 items contributing to each cell. Performance for each sex separately is nevertheless shown in table 5.

A 2 (orientation) × 2 (type of photograph) × 3 (viewpoint) mixed design ANOVA was conducted on the error rates in each condition and this revealed main effects of orientation ($F_{1,62} = 99.3$, $p < 0.001$), type ($F_{1,62} = 129.6$, $p < 0.001$), and viewpoint ($F_{2,124} = 3.3$, $p < 0.05$). No interaction terms approached significance (all other ps > 0.35).

Table 5. Accuracy (percentage wrong choices) of judgements of the sex of heads shown upright and inverted, in photographic positve and negative images.

Orientation	View	Male faces		Female faces		Faces of both sexes	
		positive	negative	positive	negative	positive	negative
Upright	profile	0.0	7.3	9.4	26.6		
	3/4	1.6	6.3	7.8	22.4		
	full face	1.0	4.7	5.2	25.5		
	overall					4.2	15.5
Inverted	profile	16.1	26.6	26.6	37.0		
	3/4	19.3	34.4	21.4	25.5		
	full face	19.8	12.0	10.9	45.8		
	overall					19.0	30.2

4.4 *Discussion*

Experiment 3 revealed highly significant and additive effects of inversion and negation on the ability to categorise the sex of faces about which there was little ambiguity in upright, positive views. The effects of inversion and negation are consistent with our suggestions that sex discrimination is determined by multiple cues. Configural processing is implicated by the effect of inversion, and 3-D processing is implicated by the effect of negation. Furthermore, the fact that performance remained comfortably above chance even when faces were both inverted and negated suggests that there is important information in high-spatial-frequency isolated details which is useful for at least some of the faces. Thus the experimental results are consistent with the use of information from three broadly different types in the determination of the sex of the face—local and/or superficial cues the perception of which should not be adversely affected by inverting or negating the images (eg hairiness or width of eyebrows), relational information (eg size of nose in relation to width of chin) and 3-D information (eg protuberance of brow).

Our suggestion that 3-D information may contribute to judgements of the sex of faces has been influenced by the observation in experiments 1 and 2 that it is only in 3/4 views that the female laser scans are classified with any degree of accuracy, combined with the detrimental effect of negation in experiment 3. In our final two experiments we investigate further the role of 3-D information by examining the extent to which judgements of masculinity or femininity of the 3-D heads can be influenced by changes to the local shape of the surfaces.

5 Experiments 4 and 5

5.1 *Introduction*

In section 1 we noted that, in addressing the question of how we tell the difference between male and female faces, a complementary method involves examining what are the actual physical differences. Burton et al (1993) show that in order to produce discriminant function analyses which approach human performance at distinguishing the sexes, 3-D as well as 2-D measures must be included.

In addition to exploring the adequacy with which a large set of measurements can discriminate statistically between the two sexes, an informal comparison of the 3-D structure of male and female heads has been made with the use of 3-D laser data. Fright and Linney (1992, 1993) describe a technique for averaging together sets of 3-D data from two or more heads. Using this technique we constructed an 'average' male and an 'average' female head by combining the laser scans of 9 adult males and 9 adult females, including the heads used in experiments 1 and 2 of this paper. Figure 4 shows the results of comparing the two average heads obtained in this way, and shows that the average male head has more protuberant nose, brow, chin/jaw, and upper neck (voice box) regions than the average female head. In turn, the average female head has somewhat more protrusive cheeks than the male. These observations lend some support for the claim of Enlow (1982) about the difference in shape of the 'muzzle' region. The greatest differences between the average male and average female heads were found in the region of the nose and chin. One of the interesting things about the results of this comparison is that these differences in relative protuberance would be rather poorly specified by 2-D measurements from full-face images of faces.

If the human visual system uses the relative protuberance of these regions in making judgements about the sex of the faces, then altering this should have some impact upon these judgements. In our final experiments we examine the effect of changing the shape of the nose and the chin on the rated masculinity of the 3-D heads.

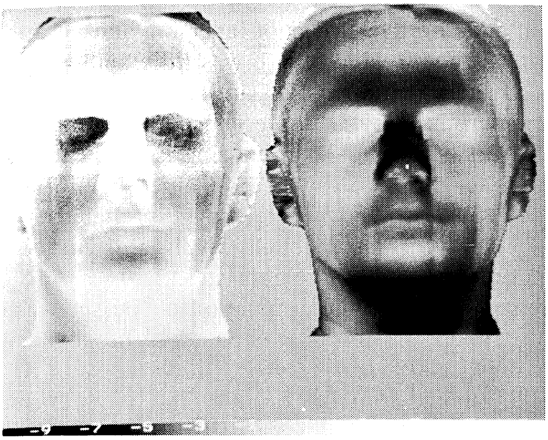

Figure 4. The 'average' male and 'average' female head, showing a comparison of their relative protuberance. Lighter shades indicate protuberance, darker shades indicate recession. To the left is the 'average' male compared with the 'average' female; Note the lighter brow, nose, jaw, and upper neck regions. To the right is the 'average' female compared with the 'average' male; note the somewhat lighter cheeks and eyes.

5.2 *Methods for feature manipulation*

Four female and four male heads were used. The noses of these heads were made either more hooked, or more retroussé, in shape by altering the gradient of the laser-measured profiles in the nose region by ±70%. The length of the section of profile to be changed was determined by selecting, interactively, the points on the profile that most closely corresponded to the turning point below the brow ridge, the tip of the nose, and the midpoint between the brows. The adjustment to the gradient in this region was applied to the midline profile in the centre of the nose, and three profiles to either side. No smoothing was necessary, as the natural reduction in the gradient of the surface itself in that region had the effect of progressively reducing the change on profiles away from the central one. The size of change effected by the ±70% change has been independently determined to be well above the 'threshold' amount of change needed for detecting a change in the shape of the nose. We refer to the +70%, normal, and −70% noses as the 'big', 'medium', and 'small' nose conditions.

Changes to chin shapes were produced in a similar manner, though to produce changes which were as visible as those to the nose, larger percentage alterations to the gradient were needed (because chins are naturally flatter). In the series used, the gradient on the chin was varied from −70% through to 120% and then 180% of the normal chin profiles. This produced the most satisfactory, readily visible, changes across a range of heads. By using +120% as the 'medium' chin we were able to produce versions which were either visibly flatter than this (−70%) or more protuberant (+180%). The chin changes were applied to 17 profiles—the central one plus 8 on either side. The control points were selected interactively and were those corresponding most closely to the tip of the chin, the point of inflection under the lower lip and a point approximately halfway between. We refer to the +180%, +120%, and −70% chin conditions as the 'big', 'medium', and 'small' chin conditions. Figure 5 shows examples of 'big' and 'small' nose and chin variations used in these experiments.

We should note that the methods used, and the resulting changes to the shapes of the noses and chins, were the best approximtion that we could make to the desired manipulations given the difficulty of making changes to these complex surfaces.

Work is in progress in our departments on making changes based more directly on the actual surface typology [cf Coombes et al (1991a, 1991b)]. The experiments we present here are preliminary explorations of the effects of surface change.

5.3 *Experiment 4: Method*

Each of the 8 heads was produced in four different versions by combining the large or the small chin with the medium nose, or the large or the small nose with the medium chin. We therefore assessed separately the effects of changing the nose (from small to large) in the context of medium values of the chin, and of changing the chin (from small to large) with medium values of the nose. Each head was produced in each of these four versions in two different viewpoints (3/4 and profile) to produce a series of 64 trials.

Heads were displayed on a Macintosh IIcx computer with the use of a program written in Supercard which displayed each head in turn in a different order for each subject. Subjects were asked to write down the code number of each face as it appeared, and then rate each face for its apparent masculinity or femininity on a scale of 1 to 10, where 1 was to be given to a very feminine face and 10 to a very masculine one. On completing the rating of a particular face, subjects requested the next face.

The subjects were 24 undergraduates, unfamiliar with the heads used in the experiment.

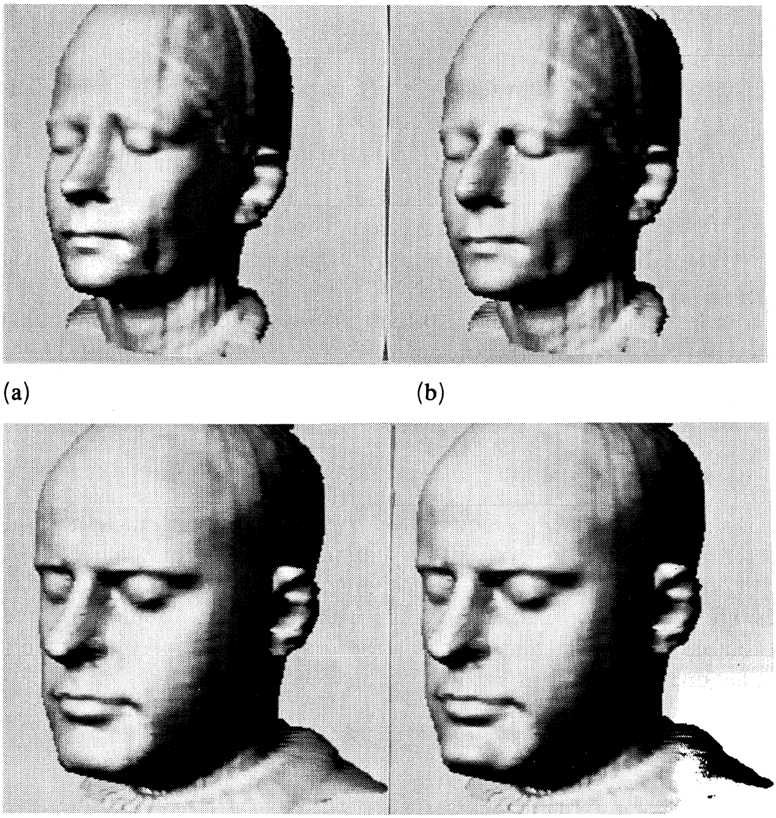

Figure 5. Examples of the nose changes applied to a female head and the chin changes applied to a male head. (a) a female head with 'small' nose (−70%) and (b) 'big' nose (+70%), (c) a male head with 'small' chin (−70%), and (d) 'big' chin (+180%).

5.4 *Experiment 4: Results*

The mean masculinity ratings in each of the conditions of interest are given in table 6. The effects of nose and chin changes were separately assessed from appropriate subsets of the data.

Table 6. The effect of varying the size of the nose and chin on mean judgements of masculinity rated on a scale from 1 (very feminine) to 10 (very masculine).

Condition	Male faces		Female faces	
	3/4 view	profile	3/4 view	profile
Big nose	8.7	8.2	5.8	4.4
Small nose	8.3	7.5	4.8	3.4
Big chin	8.6	8.1	5.3	4.0
Small chin	8.6	8.0	5.3	3.8

5.4.1 *Noses.* A 2 (set of heads) × 2 (viewpoint) × 2 (nose size) ANOVA was performed on the subjects' mean masculinity ratings in each condition. This revealed a significant main effect of sex of head ($F_{1,23} = 360.6$, $p < 0.001$) with subjects rating male heads as more masculine (mean 8.2) than female heads (mean 4.6). There was a main effect of viewpoint ($F_{1,23} = 61.25$, $p < 0.001$), with profile views rated more feminine (mean 5.9) than 3/4 views (mean 6.9). The effect of viewpoint interacted with effect of sex of head ($F_{1,23} = 6.75$, $p < 0.025$). From the means in table 6 it appears that the females were rated particularly feminine in profile. There was a main effect of noses ($F_{1,23} = 117.8$, $p < 0.001$). Big noses were rated more masculine (mean 6.8) than small noses (mean 6.0). This in turn interacted with sex of head ($F_{1,23} = 6.58$, $p < 0.025$), apparently because the bigger noses made females appear particularly masculine.

5.4.2 *Chins.* An ANOVA of the same design revealed a main effect of sex of head ($F_{1,23} = 450.1$, $p < 0.001$). Male heads were rated more masculine (mean 8.3) than female heads (mean 4.6). There was a main effect of view ($F_{1,23} = 42.0$, $p < 0.001$), with 3/4 views rated more masculine (mean 7.0) than profiles (mean 6.0). The effect of viewpoint interacted with that of sex of head ($F_{1,23} = 10.69$, $p < 0.001$): the females appeared particularly feminine in profile. There were no further significant effects. The size of the chin appears not to affect the apparent masculinity judgements ($F < 1$).

5.5 *Experiment 4: Discussion*

This experiment has shown that manipulating the shape of noses, though not the shape of chins, affected judgements of apparent masculinity in the predicted direction. This result in itself is interesting, since the 'feature salience' literature has tended to conclude that noses are virtually invisible in faces. Using 3-D representations and angled views we were able to confirm that changing the shape of the nose can make a highly significant impact on the impression created by the face.

The experiment also confirmed rating data effects that were revealed in experiments 1 and 2, where there was a tendency for laser scans of female heads to be misclassified as male, but not vice versa. The average rating with respect to male faces was clearly at the 'masculine' end of the rating scale. The average rating with respect to female faces was only just within the 'feminine' half of the scale. An apparently paradoxical finding in this experiment was that the laser heads were rated as more feminine in profile, despite our observations in experiments 1 and 3 that sex judgements are made more accurately for heads in 3/4 view than in profile. We know from experiment 1 that accuracy of judging female laser heads in profile is particularly low. Why then should these faces appear more feminine in profile? To investigate this, the data were re-analysed by items. A 2 (sex of heads) × 4 (types of nose and chin) × 2 (viewpoint) mixed design ANOVA was conducted on the mean ratings with respect to each head in each condition. This revealed significant effects of sex of heads ($p < 0.01$) and nose/chin condition ($p < 0.01$), but the effect of viewpoint just failed to reach significance. This is because only some of the items were rated more feminine in profile. The two female heads which were rated most feminine in 3/4 view were in fact given somewhat more masculine ratings in profile view. It was the other two female heads whose sex is clearly much more ambiguous in laser scans which were largely responsible for the overall effect of viewpoint. These two heads were given ratings of 5.97 and 8.21 respectively in 3/4 view, and only 3.32 and 5.47 in profile. Of the male heads, the most masculine was given the same ratings (8.9) in 3/4 view and profile, while the other three were all rated as somewhat more masculine in 3/4 view. Overall, then, the ratings given to the majority of items are not inconsistent with the overall advantage given to 3/4 views over profiles shown in experiments 1 and 2. The profile views for the majority of items elicit a rating which is more ambiguous than the rating given to 3/4 view.

5.6 *Experiment 5: Method*

The method was the same as in experiment 4, except that this time each head was produced in the four versions created by pairing big or small chins with big or small noses. A new sample of 24 subjects was tested.

5.7 *Experiment 5: Results*

The mean masculinity ratings in each of the conditions of interest are given in table 7. The effect of nose and chin changes were assessed in a single ANOVA.

A 2 (sex of head) × 2 (viewpoint) × 2 (noses) × 2 (chins) ANOVA revealed main effects of sex of head ($F_{1,23} = 227.6$, $p < 0.001$), with male heads rated as more masculine (mean 8.0) than female heads (mean 4.8). There was a main effect of viewpoint ($F_{1,23} = 117.3$, $p < 0.001$)—as in the previous experiment—with 3/4 views rated as more masculine (mean 6.9) than profiles (mean 5.9). There was a main effect of noses ($F_{1,23} = 49.9$, $p < 0.001$), with big noses rated as more masculine (mean 6.7) than small noses (mean 6.2). The effect of nose interacted with that of sex of head ($F_{1,23} = 7.6$, $p < 0.025$)—noses affected judgements relating to female heads more than those relating to male heads. There was a main effect of chins ($F_{1,23} = 24.7$, $p < 0.001$), with big chins rated as more masculine (mean 6.5) than small chins (mean 6.3). The three-way interaction between sex of head, viewpoint, and chin was also significant ($F_{1,23} = 14.0$, $p < 0.01$). The effect of chin change was most evident in profile views of males and 3/4 views of females. The interaction between viewpoint, nose, and chin just failed to reach significance ($p < 0.01$). The effect of chin change

Table 7. The effect of varying the sizes of the nose and chin together on mean judgements of masculinity rated on a scale from 1 (very feminine) to 10 (very masculine).

Condition		Males faces		Female faces	
		3/4 view	profile	3/4 view	profile
Small chin	small nose	8.4	7.1	5.0	3.8
	big nose	8.6	7.6	5.3	4.6
	mean	8.5	7.4	5.2	4.2
Big chin	small nose	8.2	7.7	5.2	4.0
	big nose	8.7	7.9	5.8	4.7
	mean	8.5	7.8	5.5	4.4

was large when combined with small noses in profile view but virtually absent when combined with small noses in 3/4 view.

5.8 *Experiment 5: Discussion*

This experiment replicated many of the effects seen in experiment 4, with a different set of subjects. The only difference was that experiment 5 revealed a significant effect of changing the shape of the chin, though the effect in terms of a change to the average ratings is rather small in size. Overall, the effects of changing the nose and changing the chin were additive in this experiment, though there was a trend towards an interaction of both these factors with viewpoint. It is too early to draw a strong conclusion that the shapes of these two regions make independent contributions to judgements of masculinity.

Overall it seems from experiments 4 and 5 that manipulating the shape of the nose and (to a lesser extent) the chin can have significant effects on the apparent masculinity of the face. However, these effects are not sufficient to shift the apparent sex of a face from male to female or vice versa. This again supports the view that it is a multiplicity of factors which combine to determine the sex of the face. Shape of nose and shape of chin make a small contribution to these multiple sources of information. However, we must caution that our conclusions about the effects of altering the shapes of the noses and chins are limited entirely to the effects of the particular shape manipulations we were able to achieve. The chin manipulation, in particular, does not really mimic the kind of protuberance in the jaw/voice-box region which is shown in figure 4. Future work with more adequate manipulations of surface shape may reveal stronger effects of changes to this region.

6 General discussion

In experiments 1 and 2, the importance of superficial cues to the sex of faces was revealed by the overall superiority of the natural images compared with the laser-scan images for classification of sex. However, the fact that the laser scans led to reasonably accurate performance in 3/4 views suggests that 3-D information may also contribute to the decision process, since 3/4 views provide both shading cues and contour/angle cues to the 3-D shape of the face. Experiment 3 also suggested the use of 3-D information, by showing significant decrements when images were shown in photographic negative. Effects of inversion in experiment 3, combined with the relatively small effects of individual masks in experiments 1 and 2, suggest that it is the relationships between different isolated features and/or measures that are important as well as local cues per se. Finally experiments 4 and 5 showed that changing the shapes of surfaces of the laser scans can alter their apparent masculinity or femininity, consistent with a role for 3-D as well as 2-D information in this task.

The experiments reported here provide converging information to show that important information about the sex of the face is conveyed by several broad (though not mutually exclusive) classes of information: (i) superficial and/or local features (such as facial hair, skin texture, eyebrows), (ii) configural relationships between features, and (iii) the 3-D structure of the face. When laser scans are presented in full face and profile, information from 3-D sources is minimised, and information from superficial features has been removed: what remains is largely the 2-D layout of features devoid of their surface textures etc. The result is very poor performance (75% average across male and female full-face laser scans). When photographs are presented in inverted photographic negatives, configural relationships and 3-D structure will be difficult to encode, but certain superficial and/or local characteristics remain potentially available (eg eyebrow thickness). The result again is that performance is severely disrupted (70% average accuracy across male and female inverted photographic negatives). Thus judgements of sex are above chance, but rather poor, when subjects are unable to rely on the multiple sources which are usually available.

The suggestion that people make use of 3-D information requires further examination. By 3-D information we mean here information about the undulations of the facial surface that is not captured by simple 2-D distances and dimensions from the full-face picture plane. This is not to say that there are no 2-D cues to such 3-D information. For example, the shape manipulations which were conducted in experiments 4 and 5 will have affected a variety of 2-D measures, such as the distance between the bridge of the nose and the tip of the earlobe, which are available from an angled view. Moreover, the variation in pattern of shading in a full-face image of a face is itself information which must be analysed in 2-D, and need not necessarily be used to produce an explicit 3-D representation. Our point is only that our evidence, combined with converging evidence reported by Burton et al (1993), suggests that more than simple 2-D distances and ratios in full-face images is needed to explain discrimination performance.

The results of these experiments fit rather well with our related work designed to examine which measurements actually discriminate between male and female faces (Burton et al 1993), where again we have found it necessary to incorporate both 2-D and 3-D measures into discriminant function analyses (DFAs) in order to produce performance approaching human accuracy. Of the 2-D measures which entered into the DFAs, some are local (width of eyebrows) while others are more relational (eg ratio of eyebrow length to the distance between the eyebrows). Since the sample of faces used in the current experiments is both smaller and different from that used by Burton et al (1993) we cannot draw direct comparisons between the results of the two studies (eg if distance between the eyebrows is the best discriminator of sex in a set of faces of undergraduates, this does not guarantee that it will be the most informative cue in a set of faces of 16 slightly older individuals who volunteered to be laser-scanned). Nevertheless we find the convergence of evidence encouraging.

What are the implications of this research for other aspects of face perception? Given the sheer complexity of information which appears to contribute to a relatively simple categorisation task which the visual system makes accurately and effortlessly, it is unwise to assume that the information used to make more subtle discriminations within categories will be any less complex. It has proved difficult to come up with good predictors of the 'memorability' or 'distinctiveness' of faces based on simple 2-D measures such as eye width or nose length, and simple ratios of such measures (eg Ellis and Shepherd 1987; Bruce et al, in press). We would speculate that this lack of predictive power arises not so much because the 'wrong measures' are being used, but because a very incomplete set of measures has been sampled. The visual system has available to it a much richer database of information than that summarised in a small set of distances between key landmark features. Perhaps it should not surprise us to learn that all the available information appears to be used.

Acknowledgements. This research was supported by grants from the SERC Image Interpretation Initiative (GR/F/33698 and GR/F/72178). We sincerely thank all those who have visited the UCL laser scanner to provide 3-D surface data for this project. We thank two referees, and Michael Morgan, for comments on an earlier version of this paper.

References

Arridge S R, Moss J P, Linney A D, James D R, 1985 "Three dimensional digitization of the face and skull" *Journal of Maxillo-facial Surgery* **13** 136 - 143

Bruce V, Burton A M, Dench N (in press) "What's distinctive about a distinctive face?" *Quarterly Journal of Experimental Psychology*

Bruce V, Burton M, Doyle T, Dench N, 1989 "Further experiments on the perception of growth in three dimensions" *Perception & Psychophysics* **46** 528 - 536

Bruce V, Healey P, Burton M, Doyle T, Coombes A, Linney A, 1991 "Recognising facial surfaces" *Perception* **20** 755-769

Bruce V, Young A W, 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" *Perception* **22** 153-176

Coombes A M, Linney A D, Richards R, Moss J P, 1991a "A method for the analysis of the 3D shape of the face and changes in the shape brought about by facial surgery" in *Biostereometrics, Technology and Applications* Ed. R E Herron, Proceedings of SPIE, Volume 1380, pp 180-189

Coombes A M,, Moss J P, Linney A D, Richards R, James D R, 1991b "A mathematical method for the comparison of three-dimensional changes in the facial surface" *European Journal of Orthodontics* **13** 95-110

Ellis H D, Shepherd J W, 1987 "Analysis of mechanisms by which human observers acquire knowledge of faces" Final Report on ESRC Grant C00232260, held at University of Aberdeen

Ellis H D, Shepherd J W, Davies G M, 1979 "Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition" *Perception* **8** 431-439

Endo M, 1986 "Perception of upside-down faces: an analysis from the viewpoint of cue saliency" in *Aspects of Face Processing* Eds H D Ellis, M A Jeeves, F Newcombe, A Young (Dordrecht: Martinus Nijhoff) pp 53-58

Enlow D H, 1982 *Handbook of Facial Growth* (Philadelphia, PA: W B Saunders)

Fright W R, Linney A D, 1992 "Registration of 3-D head surfaces using multiple landmarks" *IEEE Transactions on Medical Imaging* in press

Fright W R, Linney A D, 1993 "An average 3-D head surface" submitted for publication

Haig N D, 1986 "Investigating face recognition with an image-processing computer" in *Aspects of Face Processing* Eds H D Ellis, M A Jeeves, F Newcombe, A Young (Dordrecht: Martinus Nijhoff)

Hanna E, Bruce V, 1992 "Testing the efficacy of computer-generated cartoons" paper presented at IEE Colloquium on Machine Storage and Recognition of Faces, January 1992, IEE Digest No 1991:017

Hayes A, 1988 "Identification of two-tone images: some implications for high- and low-spatial frequency processes in human vision" *Perception* **17** 429-436

Hayes T, Morrone M C, Burr D C, 1986 "Recognition of positive and negative bandpass-filtered images" *Perception* **15** 595-602

Kemp R, McManus I C, Pigott T, 1990 "Sensitivity to the displacement of facial features in negative and inverted images" *Perception* **19** 531-543

Linney A, 1992 "The use of 3-D computer graphics for the simulation and prediction of facial surgery" in *Processing Images of Faces* (Northwood, NJ: Ablex) pp 149-178

Moss J P, Linney A D, Grindrod S R, Mosse C A, 1989 "A laser scanning system for the measurement of facial surface morphology" *Optics and Lasers in Engineering* **10** 179-190

Phillips R J, 1972 "Why are faces hard to recognise in photographic negative?" *Perception & Psychophysics* **12** 425-426

Roberts T, Bruce V, 1988 "Feature saliency in judging the sex and familiarity of faces" *Perception* **17** 475-481

Shepherd J W, Davies G M, Ellis H D, 1981 "Studies of cue saliency" in *Perceiving and Remembering Faces* Eds G Davies, H Ellis, J Shepherd (London: Academic Press) pp 105-131

Valentine T, 1988 "Upside-down faces: A review of the effect of inversion upon face recognition" *British Journal of Psychology* **79** 471-492

Young A W, Bruce V, 1991 "Peceptual categories and the computation of grandmother" *European Journal of Cognitive Psychology* **3** 5-49

Young A W, Hellawell D, Hay D C, 1987 "Configural information in face perception" *Perception* **16** 747-759