More about brows: How poses that change brow position affect perceptions of gender

Ruth Campbell, Philip J Benson†, Simon B Wallace #, Suzanne Doesbergh §, Michael Coleman
Department of Human Communication Science, University College London, Chandler House, Wakefield Street, London WC1, UK; † University Laboratory of Physiology, University of Oxford, Oxford, UK; # Institute of Psychiatry, University of London, London, UK; § Psychonomics Laboratory, University of Utrecht, Utrecht, The Netherlands; e-mail: R.Campbell@ucl.ac.uk
Received 30 April 1998, in revised form 18 November 1998

Abstract. The speeded categorisation of gender from photographs of men's and women's faces under conditions of vertical brow and vertical head movement was explored in two sets of experiments. These studies were guided by the suggestion that a simple cue to gender in faces, the vertical distance between the eyelid and brow, could support such decisions. In men this distance is smaller than in women, and can be further reduced by lowering the brows and also by lowering the head and raising the eyes to camera. How does the gender-classification mechanism take changes in pose into account? Male faces with lowered brows (experiment 1) were more quickly and accurately categorised (there was little corresponding ‘feminisation’ of raised-brow faces). Lowering gaze had a similar effect, but failed to interact with head lowering in a simple manner (experiment 2). We conclude that the initial classification of gender from the facial image may not involve normalisation of the face image to a canonical state (the ‘mug-shot view’) for expressive pose (brow movement and direction of gaze). For head pose (relative position of the features when the face is not viewed head-on), normalisation cannot be ruled out. Some perceptual mechanisms for these effects, and their functional implications, are discussed.

1 Introduction
Humankind shows sexual dimorphism in adult appearance. Men are, on the whole, taller and heavier than women, with a different distribution of body hair, of musculature, of facial appearance. To date, most experimental and metrical investigations of gender in facial appearance have explored the structural differences between men’s and women’s faces (Brown and Perrett 1993; Bruce et al 1993; Burton et al 1993; Ferrario et al 1993)—durable differences in appearance due, for the most part, to hormonally mediated differences in bone growth and skin texture. But as well as carrying critical structural information on gender, age, and attractiveness (Galton 1907; Perrett et al 1994; Bruce and Young 1998), the face is the primary visual signalling organ for emotion and for intention (Darwin 1872; Ekman 1979; Baron-Cohen 1995). Selection pressures must have affected the evolution of the face in both the perceiver and the perceived (Fridlund 1991). It is quite likely, therefore, that the way we interpret faces, and how face structures and their ‘readings’ interact, could indicate the evolution of an efficient perceptual mechanism.

Gender moderates all interactions and may often be the first and most important classification in any encounter with a new person. While hair, gait, and clothing may be reliable clues to gender, the face can be met without such support, as, for example, in photographs or video close-ups, or where a face is seen peering from a cloak or from undergrowth. On the basis of its salience, we might expect the classification of gender from faces to take account of face pose and face actions. It is therefore somewhat surprising that the perceptual classification of gender from active faces has not been explored to any extent.

A previous study (Campbell et al 1996) established that direction of gaze in a facial image affects the efficiency of gender decisions to that face. Faces looking down were
harder to classify for gender than those looking ahead. Moreover, male faces looking down were rated more feminine than the same faces looking ahead. A simple perceptual cue could account for this. This cue is reliable in discriminating sex in faces looking ahead, but is misleading when gaze is averted. It is the vertical distance between the brow and the upper eyelid (brow–lid distance), which is one among several of the most reliable structural cues to gender in photographed face images (Burton et al 1993). Brow–lid distance is smaller for male than for female faces when both are looking ahead, and increases when gaze is averted downwards. Thus, when gaze is averted, the increase in brow–lid distance can suggest a female face.

Are such effects limited just to changes in vertical direction of gaze or are they more general? We can change brow–lid distance by eyebrow plucking and, in this context, we have been unable to find reports that eyebrow plucking is used by men but not by women; suggesting that, as the very least, this cue to gender is accommodated in a universal way in grooming and cosmetic behaviour. But as well as this—and possibly in interaction with it—we can also easily change the distance between the lid and the brows by movement of the face and head. Vertical brow movement is one of the easiest actions to perform to command (Ekman 1979), and lowering the brows (contraction of the corrugator muscle, Ekman FACS action number 4) directly reduces brow–lid distance. Another easy manipulation that changes brow–lid distance is lowering the head while raising the eyes to camera. How is gender judgment of such face poses affected by such changes in pose? Does brow raising ‘feminise’ a face and brow lowering ‘masculinise’ it? Does lowering the head while looking at the camera, which should also reduce brow–lid distance in the photo image, ‘masculinise’ the face? These questions are primarily empirical, but behind them lie interesting questions concerning the processing of the facial image.

1.1 Classifying gender
It might be assumed that, in classifying face images, human observers preprocess the stimulus to a canonical facial view and disposition. This is exemplified in the ‘mug-shot’ portrait. The canonical facial view is full face,(1) neutral in expression, and posed with head and eyes to camera. It may follow that the template or prototype that informs classification of the face in terms of its structural qualities (for identity, gender, age, etc) requires preprocessing to this format to be accurate and robust. On this sort of account, any action of the head or face that deviates from the canonical view will take additional time and effort, and we would expect classification of gender from noncanonical face images to be less accurate than for the canonical face and require extra time to resolve. In addition, the extent of deviation from the canonical view should be a good predictor of extra processing effort. Our earlier studies in which vertical direction of gaze was manipulated did not address this question. Looking down conceals the eyes. While the eyes are not themselves gender dimorphic, nevertheless the fact that such faces took longer to classify may be because the face-classifying mechanism has to do extra work to ‘fill in’ for the missing feature.

1.2 An image-processing perspective
In contrast to a serial-stage model, where normalisation followed by matching to canonical template are discrete processing stages, perspectives from computer vision can offer a different approach. Here, the facial image itself is the source of all the necessary information for decision, and regularities across different images are captured by decision algorithms which (once registered appropriately) do not require normalisation to a prototype. There are a variety of ways in which face images may reveal appropriate

(1) It should be noted that the full-face view is not the most reliable for judging gender: three-quarter view, which affords more information about face structure when lit appropriately, can be at least as good—and sometimes better (see for example Hill et al 1995).
information for gender decisions. For example, at its very simplest, coarse horizontal frequency filtering of individual face photographs can generate a 'horizontal-bar-code' description of a face from which gender can be quite well categorised (Watt and Dakin 1993). Brow–lid distance could be extracted by relatively low-level visual mechanisms tuned to horizontal patterns of tonal contrast in faces. Watt (1994) describes how such spatial-frequency analysis may subserve image segmentation. Furrowed brows generate a thicker, lower, darker bar in this sort of analysis, while raised brows generate a thinner, lighter, higher bar. If such spatial-frequency filtering were used in categorising face images for gender, then any poses that reduce brow–lid distance could trick the mechanism into categorising a face as masculine, while raised brows could lead to a 'feminised' bar-code description.

On its own, a 'bar-code-filter' approach cannot be sufficiently flexible to lead to accurate coding of gender in faces, although it may provide a useful 'fast-and-dirty' first pass. At a different level of visual computation than spatial-frequency analysis and filtering, classifying sex from face images has proven amenable to statistical and to neural-net modelling. Principal component analysis (PCA) is a means of statistically compressing data, including face images input as pixels. Using a set of facial images as input, PCA derives a set of dimensions (principal components or eigenvectors), typically eight to fifteen in number, that are sufficient to reconstruct the whole set (see O'Toole et al 1994, 1998); typically, a set of 150 or so face images can be distinguished this way. If human recognition and categorisation of faces uses similar means, then the characteristics (eg eigenvectors and their eigenvalues) of such a system should correlate with human-performance measures. This has been demonstrated for face recognition (O'Toole et al 1994) and for the classification of gender (O'Toole et al 1998). These findings will be considered in more detail in section 6.3. At this stage it is sufficient to note that just the first three eigenvectors accounted for most of the gender classification.

A number of computational models of face categorisation have made use of PCA-based approaches together with specific decision and learning algorithms, often based on connexionist learning principles (neural-net algorithms). These are active 'automatic face-categorisation machines' which might offer insights into how human observers categorise faces. For example, Golomb and Sejnowski (1995; Golomb et al 1991) have developed a number of variants on a two-phase (image compression and sex classification) net: SEXnet, where output from the compression phase served as input to the classification phase. In one version, the neural net was trained on eighty faces by using standard back-propagation procedures, and was tested on ten. The images were input as tonal (256-grey-level) images of neutral—ie mug-shot—faces. SEXnet was successful at categorising gender in these images after initial training, making errors on those faces that human participants misclassified and correctly identifying one (deliberately) misclassified face. Both simple two-layer (perceptron) and three-layer (hidden-unit) networks performed well on this task: discrimination of sex from face images does not require powerful nonlinear computations. The decision architecture allows feedback from classified images to set the decision weights on the image characteristics (positions on the image of different tonal information) to improve classification performance. The 'representations' of face images in a trained net are patterns of weights on the mapped connections. The general principle of neural-net modelling emphasises that learning of categories develops sensitivity to the characteristics of the exposure set. These need not always be the characteristics that a structural analysis of the stimulus might suggest; associated qualities may be just as useful as those that are indicated by structural analysis. For instance, if the image-analysis net was trained on female faces that all happened to be blond, it would (mistakenly) learn blondness as a cue to gender classification.
SEXnet has not yet been applied to faces of different pose, but a simple prediction is that training on neutral poses and testing on brow-changed ones would trick the system in the direction of misclassifying smaller brow–lid distances as male, larger ones as female. But is this also true of human classification of gender?

1.3 Multiplicity of face meanings—interactive processes

The most compelling aspect of displays such as brow raising or lowering, or dipping the head and raising the eyes, is that they are communicative. Brow raising and furrowing have been explored as components of facial expressive acts (Ekman 1979). Furrowed brows form part of the ‘anger’ face, but can also signal ‘sadness’ and other intentional/motivational states such as ‘puzzlement’, while raised brows signal ‘surprise’ (Ekman 1979; Campbell et al 1999). We do not know whether the classification of such communicative face poses depends on the perceived gender of the face. The relationship between gender decision and expressive face movements may be interactive, with some interplay between the type of facial action signalled and the inferred masculinity of the image. The same may be true for vertical actions of the head and eyes. Such actions are constructed intentionally: they have a signalling function. A face with head lowered and gaze to camera can indicate momentary interest in the viewer. According to social context and the disposition of other face parts, aggressive or flirtatious intent (for example) may be communicated. Such poses, too, may be gender constrained: a head down–eyes up face in a woman may be construed as ‘flirty’, while the identical pose in a man’s face may be construed as ‘aggressive’.

Now while interactions of gender and of face pose may be evident where faces are rated for degree of masculinity in tasks which are not time stressed, the simple classification of gender as a speeded categorisation task may not be so sensitive to them. Indeed, a serial-stage model of classification could not allow such effects to occur, for in order to decide how masculine or feminine a face is as a function of its expressive pose, classification of the face as male or female has to happen first. However, an image-based approach, in which the learnt contingencies of categorisation determine skilled performance, could allow such interactions into the system at the learning stage.

1.4 Experimental predictions

Considerations such as these allow us to make some specific predictions about the relations between categorisation and rating of faces for gender as a function of their pose. We have outlined two different mechanisms for visual classification: one is a serial-stage process, involving precategorical procedures to normalise the image to a canonical prototype; the other is an image-based process which makes use of learned contingencies that discriminate gender in a training set.

On the basis of our previous study, reduction in brow–lid distance should be a powerful cue to male gender, while enhancement of brow–lid distance should cue female gender. Can the brow–lid-distance cue work under these different conditions where movement of the brows or head, rather than direction of gaze, is the experimental variable? The study is in two parts, reported in two sections. The first part concerns movements of the brows, the second movements of the head. Within each part, we first collected and analysed physiognometric data from people asked to pose facial acts to camera. These face images were then shown to viewers to categorise for gender. For speeded categorisation, the serial-stage model predicts that, if face images are normalised prior to the gender decision, then all deviations from the canonical face pose should be categorised less efficiently than that pose. However, an image-based model predicts that if brow–lid distance enters into these decisions there should be a systematic effect of brow movement: smaller brow–lid distance (lowered brow, lowered head) should bias classification (faster, more accurate) in favour of male faces, while greater brow–lid distance should lead to a similar effect for female faces.
2 Experiment 1a: Measuring brow–lid distance for raised and lowered brows

In this experiment we simply measured brow–lid distance in a set of male and female faces when the people posing the face actions were asked to raise and to lower their brows.

2.1 Stimuli

The faces of twelve male and twelve female participants, who had not taken part in our previous experiment (Campbell et al 1996), comprised the experimental set. All faces were of clean-shaven individuals ranging in age from 15 to 55 years and of a number of different ethnic groups. Makeup and face jewellery (nose rings, earrings) were permitted. Four of the sitters (two male) had nose studs; two females wore small ear studs. The sitters were video recorded in a day-lit room and the images were captured directly onto computer (IBM-PC). Each participant was asked to adopt a neutral expression for each pose and to look at the camera. Instructions to raise and to lower the brows were given by the experimenters, and shown by example to the sitters. Contraction of corrugator (FACS action number 4; Ekman 1979) lowered the brows, while contraction of the frontalis muscle (FACS 2, lateral part; FACS 1, medial part) accompanied by eyelid raise (oribicularis orii, FACS 5) raised the brow. Three experimenters (MC, RC, and SW) were present throughout recording and agreed the best image to be computer stored for each condition. Examples of images used are shown in figure 1.

The images were transferred to a Silicon Graphics R4000 Indy workstation and were analysed by using procedures developed for aligning and measuring facial landmarks (Benson and Perrett 1991). The vertical midline was first established by using the pogonion (the lowest point of the chin) and points equidistant between the inner canthi of the eyes and the nostrils, and the image was rotated to a vertical attitude. The distance between the upper eyelid and the lower outline of the visible eyebrow varies because the curvature of the brow is not parallel to the eyelid. Therefore the brow–lid distance was sampled at different points. Measures were made by using an interactive cross-hair cursor adapted for the experiment. Three vertical measurements were made from each eye area to the corresponding brow point, making six measurements for each face. These originated at the left and right outer canthus, the lid point above left and right pupil, where visible, and the left and right inner canthus. A further measure of eye width was taken which was used to normalise the vertical measures. This controlled for any variation in distance of the sitter from the camera, and possible small lateral differences in angle of the head. The measure used was the mean width of each pupil (measured on left and right eye). This is shown diagrammatically in figure 2.
2.2 Results
The minimum vertical distance between lid and brow was at the pupil; the maximum distance usually at the outer canthi. For each facial image the average value of the six vertical measures was divided by the mean of the two (left and right) eye-width measures to give a mean normalised distance measure for each face in each condition.

Means of these means for female and for male faces are shown in table 1. These were subjected to analysis of variance (SPSS – GLM procedure), in which the between-subject factor was gender of face and the within-subject factor was brow action (relaxed, raised, and lowered brow).

Brow–lid distance was greater in female than in male faces ($F_{1,22} = 4.02, p = 0.05$). Distance varied systematically with pose ($F_{2,44} = 153.0, p < 0.001$) and planned contrasts show each of the brow positions to be significantly different from the other ($p < 0.001$). There were no interactions of brow pose with gender.

These measurements confirm that brow–lid distance is changed in the expected direction by posed actions of the brows and that the distances are not moderated by the gender of the face. Furrowing the brows reduces brow–lid distance, while raising them increases it.

### Table 1. Experiment 1a: mean normalised vertical brow–lid distances, with standard deviations in parentheses, over the twenty-four stimulus faces.

<table>
<thead>
<tr>
<th></th>
<th>Eyes ahead</th>
<th>Brows raised</th>
<th>Brows lowered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twelve female faces</td>
<td>0.50 (0.13)</td>
<td>0.71 (0.11)</td>
<td>0.38 (0.09)</td>
</tr>
<tr>
<td>Twelve male faces</td>
<td>0.44 (0.09)</td>
<td>0.67 (0.11)</td>
<td>0.29 (0.07)</td>
</tr>
</tbody>
</table>

3 Experiment 1b: Categorising brow-raised and brow-lowered faces for gender
Having shown the expected difference in brow–lid distance for male and for female faces in this set, and that these distances change in a simple linear way with instructions to raise and lower brows, we must next determine how quickly and accurately they are categorised. The experimental question is: do these changes affect classification of gender, and, if so, how?

3.1 Procedure and design
The experiment was run in the laboratory with network-linked IBM-PC (386) machines. Thirty-two (eighteen female) participants were selected from students completing their experimental undergraduate course requirement in Psychology. They were presented with instructions that if the face was a female they should press the m (or z) key;
if male press the z (or m) key as quickly and as accurately as possible. All face stimuli were unfamiliar to participants. Fifteen trials were given for (unscored) practice, with faces that were not used in the experimental set. The experimental software presented each of the twenty-four images in each of the experimental conditions randomly without replacement. The faces were displayed as grey-scale images on a light 16-inch monitor screen. Each image occupied a visual angle of about 4 deg in the centre of the screen. Each face was displayed for 750 ms. Responses were automatically logged. These included trial response times and response class (correct and incorrect) for each participant.

3.2 Results
Preliminary analyses compared male and female subjects for speed and accuracy on this task. There were no significant differences as a function of the gender of the subject, and all further analyses report data collapsed across this variable.

Separate analyses were conducted for speed and for accuracy of response. These are summarised in table 2 and the means shown graphically in figures 3 and 4. The data reported here are dependent measures summed across thirty-two subjects for each of the twenty-four stimulus faces: that is face stimuli were cast as subjects (‘cases-as-subjects’ analysis). This was in order to align findings directly with the face characteristics shown in table 1 (experiment 1a). It should be noted that analyses of subjects (means across all

<table>
<thead>
<tr>
<th></th>
<th>Eyes ahead</th>
<th>Brows raised</th>
<th>Brows lowered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female faces</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reaction time</td>
<td>524</td>
<td>514</td>
<td>559</td>
</tr>
<tr>
<td>(82)</td>
<td>(77)</td>
<td>(73)</td>
<td></td>
</tr>
<tr>
<td>errors</td>
<td>3.42</td>
<td>3.58</td>
<td>5.33</td>
</tr>
<tr>
<td>(2.96)</td>
<td>(1.72)</td>
<td>(2.70)</td>
<td></td>
</tr>
<tr>
<td>Male faces</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reaction time</td>
<td>543</td>
<td>531</td>
<td>566</td>
</tr>
<tr>
<td>(97)</td>
<td>(86)</td>
<td>(107)</td>
<td></td>
</tr>
<tr>
<td>errors</td>
<td>5.0</td>
<td>4.66</td>
<td>3.08</td>
</tr>
<tr>
<td>(5.42)</td>
<td>(5.10)</td>
<td>(1.62)</td>
<td></td>
</tr>
</tbody>
</table>

3.2.1 Reaction time

Figure 3. Experiment 1b: mean reaction time for gender decisions for changes in brow position.

Figure 4. Experiment 1b: mean errors (out of 32) for gender decisions for changes in brow position.
stimuli in each condition) gave identical patterns of significance to those reported here. The factors in the mixed-design analysis of variance (SPSS-GLM procedure) were sex of face (between subjects) and brow action (normal, raised, or lowered—within subjects).

3.2.1 Reaction times. Means of mean reaction times for each condition are plotted in figure 3. Analysis of variance showed no significant main effect of the brow-position factor ($F_{2,44} = 0.027$) and no significant main effect of gender ($F_{1,22} = 0.21$). The interaction between the factors was significant at $F_{2,44} = 3.61, p < 0.04$. Planned orthogonal contrasts showed the significant effect to be located in the interaction of gender for normal or raised vs lowered brows ($F_{1,22} = 4.31, p < 0.05$), but not for normal vs raised brows ($F = 0.03$).

3.2.2 Accuracy. Errors (out of 32) were the dependent variable for this analysis (see table 2). Means are shown in figure 4. A similar analysis for accuracy gave similar results to that for reaction time. There was no main effect of brow and no main effect of gender ($F = 0.012$). There was a significant interaction of brow and gender ($F_{2,44} = 3.85, p < 0.03$), which planned contrasts showed to be due to the interaction of gender with normal or raised vs lowered brows ($F_{1,22} = 5.33, p < 0.05$). Once more, the contrast between raised and normal brows was not significant.

The finding of a pronounced gender/brows interaction, with no main effect of brow position suggests that an image-based rather than a serial-stage model is used to judge gender in photographs of faces showing changes in brow action.

4 Discussion of experiments 1a and 1b

Although raising and lowering the brows produced similar changes in distance from the resting face (table 1), the pattern of response to the two types of action was rather different. Raising the brows hardly changed the pattern of errors and had a nonsignificant effect on speed. This effect was in the direction of speeding ‘female’ responses and slowing ‘male’ ones ($t$-test, one-tailed, $p = 0.07$). Lowering the brows, however, made a significant difference to both speed and accuracy of gender decisions as a function of the gender of the face. Male faces were classified more accurately and speedily in this pose, while classification of female ones was slowed and more error prone.

In terms of the frameworks outlined in section 1, an image-based rather than a serial-stage model fits this model well. As with changes in direction of gaze, changes in brow position can trick the gender-classification mechanism. The effects are entirely congruent with those described in our previous report (Campbell et al 1996) but they are somewhat asymmetric: while reducing brow–lid distance definitely ‘masculinises’ the face, raising the brows ‘feminises’ it only slightly (nonsignificantly). At this stage it is not entirely clear whether this asymmetry was a function of the face set used. In a parallel set of studies (Campbell et al, in preparation), ratings of masculinity for this face set have been obtained. They all appear to be judged as rather ‘feminine’ faces, and further ‘feminising’ by raising the brows may have relatively little effect.

The next experiment explores whether an image-based account can hold for other poses that lead to a change in brow–lid distance in the facial image.

5 Experiment 2a: Measuring brow–lid distance—changes in vertical orientation of the head and eyes

This study established the extent of change in brow–lid distance when the head and eyes were averted down; the individuals who posed different brow actions in experiment 1a took part. Just two poses were examined: head or eyes to camera (‘ahead’) and head or eyes down. As in the previous experiments, our first concern was to measure brow–lid distance accurately for face images under these four conditions.
We predicted that there would be a gender difference for faces looking ahead to camera: brow–lid distance should be greater for women’s than men’s faces, in line with our earlier findings. This would provide a replication of those findings on a new set of faces. Looking down (gaze averted) should increase this distance, while inclining the head may decrease it. Similarly, lowering the head, with gaze to camera, should reduce brow–lid distance, while lowering the eyes when the head is lowered should again increase brow–lid distance. This study should also indicate how the two actions may interact in terms of the brow–lid-distance measure.

Further vertical measures of the face that change with vertical movement of the head were also made. Brow–lid distance is not the only vertical measure on the face that varies systematically with gender: length of nose, of cheek, and of chin, as well as size of forehead, are all sexually dimorphic and could change with the aversion of head.

5.1 Stimuli

The faces of twelve male and twelve female participants who were sitters for experiment 1 comprised the experimental set. Images were captured in the same studio session as for experiment 1, by using identical recording and capture techniques. Each participant was asked to adopt a neutral expression for each pose. A mark was drawn on the floor for fixation, in line with the participant’s seat and the camera. When looking directly at this mark, the sitter’s head was averted down by about 45°. Instructions were given to each sitter in the following sequence: “look directly at the mark on the floor” (head down, eyes down); “keeping your head still, look at the camera” (head down, eyes up); “raising your head, look directly at the camera” (head up, eyes up); “keeping your head still, look down at the mark on the floor” (head up, eyes down). Three experimenters (MC, RC, and SW) were present throughout recording and agreed the best image to be stored for each of the four conditions. By using available computer software, these full-face images were then masked to eliminate hairstyle and clothing before being stored as grey-scale images. Figure 5 shows one set of final images for one individual.

![Figure 5. Examples of faces used in experiment 2. From left to right: (a) head up, eyes up; (b) head up, eyes down; (c) head down, eyes up; (d) head down, eyes down.](image)

The images were transferred to a Silicon Graphics R4000 Indy workstation and were analysed by using procedures developed for aligning and measuring facial landmarks (Benson and Perrett 1991) and outlined in section 2.1.
5.2 Results
The minimum vertical distance between lid and brow was at the pupil; the maximum distance usually at the outer canthi. For each facial image the average value of the six vertical measures was divided by the mean of the two (left and right) eye-width measures to give a mean normalised distance measure for each face in each condition.

Means of these means for female and for male faces are shown in figure 6, with the full numerical values shown in table 3. Analysis of variance (SPSS-GLM – repeated measures) was performed. The between-subject factor was the sex of the face, the within-subject factors were position of head (upright or averted) and eyes (to camera or averted). All three factors generated significant values. Brow – lid distance was greater in female than male faces \( (F_{1,22} = 9.74, p < 0.005) \). Faces looking down (head position) had smaller brow – lid distances than faces looking ahead \( (F_{1,22} = 195.4, p < 0.001) \) and faces with gaze averted (eyes down) had greater brow – lid distances than faces looking to camera \( (F_{1,22} = 228.8, p < 0.001) \). There were no significant interactions between any of the factors.

![Figure 6. Mean normalised brow – lid distance for faces in experiment 2a (see table 5).](image)

**Table 3.** Experiment 2a: mean normalised vertical brow – lid distances, with standard deviations in parentheses, over the twenty-four stimulus faces (twelve female, twelve male).

<table>
<thead>
<tr>
<th></th>
<th>Head up</th>
<th>Head down</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>eyes up</td>
<td>eyes down</td>
</tr>
<tr>
<td>Female faces</td>
<td>2.04</td>
<td>2.53</td>
</tr>
<tr>
<td></td>
<td>(0.47)</td>
<td>(0.51)</td>
</tr>
<tr>
<td>Male faces</td>
<td>1.70</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td>(0.27)</td>
</tr>
</tbody>
</table>

5.2.1 Analysis of other vertical measures of the face image. In addition to normalised measures of brow – lid distance, measures of the following vertical distances were made for each face: maximum length of forehead (brow length) from highest point of the eyebrow to the hairline; maximum cheek length (cheek length) from the perpendicular dropped from the midpoint of the lower eyelid to the face margin; the length of the nose (nose length) and the maximum length of the chin taken from the midpoint of the lower margin of the lower lip to the margin of the face at the pogonion (chin length). Each of these measures was again normalised for size by dividing the measure by the minimum eye distance. Mean values for each view for male and for female faces are shown in table 4.
Mixed-design analysis of variance was performed on these data in which the between-subject factor was gender and the within-subject factors were pose (upright/averted) and measure (four levels). There was a main effect of gender ($F_{1,22} = 5.51$, $p < 0.03$), of pose ($F_{1,22} = 107, p < 0.01$), and of measure ($F_{3,22} = 91.8, p < 0.001$). Female faces show (i) greater brow–lid distance, (ii) shorter brow ($p = 0.05$), and (iii) slightly shorter chin and cheek. With head down (i) is reduced (more masculine) while (iii) is reduced further (more feminine) and (ii) shows an interaction with gender. That is, the range of vertical feature distances that is gender sensitive in upright faces changes in ways that give different gender information depending on the feature. In particular, chin and brow–lid distance change in contradictory ways with respect to the gender feature that they signal.

5.3 Discussion
As predicted, brow–lid distance in this set of faces varies systematically with gender, head position, and eye position. Women’s faces show greater brow–lid distance than men’s. Brow–lid distance was reduced by downward aversion of the head and increased by downward aversion of the eyes. These effects of gender and of direction of gaze confirm our earlier findings (Campbell et al 1996) on a completely new set of faces. Additionally, downward aversion of the head to camera reduces brow–lid distance. Last, there is no interaction between these measures, nor any interaction with gender. These images, then, appear suitable for testing if there is normalisation for head pose in this task. Image-based classification would predict that accuracy and speed of classification for male faces should follow the order of brow–lid distance from smallest to greatest—while the opposite may be true for female faces. By contrast, a serial-stage-normalisation model might predict that both head-averted and eyes-averted pose should be less efficient than the canonical ‘mug-shot’ view (head and eyes to camera).

However, these measurements also showed that other vertical distances varied between male and female faces and were affected by head pose. At this stage it should be borne in mind that these could affect the interpretation of gender classification, since these measures do not signal gender in a way compatible with the changes in brow–lid distance.

6 Experiment 2b: Gender classification of faces where gaze and head pose are changed
If brow–lid distance predicts gender classification directly, following an image-based process as suggested by experiment 1b, the measurements of the twenty-four faces (experiment 2a) give rise to a simple prediction. For men’s and for women’s faces there is an orderly change in brow–lid distance as a function of head and gaze aversion, with no interaction between the variables. Thus efficiency of processing (increase in speed

| Table 4. Experiment 2a: mean face height (vertical) measures for faces looking ahead and down, all normalised for interpupil distance (with standard deviations in parentheses). |
|-----------------|-----------------|-----------------|-----------------|
|                | Female          | Male            |
|                | ahead | down | ahead | down |
| Cheek length   | 1.47  | 1.17 | 1.51  | 1.23  |
| (0.10)         | (0.11)| (0.14)| (0.09)|       |
| Brow (forehead) length | 0.85  | 0.78 | 1.03  | 1.05  |
| (0.41)         | (0.37)| (0.15)| (0.20)|       |
| Nose length    | 0.83  | 0.83 | 0.83  | 0.83  |
| (0.07)         | (0.06)| (0.07)| (0.07)|       |
| Chin length    | 0.67  | 0.48 | 0.73  | 0.52  |
| (0.06)         | (0.06)| (0.09)| (0.06)|       |

Face pose and gender perception 499
and/or accuracy of classification) should be directly related to brow–lid distance for female faces, inversely related to brow–lid distance for male faces. Our first prediction, therefore, is that there will be an interaction of gender of face with pose—and that both head and eye direction will change processing efficiency systematically. At a finer level, if brow–lid distance is the critical clue to gender, an orderly gradient of classification efficiency can be predicted: for women's faces the fastest reaction times are predicted to be in the following order from best to worst: (1) head up–eyes down (largest brow–lid distance), (2) head up–eyes up, (3) head down–eyes down, (4) head down–eyes up. Men's faces should follow the opposite pattern. Any other pattern would suggest a more complex relationship between gender classification and brow–lid distance.

6.1 Method

6.1.1 Subjects. Twenty-one volunteer participants (eleven female) who had not performed experiment 1b and who were not familiar with the face set performed the task as part of their experimental undergraduate course requirement.

6.1.2 Procedure and design. The experiment was run in a very similar fashion to experiment 1b, with network-linked IBM-PC (386) machines in the laboratory. Participants were presented with instructions that if the face was a female one they should press the m (or z) key of the computer keyboard, and if it was a male face press the z (or m) key as quickly and as accurately as possible. Fifteen trials were given for (unscored) practice, with faces that were not used in the experimental set. The experimental software presented each of the 4 × 24 experimental images randomly without replacement in one run to each participant. The faces were displayed as grey-scale images on a light 16-inch monitor screen. Each image occupied a visual angle of about 4 deg in the centre of the screen and was displayed for 750 ms. Responses were automatically logged. These included individual trial response times and response class (correct and incorrect for actual gender). Total experimental time for each participant was less than 15 min and they were debriefed after the study.

6.2 Results

Once more, results are reported for face images as subjects. That is mean reaction times and accuracy scores were taken for each face across all twenty-one participants. Mean scores for the twelve experimental trials in each condition, for male and for female faces, were obtained. Means and standard deviations are shown in table 5.

Table 5. Speed (reaction time, in ms) and accuracy (errors out of 21) of gender decisions: mean scores, with standard deviations in parentheses, for twenty-four faces (twelve female, twelve male).

<table>
<thead>
<tr>
<th>Brow–lid distance rank order</th>
<th>Head up</th>
<th>Head down</th>
</tr>
</thead>
<tbody>
<tr>
<td>female faces reaction time</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>errors</td>
<td>741.05</td>
<td>723.75</td>
</tr>
<tr>
<td>(62.46)</td>
<td>(58.72)</td>
<td>(97.01)</td>
</tr>
<tr>
<td>male faces reaction time</td>
<td>927.25</td>
<td>823.90</td>
</tr>
<tr>
<td>errors</td>
<td>2.50</td>
<td>2.50</td>
</tr>
<tr>
<td>(3.23)</td>
<td>(2.74)</td>
<td>(2.34)</td>
</tr>
</tbody>
</table>
6.2.1 Reaction times. Repeated-measures analysis of variance (SPSS-GLM) was performed on these scores. The factors were (within subjects) head position (up or down) and eye position (up or down), and (between subjects) gender of face. The main effects of head position ($F_{1,22} = 6.07$, $p = 0.02$) and of eye position ($F_{1,22} = 10.88$, $p < 0.01$) were significant. The interaction of eyes and gender ($F_{1,22} = 6.24$, $p = 0.02$) was also significant. There was also a significant main effect of gender of face ($F_{1,22} = 5.84$, $p < 0.03$). No further effects were significant. In this set of faces, female faces were faster to classify, as were head-up and eyes-up faces. The eyes x gender interaction suggests that, whatever the position of the head, averting the eyes downwards slows decisions on male faces.

6.2.2 Errors. Similar analysis of errors (see table 5) showed that only the interaction of head and gender was significant ($F_{1,22} = 6.02$, $p < 0.03$). Male faces were classified less accurately when the head was up. This odd result may simply reflect a few anomalous images within the series. Overall, error rates were not significantly above zero ($F_{1,22} = 0.07$) and in the two head-up conditions which generated most errors these were confined to just three images out of twenty-four. We offer no further comment on these data.

6.3 Discussion
In contrast to our earlier study (Campbell et al 1996), but confirming experiment 1 in this series, this set of faces generated a ‘female’ bias: female faces were faster and more accurately classified overall. As suggested above, this probably reflects the relative youth of the faces and, to a lesser extent, the presence of face jewellery. Also, in distinction to the study published earlier, these faces generated a gender-by-gaze interaction: looking down made female faces easier to judge, male ones harder to judge. Despite being moderated by gender in this experiment, our earlier findings concerning direction of gaze were confirmed here: looking down ‘demasculinises’ the face.

The reaction-time data from this study were systematic: but which hypothesis do they support? The prediction from the image-based hypothesis was that efficiency of categorisation should follow the pattern indicated by the measurements of experiment 2a; that is that there should be a marked interaction of gender with pose, and the gradient of efficiency should track the measured brow–lid distances. From best to worst for male faces that would be: head down–eyes up, head down–eyes down, head up–eyes up, and head down–eyes down (vice versa for female faces). The normalisation hypothesis, by contrast, suggests that discrete preprocessing may be required to control first for head and then for gaze prior to gender classification: that is, the gradient of efficiency should be independent of gender and should be in the order head up–eyes up, head up–eyes down, head down–eyes up, and head down–eyes down.

In fact, for these faces, the outcome in relation to these hypotheses was mixed. While the image-based hypothesis was supported by the interaction of gender with direction of gaze, there was no indication of a three-way interaction between gender, head, and eye pose which would have been strong confirmation for an image-based analysis. In favour of the normalisation hypothesis, the main effect of head was independent of gender and also of direction of gaze, suggesting separate processing stages are engaged for the normalisation of head pose and of eye gaze. A best fit therefore appears to be that face images are normalised for head pose, but not for direction of gaze when speeded gender decisions are made.

How else might these effects be explained? There are other vertical distances in faces than brow–lid distance that are gender related and change with head position (see table 4). Female faces in this set conform with other structural analyses (eg Ferrario et al 1993) in having shorter brow-to-hairline (forehead) distance and marginally shorter lip-to-chin and cheekbone-to-jawline (cheek) distances compared with the male faces (all measures were normalised for interpupil distance to control for different
face-image sizes). With head down, brow–lid distance is reduced (more masculine) while chin and cheek length are reduced further (more feminine) and brow–hairline length shows an interaction with gender. The vertical feature distances that are gender sensitive in upright faces change when the head is inclined, but in ways that may give different gender information depending on the particular measure. While lowering the head might make a female face appear more masculine in terms of the brow–lid-distance cue, it could feminise the face even more in terms of the chin-length and/or cheek-length cue. In this way, apparent normalisation for head pose may yet be accounted for by an image-based hypothesis when vertical distances other than brow–lid distance are taken into account.

While the present study has explored the effects of change of experimental face-pose changes on gender classification, a recent study has discovered possible effects of pose in gender classification of studio-style photographic portraits. This was a PCA analysis performed by O'Toole et al (1998). In that study, a single large set of full-face, gaze-to-camera, head-and-neck portraits of young American adults comprised the data set for PCA compression. The face images were also classified for gender, attractiveness, and recognisability by a team of human observers, and a statistical model was developed to indicate the best weighting of the principal components—the eigenvectors—for each of these (human) categorisation functions. Of the twelve eigenvectors that fully described the set, the first three were highly predictive of gender classification. While the first and second eigenvectors seemed to capture the generalised ('average') face values and hairlength/style, the third eigenvector reflected a rather different aspect of the face; one which the authors were surprised to find in this set of images where gaze and aspect of the face seemed to be fully controlled:

"our interpretation of ... this model measure (third eigenvector) is that it conveys information about a facial mannerism. Thus it seems possible that a face can be made to appear (at any instant) more feminine or more masculine via some very simple facial mannerisms. For example, looking straight ahead and seeking direct eye contact may lend any face a more masculine appearance, whereas averting the eyes and gazing downward may lend a face a more feminine appearance" (O'Toole et al 1998, page 157).

In other words, an image-based (PCA) approach to the categorisation of gender from faces independently predicts the effects reported here and in Campbell et al (1996). O'Toole et al's approach made no use of specific structural cues to gender such as the brow–lid-distance cue: its salience emerged naturally from the statistical structure of the image set and also by its classification by human observers.

7 General discussion
Two experiments showed that effects of pose on gender classification of face images are robust and systematic. Image-based process seems to apply to changes in brow position and direction of gaze and, while changes in vertical inclination of the head can be accounted for by normalisation prior to classification, direct image-based processing may have a role here too. Whether such effects are limited to the classification of simple photographic face images or can apply to more naturalistic displays of faces too (eg video clips) is a question for further research. These studies have explored possible perceptual mechanisms for judging gender from face images quickly—but what is the functional significance of these findings?

Changes in pose of the head may affect gender perception in two ways: they alter the signalling properties of the source face, that is gender-dimorphic behaviour may be perceived, but as we have seen, structural aspects of the face, which carry gender-salient information, are affected by pose too. In section 1, it was pointed out that image-based approaches using a learning algorithm (typically a neural-net algorithm) to hone categorisation skill are only as good as the set on which they have been trained.
For humans looking at male and female faces, learning contingencies may be affected by a range of factors. For instance, male faces may be typically seen with lowered brows, female ones with raised or normal brows. There may be gender-related learning that is linked to the salience of the judgment—the costs of misclassifying a low-browed face as female may be less than those of classifying a high-browed face as male.

As infants, we may be exposed to more female than male adult faces (mother, carer, teacher) and become more adept at discriminating expression in such faces. This could account in part for the asymmetry in the effects of raising the brows in experiment 1b, where female faces were less affected than male faces—although this effect may also have reflected a rather more ‘feminine’ cast to this set of faces than in our previous study. Such speculations are essentially neutral concerning the cause of gender-typical behaviours by men and women. They just assume that fairly simple perceptual mechanisms can tune in to them. A more radical suggestion is that such behaviours are not only designed to be picked up by simple perceptual processes, but that they interact with facial displays in a manner that betrays their evolutionary history. The cosmetic enhancement of brow–lid distance could be evidence for this. This appears to take a universal form: brow thickening, thinning, and scarifying seems to be done in a manner congruent with enhancing the ‘natural’ gender difference, and not obscuring it—at all times, and in all cultures. This suggests that, as far as the brow region is concerned, all humankind uses a similar means to enhance the gendered appearance of the face.

So could face acts that affect the disposition of these features be related to gender signs? According to Darwin’s speculations concerning the origin of the emotions (Darwin 1872), brow lowering may have a basis in two functions—the protection of the eyes (this may be needed when anger is displayed—intention to fight) and as a concomitant of vergence of the eyes to inspect close material (the ‘intently thoughtful’ face). The narrowing of the lips occurs in similar intentional settings. Neither of these sets of face acts can be construed as gender specific in themselves, but they might tend to be associated with male behaviour. They also result in modifying the classification of gender of the face of the actor. Is it accidental that lowered brows and narrowed lips characterise the structure of male faces as well as a ‘typically masculine’ behaviour pattern? Human behaviour as well as human appearance is sexually dimorphic, tuned in a variety of ways by culture. While many of the signs of gender cannot be changed by moving the face muscles (eg the breadth of brow or chin) one way a man can look more manly is by adopting a facial display that enhances masculinity. Women may appear more womanly by adopting the opposite pattern (Darwin’s ‘principle of antithesis’ in action)—including the vertical aversion of gaze.

We speculate that displays of gender and of intention have developed in an interactive fashion, to take advantage of a perceptual mechanism that can be tricked by one (intention) in processing the other (gender), and that this could be a basis for the present form of several facial displays. These could include ‘aggression’ and ‘resolution’ for the lowered-brow male face; ‘primness’, ‘girlishness’ for the raised-brow and ‘demureness’ for the lowered-eye female face—all of which may be construed as typical of one or other gender. Further studies requiring judges to rate displays like those in the present experiment for a variety of intentions may help elucidate this further.

Acknowledgements. We thank Rob Davis, Goldsmiths College, University of London for developing the software to run these experiments. Philip Benson acknowledges the support of the MRC and the Oxford Centre in Brain and Behaviour, and the McDonnell–Pew Centre for Cognitive Neuroscience. We thank Paul Ekman for discussion, and Vicki Bruce and several anonymous reviewers for their constructive comments on an earlier version of these studies.
References
Fridlund A J, 1991 “Evolution and facial action in reflex, social motive and paralanguage” *Biological Psychology* **32** 3 – 100
Galton F, 1907 *Inquiries into Human Faculty and its Development* 2nd edition (London: Dent)