
Does face recognition rely on encoding of 3-D surface? Examining the role of shape-from-shading and shape- from-stereo

Chang Hong Liu, Charles A Collin, Avi Chaudhuri

Department of Psychology, McGill University, 1205 Dr Penfield Avenue, Montréal, Québec H3A 1B1, Canada; e-mail: chang@psych.mcgill.ca; web: www.psych.mcgill.ca/labs/cvl

Received 20 October 1999, in revised form 29 February 2000

Abstract. It is now well known that processing of shading information in face recognition is susceptible to bottom lighting and contrast reversal, an effect that may be due to a disruption of 3-D shape processing. The question then is whether the disruption can be rectified by other sources of 3-D information, such as shape-from-stereo. We examined this issue by comparing identification performance either with or without stereo information using top-lit and bottom-lit face stimuli in both photographic positive and negative conditions. The results show that none of the shading effects was reduced by the presence of stereo information. This finding supports the notion that shape-from-shading overrides shape-from-stereo in face perception. Although shape-from-stereo did produce some signs of facilitation for face identification, this effect was negligible. Together, our results support the view that 3-D shape processing plays only a minor role in face recognition. Our data are best accounted for by a weighted function of 2-D processing of shading pattern and 3-D processing of shapes, with a much greater weight assigned to 2-D pattern processing.

1 Introduction

1.1 *Faces can be identified from shading information alone*

Faces are 3-D objects. As discovered by ancient sculptors, facial resemblance can be achieved by imitating the 3-D shape alone without the need to include texture and colour. Apart from aesthetic reasons, statues today are often uncoloured for this reason. Indeed, most statues are presented in their natural-material colours. Similarly, studies using sculpture-like 3-D laser-scanned faces that are devoid of facial colour and texture information have confirmed that, even though more difficult, faces can be identified from their shading information alone (Bruce et al 1991; Bruce and Langton 1994; Hill and Bruce 1996; Troje and Bühlhoff 1996, 1998; Tarr et al 1998; Liu et al 1999). This fact is useful in understanding the nature of face processing, because it shows that shading information alone may play an equally important role in identifying real faces. The literature on face recognition in the past thirty years has provided rich evidence for such a hypothesis.

The role of shading information is well demonstrated by a number of phenomena. First, studies have shown that a line drawing of a face containing no shading information is more difficult to identify than a photograph of the same person (Davies et al 1978; Bruce et al 1991; Leder 1996). Second, altering lighting direction between learning and test can impair recognition performance (Johnston et al 1992; Hill and Bruce 1996; Enns and Shore 1997; Braje et al 1998; Troje and Bühlhoff 1998; Liu et al 1999). Third, reversing the luminance contrast of face stimuli (eg by using photographic negatives) results in poor performance (Galper 1970; Galper and Hochberg 1971; Phillips 1972; Hayes et al 1986; Hayes 1988; Johnston et al 1992; Bruce and Langton 1994; Kemp et al 1996; Gauthier and Tarr 1997; Liu and Chaudhuri 1997, 1998). And, finally, facial information is best described by a band of middle spatial frequencies, a large portion of which function to provide shading information (Näsänen 1999).

Frequency overlap between face images is a crucial predictor of recognition performance (Liu et al, in press). What is remarkable about these findings is that, under many circumstances, transformation of face images (such as reversing luminance contrast or altering lighting direction) does not alter facial features or their configurations. And yet, face recognition becomes variably affected. This suggests that face recognition is indeed sensitive to shading information.

1.2 *Shape-from-shading and shape-from-stereo*

The important role of shading information may suggest that the brain derives 3-D information from this cue. According to this view, shading information is used by the recognition system to reconstruct a 3-D representation from 2-D input. If this is so, the impairments discussed earlier may be attributed to a disruption of 3-D shape processing from shading. Obviously, shading is only one of many cues from which 3-D information can be extracted. Studies based on shading information alone can hardly allow us to fully estimate face perception in the real world. In natural images, the depth cues about the 3-D surface of a face (such as stereopsis, motion, and texture gradients) are rich and simultaneously available. How do these different depth modules interact and contribute to face processing?

In attempts to address this question, there has been increasing attention paid to the role of motion (eg Bassili 1978; Nelson and Horowitz 1983; Stucki et al 1987; Berry 1990, 1991; Soken and Pick 1992; Hill et al 1997; Knight and Johnston 1997; Pike et al 1997; Christie and Bruce 1998; Ramachandran et al 1998; Lander et al 1999). However, few studies have addressed the role of other depth cues and the way these cues are integrated in face perception. In this study, we decided to investigate how shape-from-stereo interacts with the process of shading in face perception. We chose to examine shape-from-stereo because this depth cue is the most salient in constructing 3-D shape (Bülthoff 1991). Therefore, this depth module should be the most useful if face processing does indeed involve computations of 3-D surface information.

The role of stereopsis in face recognition has not been studied extensively. The scope of the studies has so far been limited to perception of hollow faces rather than normal ones. Hill and Bruce (1993) found that the hollow-face illusion disappeared more quickly under binocular viewing than monocular viewing conditions as the distance between their observers and the hollow mask was shortened. Van den Eenden and Spekrijse (1989) found that, once texture disparity is neutralised, a face can be perceived as concave when viewed pseudoscopically. These studies demonstrate that, under certain circumstances, information provided by stereopsis can override the convexity assumption.

In a few studies, the interaction of shape-from-shading and shape-from-stereo has been investigated with simple 3-D forms. In these studies, shape-from-stereo has often been shown to be useful when combined with other depth modules (Bülthoff and Mallot 1988; Bülthoff 1991; Parker et al 1995; Uttal et al 1996). Although perception of 3-D structures from shading alone can be highly ambiguous and inaccurate (eg Erens et al 1993; Johnston and Passmore 1994a, 1994b; Mamassian and Kersten 1996), Bülthoff and Mallot (1988) have shown that subjects are able to judge the 3-D shape of an ellipsoid more accurately when shading is combined with stereo than from any of the other depth cues alone. Edelman and Bülthoff (1992) have shown that stereo information also moderately improves recognition of simple wire-like or amoeba-like objects.

However, not all studies show benefits of combining stereo and shading cues. An instance can be found in the study of the asymmetric visual search effect for top-lit or bottom-lit cubes by Sun and Perona (1997). It is known that searching for a top-lit cube or sphere among bottom-lit ones is more difficult than searching for a bottom-lit cube or sphere among top-lit ones. Because a shaded cube or sphere can be seen either as convex or concave, the asymmetry effect is commonly attributed to the preference of

early vision for convex and top-lit interpretations. Sun and Perona (1997) examined whether eliminating the ambiguity of 3-D shape from shading by stereo cues could reduce the asymmetry effect. They found that the asymmetry effect in visual search for top-lit or bottom-lit cubes not only persisted but can be even stronger under stereo conditions. Thus, their data show that stereo information does not override the top-lit and convexity assumptions preferred by early vision. However, equally notable was a result that, when stereo information was consistent with a convex top-lit interpretation, there was a clear benefit from such information. Whether stereo information can facilitate or interfere with the task depends on whether the preference of early vision is satisfied or violated.

The question remains whether these findings apply to complex 3-D objects such as faces as well. If accurate estimation of 3-D surface is indeed important in face perception, then adding stereo information to shading should improve face perception, provided that the stereo input is consistent with the assumptions preferred by early vision. On the other hand, because face perception may involve processes other than 3-D shape estimation, the benefits of stereo and shading combination found in judging surface orientation of simple objects may not apply to face-identification tasks. In the visual-search literature, there is already evidence that the search asymmetry effect does not apply to top-lit and bottom-lit faces (Troje and Symons 1998). A simple reason for this difference may be that, although simple discs defined by shading can be interpreted as either convex or concave, faces are more likely to be perceived only as convex.

To study this issue, we examined two effects that are commonly thought to be due to limitations in processing 3-D shape information from shading. One of them is the effect of lighting direction and the other, contrast reversal. Our main question is whether the deficits that are so evident in these conditions can be reduced when shading is combined with stereo information.

1.3 *The effect of lighting direction*

Consistent with the top-lit preference found in simple forms, top-lit faces are easier to identify than bottom-lit faces (Johnston et al 1992; Hill and Bruce 1996; Enns and Shore 1997; Braje et al 1998; Liu et al 1999). It is thought that the proficiency in processing 3-D shape information in top-lit objects may be the result of continued exposure to a predominant top-lit environment (Ramachandran 1988).

Stereo processing, on the other hand, relies only on disparity information. If the stereo module is not affected by luminance difference or lighting directions, identification of bottom-lit faces with stereo information should be better than with shading information alone. However, it is not clear how 3-D information from the two modules is combined. One possibility is that the two modules work independently on 3-D descriptions and each sends an output for further integration. Another possibility is that the two modules are interactive, so that the assumptions that affect one module will also affect the other. In the first case, because the preference for top-lit stimuli used by the shading module does not affect 3-D shape processing in the stereo module, identification of bottom-lit faces may be easier with than without stereo information. In the second, the top-lit preference used by the shading module could interfere with 3-D processing in the stereo module, hence adding stereo information may add little benefit in identification of bottom-lit faces. To address this issue directly, we compared identification performance for top-lit and bottom-lit faces with or without the presence of stereo information in our first experiment.

1.4 *The effect of contrast reversal*

The difficulty in recognising faces in photographic negatives is another effect that is attributable to a disruption of shape-from-shading processes. Liu and Chaudhuri (1997) have distinguished two kinds of photographic negative effects in which the sign of contrast between learning and test can be either the same (congruent) or reversed

(incongruent). The distinction is important because different mechanisms may underlie these effects. For the congruent contrast, in which faces are learned and tested in the same contrast polarity, the difference between recognition performance for positive and negative faces may be accounted for in part by the bottom-lit appearance when a top-lit face is presented in negative (Johnston et al 1992; Liu et al 1999). For the incongruent contrast condition, in which faces to be matched are displayed in reversed contrast, the identification deficit may be attributed to a disruption of shape-from-shading information in the contrast-reversed images in addition to the difficulty caused by apparent bottom lighting. Not surprisingly, the incongruent condition is more detrimental to recognition than the congruent condition (Liu and Chaudhuri 1997).

In this study, we examined whether the mismatch of the 3-D descriptions derived from incongruent conditions can be corrected or compensated by stereo information. As discussed in the last section, this should also depend on how 3-D shape processing in the stereo and shading modules is integrated. If stereo processing is independent of lighting pattern, the similarity between the 3-D shapes of a pair of contrast-reversed images may be corrected by stereo information. However, if stereo and shading modules are highly interactive, the processing of 3-D information in stereo could also be interfered with by contrast reversal. We tested the two alternatives in our second experiment.

2 Experiment 1

The purpose of this experiment was to determine whether adding shape-from-stereo cues can reduce the lighting-direction effect on face recognition found in earlier studies. To separate the contributions of shading from pigmentation, we employed the same untextured face stimuli that were used in our earlier study on lighting direction (Liu et al 1999). An example of these faces is presented in figure 1.

Our earlier study found that face recognition in both photographic positive and negative was affected by lighting directions. Top-lit faces were identified better than bottom-lit ones when they were presented in positive. The opposite pattern was found when faces were presented in negative. We reasoned that this occurred because of reversals in the apparent lighting of negatives (ie bottom-lit faces appear top-lit in negative and vice versa). Our results showed that faces were easier to identify in top lighting, and that the photographic negative effect was partially due to apparent bottom lighting. The advantage for apparent top-lit faces is consistent with the preference for top-lit interpretation demonstrated in the shape-from-shading literature.

In this experiment, we used a similar design, but instead presented half of the faces in stereo. Our goal was to find whether the effect of lighting direction can be reduced by the presence of shape-from-stereo information.

2.1 Method

2.1.1 Subjects. Sixty-seven undergraduate students from McGill University participated. All subjects were required to pass a screening test for normal stereo vision. We used Titmus Stereo Tests (Titmus Optical Co. Inc., ca1960) for this purpose. The passing criterion was a stereo acuity of 40 s of arc. All but two subjects passed the test. The remaining sixty-five subjects were randomly assigned to two groups. Thirty-three subjects were shown positive face images. Thirty-two were shown negative face images. The ages ranged from 18 to 33 years (median = 21 years).

2.1.2 Materials. 3-D laser-scanned faces, originally developed at University College London, were created by a laser beam which recorded the 3-D structures of faces. Over 20 000 x , y , z coordinates were recorded for each face. This provided a very detailed description of the facial surface. The coordinates were connected by polygons, each having four vertices. A full description of these 3-D models can be found in a number of published studies (Bruce et al 1991; Bruce and Langton 1994; Hill and Bruce 1996).

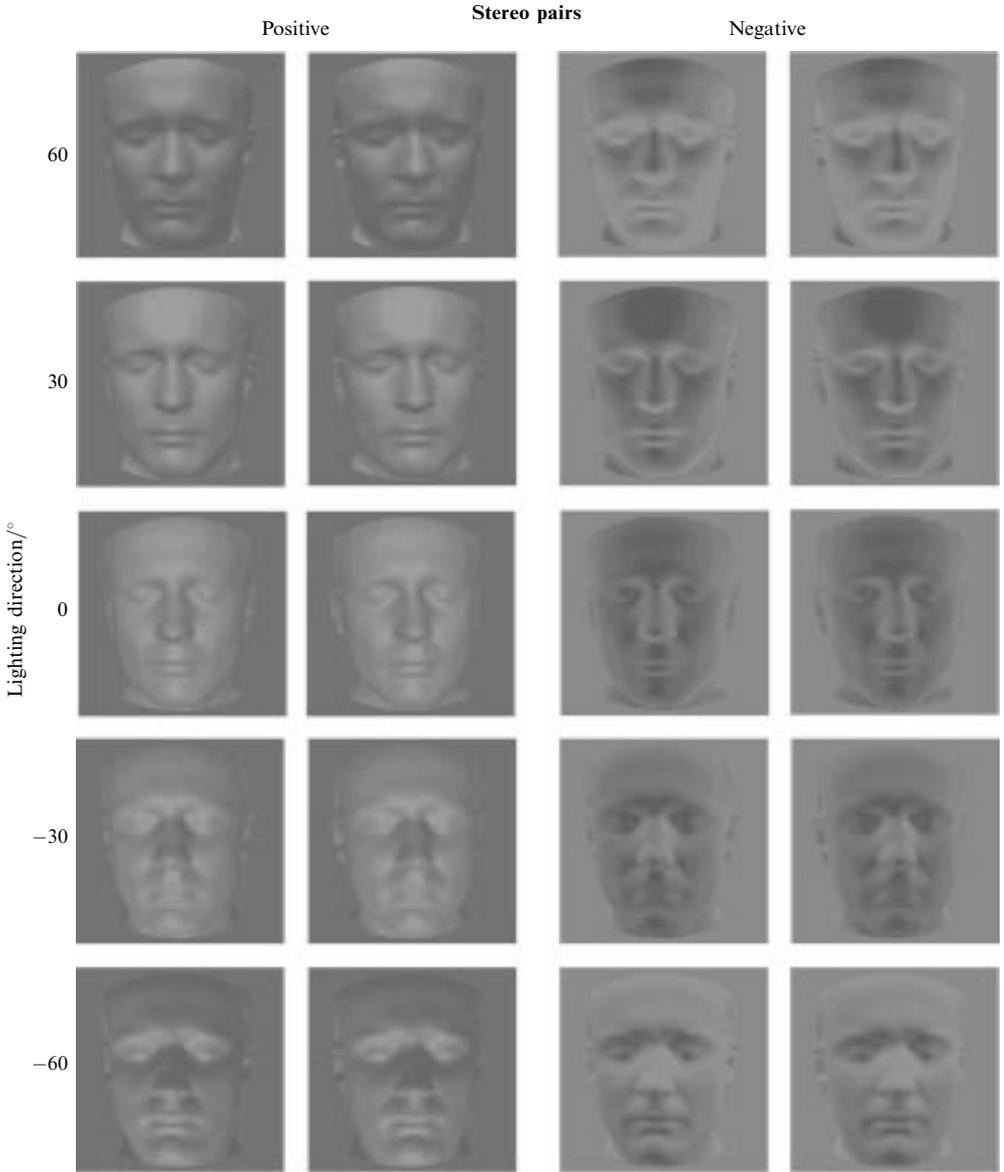


Figure 1. An example face manipulated for different conditions used in experiments 1 and 2. These can be free-fused (crossed fusion) to get a sense of the stereo condition. The examples for shading-alone and three-quarter view conditions are not shown here. Images to be matched had the same contrast polarity (experiment 1) or opposite contrast polarity (experiment 2).

Two sets of face images were created from the models, one for shading-alone and the other for shading-plus-stereo condition. We used nine male and nine female models. Two of these were designated for use in the practice session. The set for shading-alone condition was adapted from our previous study (Liu et al 1999). We rendered the models using Geomview 2.0 (www.geom.umn.edu/projects/visualization/). A Gouraud shading function was used to apply gradually interpolated shading to each facet and thus smooth the appearance of the polygonal surface. Two views, one full-face (0°) and one three-quarter (45°) of each 3-D face model were captured. We used an elliptical window to clip off the parts of face above the hair line and below the chin.

The virtual camera was one metre away from the face. The projection from the 3-D models to the camera assumed perspective. The standard OpenGL illumination model was used with a single light source at infinity and a small amount of ambient light. The relative intensities of between the light source and the ambient light were 0.8 and 0.2, respectively. The faces were rendered with a diffuse reflection factor of 1.0, a specular reflection factor of 0.3, an ambient reflection factor of 0.3, and a specular reflection exponent of 13.7. For both full-face and three-quarter views of each model, five images with different lighting directions were captured. The lighting directions were $+60^\circ$, $+30^\circ$, 0° , -30° , and -60° relative to the horizontal meridian of the face, with positive values indicating angles above the meridian and negative values below the meridian. All images were then reduced from 16-bit RGB format to 8-bit gray scale with Graphic Converter 2.2 for Macintosh (www.lemkesoft.de).

The images for stereo conditions were taken by two simulated cameras. The camera positions were displaced horizontally by 6 cm, which approximated average interocular distance. All other aspects of the procedure for generating the stereo images were the same as that used for the non-stereo image set.

A potential problem with this rendering technique is that different lighting directions resulted in different overall luminance. For example, faces with a lighting direction at -60° were dimmer than at -30° and 0° in the positive images. This systematic overall luminance difference could become a confounding variable when the effect of lighting direction was evaluated. We dealt with this problem by equalising the mean luminance for the images of different lighting directions. This was done by calculating the mean pixel value for each of the positive images to derive a grand mean luminance. The grand mean was then used to scale all images to this value.

Both the stereo and non-stereo images were also created in negative. One way to generate a negative image is to simply subtract the pixel values of a positive image from 255—the maximum pixel value of an 8-bit image. However, this method alters the contrast amplitude in the negative owing to the nonlinear relationship between pixel values and the display luminance. The result is that the negative images tend to have less contrast than the positive images. To produce negative images that preserve the same contrast amplitude as in the positive images, the gamma function of the monitor was linearised. The screen luminance was measured by a photometer (Hagner Universal Photometer S2) at all 256 gray levels. The result of this calibration was then used to convert pixel values of the scaled images to luminance values. The negative images were then created by applying a 180° phase shift in the Fourier domain of the luminance values.

A total of 1080 images were used in this study (18 faces \times 2 views \times 5 lighting directions \times 3 non-stereo singles and stereo pairs \times 2 contrast polarities). The faces were embedded in a neutral gray background that filled the screen. All face images were scaled to 256×256 pixels ($95 \text{ mm} \times 95 \text{ mm}$). The images were displayed on a 21-inch Apple monitor and viewed through a stereoscope. The screen resolution was set to 832×624 pixels with millions of colours. At each presentation, two images were shown on the screen. The distance between the two images was 20 pixels. For the non-stereo condition, the two images were identical. They were the same image presented at both locations. An example face manipulated for different lighting directions in stereo pairs is shown in figure 1.

The size of the stereoscope was $44 \text{ cm} \times 12 \text{ cm} \times 23 \text{ cm}$. It was a simple reflector-type stereoscope with eight front-surfaced mirrors. The back side of the stereoscope was attached to the screen surface. The two mirrors that were directly projected to the eyes were adjustable and used to bring the stereo pairs into fusion. The images were viewed from a distance that was the same as the stereoscope window (43.3 cm), taking into account the distance between the mirrors. Each face image subtended approximately $12.4 \text{ deg} \times 12.4 \text{ deg}$ of visual angle.

2.1.3 Procedure and design. We employed a $2 \times 2 \times 5$ mixed design. The three factors were two levels of depth cues (shading-alone, shading-plus-stereo), two levels of contrast polarity (positive or negative), and five levels of lighting direction ($+60^\circ$, $+30^\circ$, 0° , -30° , and -60°). Contrast polarity was a between-subject factor; the rest were within-subject factors.

Subjects were tested individually on a Power Macintosh G3 computer. Instructions were given on the monitor. To ensure appropriate fusion, subjects were required to perform a simple stereo test using random-dot stereograms. The stereograms were 256×256 pixel random-dot patterns. The intensity level of pixels was binary (0 or 255). A 50×50 pixel square in the centre of each stereo half was moved to the left or right to create a disparate region. The initial size of disparity for the test was 10 pixels. Subjects were asked to judge whether the square was in front or behind the background. Each correct response was followed by a decrease of the disparity by half (rounded to an integer). The incorrect response was followed by an increase of the disparity by half. The test stopped after three correct responses to the smallest disparity (one pixel). All subjects were able to complete this task once stereo fusion was achieved.

After the short stereo test, subjects were given 16 trials of practice for the sequential face-matching task. They were then given feedback about their performance (accuracy and mean reaction time) for the practice. The actual experimental trials took place immediately after the practice session.

At each trial, a pair of faces was presented sequentially. A 256×256 pixel random-dot pattern was presented between the two face images. Each pixel of the pattern was assigned a random grey level from 0 to 255, by using a flat probability distribution. The first face was presented for 1.5 s, followed by the noise pattern, which was presented for 1 s. The second face was then shown and remained on the screen until the subject responded. The subject was instructed to judge whether the first and the second face were of the same person and to respond as quickly and as accurately as possible. The two faces were always presented in different views. On one half of the trials, a full-face view was presented first followed by the three-quarter view. This order was reversed for the other half.

Half of the male and female faces were randomly selected and presented in the stereo condition. The rest were presented in the non-stereo condition. At shading-alone trials, the pair of images presented to the two eyes through the stereoscope were identical. Half of the faces were randomly assigned as targets, the rest were assigned as distractors. Both faces in a pair were always of the same gender, whether or not they were matched. The presentation order for the eight target faces were randomised only once for each subject. The subsequent repetitions of the target faces in different conditions followed the same order. This was intended to equalise the length of delay between the repetitions of each face in different conditions, and hence to reduce the variability due to uncertainty in recency effects. The other factors (views, depth cues, and lighting directions) were fully randomised. These conditions form a total of 80 trials ($2 \text{ faces} \times 2 \text{ stereo} \times 2 \text{ genders} \times 2 \text{ view orders} \times 5 \text{ lighting directions}$).

2.2 Results and discussion

2.2.1 Accuracy. The percentage accuracy data are shown in figure 2. A three-way analysis of variance (ANOVA) showed significant main effect of contrast polarity ($F_{1,63} = 3.96$, $p < 0.05$), with positive images being identified 3.1% more accurately than negative images. There was also a significant interaction between contrast polarity and lighting direction ($F_{4,252} = 5.36$, $p < 0.0001$). The rest of the main effects (depth cues and lighting directions) and interactions (contrast polarity versus depth cue versus lighting direction, contrast polarity versus depth cue, and depth cue versus lighting direction) were not significant.

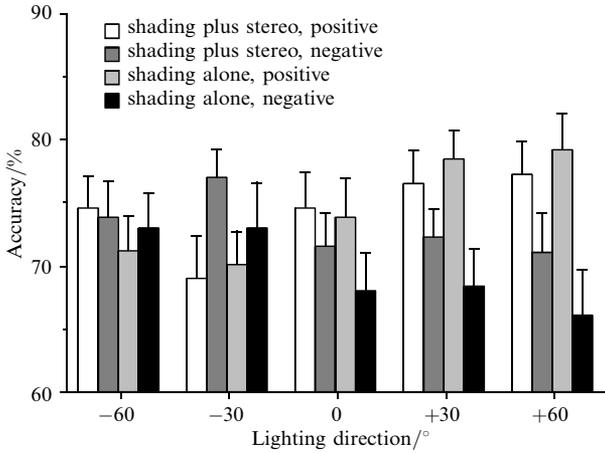


Figure 2. Percentage accuracy as a function of polarities, depth cues, and lighting directions in experiment 1. The error bars represent standard error.

Tukey Honest Significant Difference (HSD) a posteriori tests ($\alpha = 0.05$) revealed that when the lighting direction was at $+30^\circ$ and $+60^\circ$, faces in positive were identified more accurately than faces in negative at $+60^\circ$. Positive faces at $+60^\circ$ were also recognised better than negative faces at $+30^\circ$ and 0° . There was no other difference between results from positive and negative conditions. Within the positive conditions, faces in $+60^\circ$ were recognised better than faces in -30° , whereas the results for the other lighting directions were not different from one another. Within the negative conditions, results for all the lighting directions were not different from each other.

2.2.2 Reaction time. The median for each condition was computed for each subject. Only the reaction time for the correct responses were included. The data from two subjects, one from each group, were dropped because of the missing cells created by their incorrect responses in those conditions. The means and standard errors are shown in figure 3.

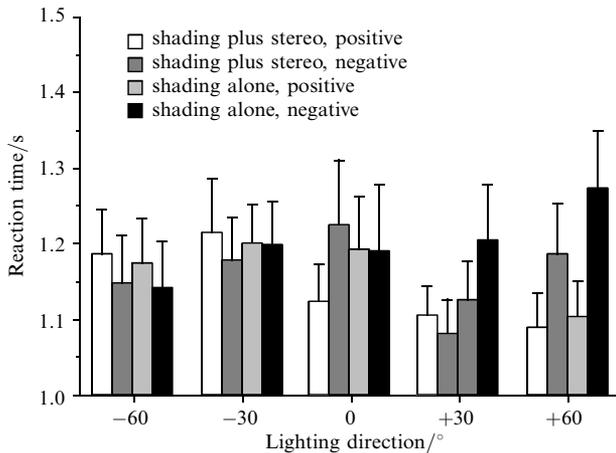


Figure 3. Reaction time as a function of polarities, depth cues, and lighting directions in experiment 1. The error bars represent standard error.

To normalise the reaction times, a logarithmic transformation was applied to the data set before it was submitted to ANOVAs. The only significant result was the interaction between contrast polarity and lighting direction ($F_{4,244} = 2.57, p < 0.04$). Tukey HSD a posteriori analysis revealed that the interaction was due to a marginally ($p = 0.06$) faster response (134 ms) to positive than to negative images when faces were lit at $+60^\circ$.

Neither accuracy nor reaction time data show any interaction between depth cue and lighting direction. However, both reveal a significant interaction between contrast polarity and lighting direction. That is, top-lit faces are easier to recognise in positive than in negative. Thus, although shape-from-stereo and shape-from-shading were combined in this experiment, the basic findings were consistent with what we have found previously with shading information alone (Liu et al 1999). A small discrepancy was found for bottom-lit faces in which the accuracy for positive and negative faces was similar, rather than confirming the advantage of negatives in this situation. This may be due to an overall drop of performance in the negative-face condition. The shape of the accuracy data for the negative face shown in figure 2 is very similar to our previous finding. The reason for the overall drop in performance is not clear. One possibility is that the high contrast between the screen and the dark interior of the stereoscope had a greater detrimental effect on identification of faces in negative than in positive images.

The lack of interaction between depth cue and lighting direction indicates that stereo information does not have any influence on the output of shading mechanisms. Further, the lack of interaction between depth cue and contrast polarity suggests that the photographic negative effect is not influenced by stereo processing. However, this should be interpreted cautiously because the effect of size of contrast polarity was rather small (3.5%). The small size is consistent with some previous findings with congruent contrast polarities being used for learning and test (eg Liu and Chaudhuri 1997). In fact, photographic-negative effects are often not detected in studies in which untextured faces are used (Bruce and Langton 1994; Liu et al 1999). To find whether stereo information modifies the photographic-negative effect, we tested performance for incongruent contrast conditions in the next experiment. We have previously shown that incongruent contrast between learning and test produces a strong effect under non-stereo conditions (Liu and Chaudhuri 1997).

3 Experiment 2

Disruption of shape-from-shading is often considered to be a major cause of the photographic-negative effect (reviewed by Kemp et al 1996). Although there is evidence for a greater role of pigmentation reversal (Bruce and Langton 1994; Liu et al 1999), that conclusion may be limited to conditions where the image contrast between learning and test is congruent, ie they are both negative. A disruption of shape-from-shading may be involved in incongruent learning–testing conditions where much stronger contrast-reversal effects have been reported. One possible explanation for this is that the 3-D surface map recovered from a positive image is different from the one recovered from its negative version, and therefore produces a mismatch between the two images of the same person. However, because the incongruent contrast conditions used in prior studies contained information of pigmentation, it is unclear whether the strong contrast-reversal effect was due to pigmentation reversal, shading reversal, or both. In no study to date incongruent learning–testing conditions have been examined with the contribution from each of these factors separated from each other. In this experiment, we used untextured face images to eliminate the contribution of pigmentation.

Our purpose was to investigate whether the hypothesised disruption of 3-D shape information from shading could be compensated for by the inclusion of stereo information. If the stereo module is relatively independent from the shading module, the positive and negative versions of a face should result in a matched 3-D surface, and therefore reduce the contrast-reversal effect due to disruption of shape-from-shading.

We were also interested in whether the effect of contrast reversal was related to angular difference between the apparent lighting of positive and negative pairs of images. Obviously, a lighting direction of 60° creates a much greater difference between

a positive image and the apparent lighting angle of its negative version than some smaller lighting angles, such as 30° or 0° . To investigate this issue, we used the same five lighting directions as in experiment 1.

3.1 Method

3.1.1 Subjects. Twenty-seven undergraduate students from McGill University participated in this study. All subjects passed the same screening criterion as in experiment 1 on the Titmus Stereo Test. All had normal or corrected-to-normal vision. The ages ranged from 17 to 32 years (median = 20 years).

3.1.2 Materials. These were the same as in experiment 1.

3.1.3 Procedure and design. We used a 2 polarity \times 5 lighting direction within-subject design. The procedure was the same as in experiment 1 except that the contrast polarity was always reversed between the two sequentially presented face images. Half of the time the positive image was presented first.

3.2 Results and discussion

3.2.1 Accuracy. The percentage accuracy data are shown in figure 4. A two-way ANOVA showed no significant result for the main effects of depth cues ($F_{1,23} = 2.51$, $p = 0.13$) or lighting direction ($F_{4,92} = 1.19$, $p = 0.32$). There was also no significant interaction ($F_{4,92} = 0.73$, $p = 0.57$).

An inspection of the overall results in figure 4 suggests an advantage for the stereo conditions. The grand mean was 71.9% for stereo and 67.8% for non-stereo conditions. We suspected that the reason the 4.1% difference was not significant may be due to a lack of power. To see whether this was the case, we combined the data from experiments 1 and 2 and performed a 3 (positive, negative, contrast reversal) \times 2 (stereo, non-stereo) \times 5 (lighting direction) ANOVA. As suspected, with the increased power, a small but significant advantage of stereo conditions (2.5%) was detected ($F_{1,86} = 4.55$, $p < 0.04$).

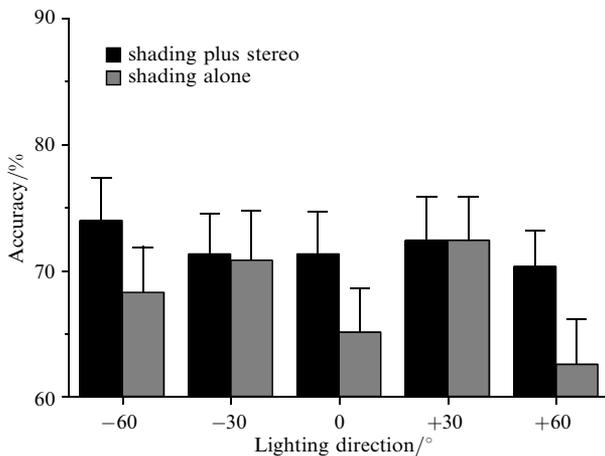


Figure 4. Percentage accuracy as a function of depth cues and lighting directions in experiment 2. The error bars represent standard error.

3.2.2 Reaction time. As in experiment 1, the median reaction time for the correct responses to each condition was computed for each subject. The means and standard errors are shown in figure 5. A logarithmic transformation was applied to the data set before it was submitted to ANOVA. Both main effects of depth cue and lighting direction were not significant. However, there was a significant interaction between these two ($F_{4,92} = 2.92$, $p < 0.03$). Pairwise comparisons with Tukey HSD tests did not yield any significant difference. Figure 5 seems to indicate that the interaction was caused by the faster response for the stereo than the non-stereo conditions (139 ms)

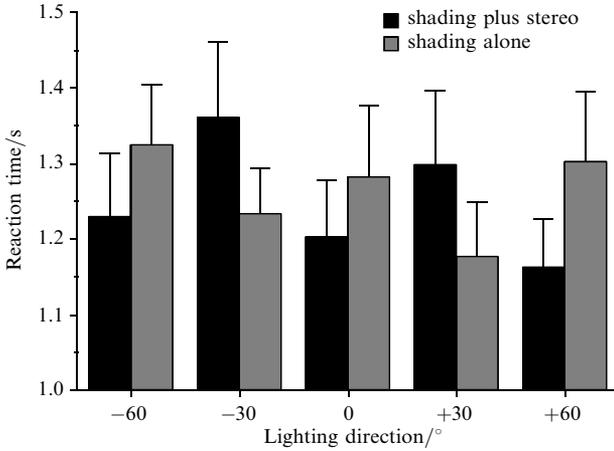


Figure 5. Reaction time as a function of depth cues and lighting directions in experiment 2. The error bars represent standard error.

when the lighting condition was $+60^\circ$. The error bars here had less overlap than in the results from any of the other lighting directions.

The results show that stereo information has little influence on the contrast-reversal effect. Also, the angular difference between apparent lighting of positive and negative images had little effect on identification. However, there were signs that stereo information was used by the subjects. First, the performance was slightly better when there were stereo cues than when they were absent. Second, the reaction-time data indicate that lighting angle difference affected the stereo condition less than the non-stereo condition.

4 General discussion

Studies on the role of shape information in face processing have suggested that face recognition may depend upon effective encoding of 3-D surface information. However, shape-from-shading is often ambiguous when used alone. An interesting question, therefore, is whether a combination of shading along with other depth cues that are abundant in real situations can produce a better estimation of 3-D information. Studies with simple 3-D objects seem to confirm this expectation because combining different depth cues often results in a more veridical estimation of 3-D surface geometry than if the individual cues are used alone (Bülthoff 1991; Uttal et al 1996). The question addressed in this study was whether this benefit also applies to face recognition.

It has been previously shown that face recognition based on shading alone is susceptible to bottom lighting and contrast reversal. The aim of our study has been to test whether these effects can be reduced when stereo information is combined with shading. Surprisingly, our main results from the two experiments showed only a weak effect of cue combination. Performance in both shading-alone and shading-plus-stereo conditions were affected evenly by lighting direction and contrast polarity, implying that adding stereo information did not change the effect found in the shading-alone condition. However, there was also evidence that stereo information was not totally neglected in our face-identification task either. First, the reaction time for the shading-plus-stereo condition was slightly faster than for the shading-alone condition when faces were lit from $+60^\circ$. Furthermore, although recognition performance in the two conditions was quite close, nevertheless there was a slight advantage (2.5%) in the shading-plus-stereo condition.

Because face perception may require different representations for different kinds of task, the generalisability of our results awaits further experimentation in other types of tasks such as sex identification. Bruce et al (1993) found that shading and pigmentation

play different roles in a sex-identification task. It would be interesting to examine whether the role of shading in this task is mediated by estimation of 3-D shape information or merely by the 2-D shading pattern. Also, it is not known whether adverse lighting and contrast conditions affect the precision of 3-D surface estimation from stereo. It may be that our subjects simply failed to extract accurate 3-D information from stereo under these conditions, or they estimated depth accurately but were unable to use this information to improve their performance. We are planning to perform a direct test of this by asking subjects to judge surface orientation of the face in these conditions presented stereoscopically.

4.1 *The role of shape-from-shading and shape-from-stereo in face processing*

Because 3-D shapes are estimated by different depth modules, questions arise as to how these estimations may be integrated. Bülthoff and Mallot (1988) listed several possible means of interaction. One of them is that unequivocal information from one depth cue can override the others. In 3-D shape-judgment tasks, they found that judgments based on shape-from-stereo are more accurate than ones based on shape-from-shading. They conclude that shape-from-stereo overrides shape-from-shading. Our results, however, suggest a greater role of shape-from-shading than that of shape-from-stereo in the face-identification task. The output from the shape-from-shading module appears to dominate the percept because none of the shading effects (ie lighting direction and contrast reversal) were corrected by information contained in the stereo module. In fact, this finding may be considered as a case in which shape-from-shading overrides shape-from-stereo. This is rather intriguing because the stereo cue is not equivocal. The apparent discrepancy between our result and that of Bülthoff and Mallot's may imply that the relative importance of depth modules is task-specific. In 3-D shape-judgment tasks, shape-from-stereo is likely to play a greater role than shape-from-shading; whereas in face-recognition tasks, shape-from-shading may play a greater role instead.

Another possibility is that shape-from-stereo cannot override shape-from-shading if some preferences assumed by the shading module are violated. This hypothesis has been supported by the work of Sun and Perona (1997) who found that shape-from-stereo failed to correct the search asymmetry effect due to bottom lighting. Their conclusion is that early 3-D mechanisms make a default assumption for top-lighting conditions independent of conflicting stereo information. Our result is consistent with this theory, even though there was a difference in our task demands. Both studies required that subjects perform tasks under bottom-lighting conditions and both have found that the difficulty due to bottom lighting cannot be compensated by stereo information.

Our results also show that contrast polarity may be an additional assumption used by the shape-from-shading module. However, this assumption may only apply to face stimuli because the contrast-reversal effect in recognition of non-face objects is lacking (Subramaniam and Biederman 1997). If so, it implies that stereo and shading processing is affected by top-down object knowledge.

The lack of enhanced identification performance after inclusion of stereo cues may suggest that stereo and shading modules are interactive rather than encapsulated. If the two modules had been encapsulated, the assumptions preferred in the shading module (such as positive and top lighting) would have had little influence on the processes occurring in the stereo module. As a result, the deficits due to lighting direction and contrast reversal should have been reduced.

4.2 *Can effects of lighting direction and contrast reversal be attributed to a disruption of 3-D shape processing from shading?*

The results of our study suggest that shape-from-stereo, although more reliable than shape-from-shading in estimating 3-D surface, contributes only minimally to face recognition. This reliance on a less robust 3-D cue is incompatible with the view that

face identification requires 3-D surface construction from various depth modules. But, such processing may have some degree of involvement in face perception, given our finding that recognition in shading-plus-stereo was slightly better than in shading-alone condition, although the magnitude of the effect shows that the involvement is very small.

This conclusion seems to contradict the standard explanation for the lighting-direction and contrast-reversal effects. A widely accepted explanation is that 3-D shape-from-shading information is disrupted under these circumstances. But if the computation of 3-D shape-from-shading is not important for face perception, as the data in this study suggest, it would be hard to advocate a disruption of this process as an explanation.

4.3 *Face matching mechanisms: 2-D versus 3-D accounts*

If disruption of 3-D processing cannot explain the lighting and contrast-reversal effects, then what is the alternative? The answer may lie within 2-D or view-based mechanisms (Bülthoff and Edelman 1992; Bülthoff et al 1995). There are currently several models that do not require descriptions of 3-D surface (eg Beymer and Poggio 1996; Wiskott et al 1997; Hancock et al 1998). Yet they have been shown to produce superior face recognition. One problem with these models is that they may not be able to account for the lighting-direction and contrast-reversal effects because they are unlikely to be affected by the same stimulus parameters. However, this can be solved if a 2-D model relies on a classification scheme that produces an expertise effect. For example, studies have shown that general categorical information such as gender and race can be represented by eigenvectors in principal component analysis (Valentin et al 1997). This approach is able to produce expertise effects such as recognising faces of one's own race better than those of different races. Assuming that the direction of lighting and sign can be categorised in the same fashion, faces frequently experienced in top-lighting and positive conditions should generalise better to new faces of the same conditions than in bottom-lighting and negative conditions.

Shading information processed in this manner does not require an interpretation of depth or surface orientation. Rather, it is treated simply as a 2-D pattern. However, because 3-D shape information is not totally discarded, as shown in this and another study involving object recognition (Edelman and Bülthoff 1992), an adequate way to characterise recognition processes may be to incorporate both 2-D and 3-D processing components, with much greater weight assigned to the output of 2-D processing. According to this account, the reason that statues or the laser-scanned untextured faces are so recognisable is that by duplicating 3-D shape, sculptors or 3-D scanners can recreate a shading pattern that resembles the original pattern in the models. Therefore, it is not duplication of the 3-D surface per se that is responsible for the saliency, but rather it is the by-product of the resulting 2-D layout of the shading pattern that is the critical determinant.

Acknowledgements. This research was supported by grants from the Medical Research Council of Canada and Natural Sciences and Engineering Research Council to AC. We thank our anonymous reviewers for their insightful comments on the earlier version of this paper.

References

- Bassili J N, 1978 "Facial motion in the perception of faces and of emotional expression" *Journal of Experimental Psychology: Human Perception and Performance* **4** 373–379
- Berry D S, 1990 "What can a moving face tell us?" *Journal of Personality and Social Psychology* **58** 1004–1014
- Berry D S, 1991 "Child and adult sensitivity to gender information in patterns of facial motion" *Ecological Psychology* **3** 349–366
- Beymer D, Poggio T, 1996 "Image representations for visual learning" *Science* **272** 1905–1909
- Braje W L, Kersten D, Tarr M J, Troje N F, 1998 "Illumination effects in face recognition" *Psychobiology* **26** 371–380

- Bruce V, Burton A M, Hanna E, Healey P, Mason O, Coombes A, Fright R, Linney A, 1993 "Sex discrimination: How do we tell the difference between male and female faces?" *Perception* **22** 131–152
- Bruce V, Healey P, Burton M, Doyle T, Coombes A, Linney A, 1991 "Recognising facial surfaces" *Perception* **20** 755–769
- Bruce V, Langton S, 1994 "The use of pigmentation and shading information in recognising the sex and identities of faces" *Perception* **23** 803–822
- Bülthoff H H, 1991 "Shape from X: Psychophysics and computation", in *Computational Models of Visual Processing* Eds M S Landy, J A Movshon (Cambridge, MA: MIT Press) pp 305–330
- Bülthoff H H, Edelman S, 1992 "Psychophysical support for a 2-D view interpolation theory of object recognition" *Proceedings of the National Academy of Sciences of the USA* **89** 60–64
- Bülthoff H H, Edelman S, Tarr M J, 1995 "How are three-dimensional objects represented in the brain?" *Cerebral Cortex* **5** 247–260
- Bülthoff H H, Mallot H A, 1988 "Integration of depth modules: Stereo and shading" *Journal of the Optical Society of America A* **5** 1749–1758
- Christie F, Bruce V, 1998 "The role of dynamic information in the recognition of unfamiliar faces" *Memory & Cognition* **26** 780–790
- Davies G, Ellis H, Shepherd J, 1978 "Face recognition accuracy as a function of mode of representation" *Journal of Applied Psychology* **63** 180–187
- Edelman S, Bülthoff H H, 1992 "Orientation dependence in the recognition of familiar and novel views of three-dimensional objects" *Vision Research* **32** 2385–2400
- Enns J T, Shore D I, 1997 "Separate influences of orientation and lighting in the inverted-face" *Perception & Psychophysics* **59** 23–31
- Erens R G, Kappers A M, Koenderink J J, 1993 "Perception of local shape from shading" *Perception & Psychophysics* **54** 145–156
- Galper R E, 1970 "Recognition of faces in photographic negative" *Psychonomic Science* **19** 207–208
- Galper R E, Hochberg J, 1971 "Recognition memory for photographs of faces" *American Journal of Psychology* **84** 351–354
- Gauthier I, Tarr M J, 1997 "Becoming a 'Greeble' expert: exploring mechanisms for face recognition" *Vision Research* **37** 1673–1682
- Hancock P J B, Bruce V, Burton M A, 1998 "A comparison of two computer-based face identification systems with human perceptions of faces" *Vision Research* **38** 2277–2288
- Hayes A, 1988 "Identification of two-tone images: some implications for high- and low-spatial-frequency processes in human vision" *Perception* **17** 429–436
- Hill H, Bruce V, 1993 "Independent effects of lighting, orientation, and stereopsis on the hollow-face illusion" *Perception* **22** 887–897
- Hill H, Bruce V, 1994 "A comparison between the hollow-face and 'hollow-potato' illusions" *Perception* **23** 1335–1337
- Hill H, Bruce V, 1996 "The effects of lighting on the perception of facial surfaces" *Journal of Experimental Psychology: Human Perception and Performance* **22** 986–1004
- Hill H, Schyns P G, Akamatsu S, 1997 "Information and viewpoint dependence in face recognition" *Cognition* **62** 201–222
- Johnston A, Hill H, Carman N, 1992 "Recognising faces: Effects of lighting direction, inversion, and brightness reversal" *Perception* **21** 365–375
- Johnston A, Passmore P J, 1994a "Shape from shading I: Surface curvature and orientation" *Perception* **23** 169–189
- Johnston A, Passmore P J, 1994b "Shape from shading II: Geodesic bisection and alignment" *Perception* **23** 191–200
- Kemp R, Pike G, White P, Musselman A, 1996 "Perception and recognition of normal and negative faces: the role of shape from shading and pigmentation cues" *Perception* **25** 37–52
- Knight B, Johnston R A, 1997 "The role of movement in face recognition" *Visual Cognition* **4** 265–273
- Lander K, Christie F, Bruce V, 1999 "The role of movement in the recognition of famous faces" *Memory & Cognition* **27** 974–985
- Leder H, 1996 "Line drawings of faces reduce configural processing" *Perception* **25** 355–366
- Liu C H, Chaudhuri A, 1997 "Face recognition with multi-tone and two-tone photographic negatives" *Perception* **26** 1289–1296
- Liu C H, Chaudhuri A, 1998 "Are there qualitative differences in face recognition between positive and negative?" *Perception* **27** 1107–1122
- Liu C H, Collin C A, Burton A M, Chaudhuri A, 1999 "Lighting direction affects recognition of untextured faces in photographic positive and negative" *Vision Research* **39** 4003–4009

- Liu C H, Collin C A, Rainville S J M, Chaudhuri A, in press "The effects of spatial frequency overlap on face recognition" *Journal of Experimental Psychology: Human Perception and Performance*
- Mamassian P, Kersten D, 1996 "Illumination, shading and the perception of local orientation" *Vision Research* **36** 2351–2367
- Näsänen R, 1999 "Spatial frequency bandwidth used in the recognition of facial images" *Vision Research* **39** 3824–3833
- Nelson C A, Horowitz F D, 1983 "The perception of facial expressions and stimulus motion by two- and five-month-old infants using holographic stimuli" *Child Development* **54** 868–877
- Parker A J, Cumming B G, Johnston E B, Hurlbert A C, 1995 "Multiple cues for three-dimensional shape", in *The Cognitive Neurosciences* Ed. M S Gazzaniga (Cambridge, MA: MIT Press) pp 351–364
- Phillips R J, 1972 "Why are faces hard to recognise in photographic negative?" *Perception & Psychophysics* **12** 425–426
- Pike G E, Kemp R I, Towell N A, Phillips K C, 1997 "Recognizing moving faces: The relative contribution of motion and perspective view information" *Visual Cognition* **4** 409–437
- Ramachandran V S, 1988 "Perception of shape from shading" *Nature (London)* **331** 163–166
- Ramachandran V S, Armel C, Foster C, Stoddard R, 1998 "Object recognition can drive motion perception" *Nature (London)* **395** 852–853
- Soken N H, Pick A D, 1992 "Intermodal perception of happy and angry expressive behaviors by seven-month-old infants" *Child Development* **63** 787–795
- Stucki M, Kaufmann-Hayoz R, Kaufmann F, 1987 "Infants' recognition of a face revealed through motion: Contribution of internal facial movement and head movement" *Journal of Experimental Child Psychology* **44** 80–91
- Subramaniam S, Biederman I, 1997 "Does contrast reversal affect object identification?" *Investigative Ophthalmology & Visual Science* **38**(4) S998
- Sun J, Perona P, 1997 "Shading and stereo in early perception of shape and reflectance" *Perception* **26** 519–529
- Tarr M J, Kersten D, Bülthoff H H, 1998 "Why the visual recognition system might encode the effects of illumination" *Vision Research* **38** 2259–2275
- Troje N F, Bülthoff H H, 1996 "Face recognition under varying poses: The role of texture and shape" *Vision Research* **36** 1761–1771
- Troje N F, Bülthoff H H, 1998 "How is bilateral symmetry of human faces used for recognition of novel views?" *Vision Research* **38** 79–89
- Troje N F, Symons L A, 1998 "Search asymmetries for shaded disks are not present for shaded faces" *Investigative Ophthalmology & Visual Science* **39**(4) S166
- Uttal W R, Liu N, Kalki J, 1996 "An integrated computational model of three-dimensional vision" *Spatial Vision* **9** 393–422
- Valentin D, Abdi H, Edelman B, 1997 "What represents a face? A computational approach for the integration of physiological and psychological data" *Perception* **26** 1271–1288
- Van den Enden A, Spekreijse H, 1989 "Binocular depth reversals despite familiarity cues" *Science* **244** 959–961
- Wiskott L, Fellous J-M, Kruger N, Malsburg C von der, 1997 "Face recognition by elastic bunch graph matching" *Proceedings of the IEEE International Conference on Image Processing, Santa Barbara, CA* volume 1, pp 129–132

