Pergamon

PII: S0042-6989(97)00165-X

Vision Res., Vol. 38, No. 1, pp. 79–89, 1998
© 1997 Elsevier Science Ltd. All rights reserved
Printed in Great Britain
0042-6989/98 $19.00 + 0.00

How is Bilateral Symmetry of Human Faces Used for Recognition of Novel Views?

NIKOLAUS F. TROJE,*† HEINRICH H. BÜLTHOFF*

Received 2 October 1996; in revised form 26 March 1997

The role of bilateral symmetry in face recognition is investigated in two psychophysical experiments using a Same/Different paradigm. The results of Experiment 1 confirm the hypothesis that the ability to identify mirror symmetric patterns is used for viewpoint generalization by approximating the view symmetric to the learned view by its mirror reversed image. The results of Experiment 2 show that the match between this virtual view and the test image is performed directly between the images. Performance drops dramatically if the symmetry between the intensity patterns of the learning and the testing view is disturbed by an asymmetric illumination, although the symmetry between the spatial arrangement of high-level features is retained. Experimental results are discussed in terms of their relation to existing approaches to object recognition. © 1997 Elsevier Science Ltd

Symmetry Face recognition Viewpoint invariance

INTRODUCTION

Human faces form an object class that has probably undergone more investigation than any other object class. Faces of different people are very similar to each other compared with differing objects in other object classes. Human perception is, however, extremely sensitive to even small differences between faces. This reflects the biological and social relevance that face recognition has to our species. A face that has been seen only once can be recognized even after large changes in orientation, expression or illumination conditions (Bruce, Valentine, & Baddeley, 1987; Moses, 1993; Troje & Bülthoff, 1996a). Face processing, however, involves not only the identification of particular persons but also a variety of different classification tasks concerning the sex, age, race, and attractiveness, as well as mood and intention of the person.

In this paper, we focus on a property that human faces share with many other biologically relevant objects: bilateral symmetry. Bilateral symmetry of the shape of the body with respect to a vertical axis is almost universal among vertebrates. Also, the majority of invertebrates show this kind of symmetry. Exceptions are mainly found among immovable or only slowly moving animals. This led to the assumption that bilateral symmetry is caused by breaking the perfect spherical symmetry of the single cell by the vertical force of gravity and the fore/aft asymmetry of the motion direction (Gardner, 1964). However, as Tyler (1994) pointed out, this argument would also predict a bilateral configuration of the inner organs which is—at least in vertebrates—not the case. In fact, the evolutionary constraint (whatever it is) that conserves bilateral symmetry seems to work on the outer, visible parts of the body. This implies that bilateral symmetry also plays a role in recognition and communication.

Not only is bilateral symmetry universal among animals, but so is the sensitivity of visual systems to symmetric patterns. Preference for symmetric patterns has been shown in a variety of different animals (Lehrer, Horridge, Zhang, & Gadagkar, 1994; Møller, 1993, 1995; Swaddle & Cuthill, 1994;). Pigeons (Delius & Novak, 1982), bees (Giurfa, Eichmann, & Menzel, 1996), and dolphins (Fersen, Manos, Galdowski, & Roitblat, 1992) have been successfully trained to generalize symmetry. The animals could be trained to respond to either only symmetric or to asymmetric patterns, even if they had not seen the particular pattern before. Humans also show a high sensitivity to symmetrical patterns (e.g. Biederman & Cooper, 1991; Julez, 1971; Wagemans, 1995) as well as to slight deviations from symmetry (Barlow & Reeves, 1979). Sensitivity to bilateral symmetry with respect to a vertical axis is much higher than with respect to a horizontal axis (Corballis & Roldan, 1975; Mach, 1903). A variety of different models have been proposed to describe human symmetry detection. A review of the field is provided in two special issues of the journal *Spatial Vision* (Tyler, 1994, 1995).

Different explanations for the striking convergence between bilaterally symmetric shapes and sensitivity to

---

*Max-Planck-Institut für biologische Kybernetik Spemanntraße 38, D-72076 Tübingen, Germany.

†To whom all correspondence should be addressed at current address: Department of Psychology, Queen's University, Kingston, Ontario, Canada U7L 3N6.

symmetrical patterns have been proposed. Several authors (Brookes & Pomiankowski, 1994; Watson & Thornhill, 1994) argue that preferences for symmetry have evolved because the degree of symmetry signals health—the more symmetric organisms being those having a potentially better phenotypic quality. According to another theory (Enquist & Arak, 1994) symmetry is the result of coevolution between a signal emitter and a receiver because it allows a better recognition of the signal, irrespective of its position and orientation in the visual field. Osorio (1996) argues that symmetry detection might have occurred as a by-product of edge detection mechanisms because it can be performed with a similar set of filters sensitive to spatial phase.

All the investigations mentioned above used tasks in which the subjects were required to detect symmetry within a single image. A related task has undergone much less investigation: humans can not only detect symmetry within a single image but also the symmetry between two successively presented images that are mirror reflected versions of each other. A great deal of evidence for this ability comes from related studies on mental rotation (Cooper, Schacter, Ballesteros, & Moore, 1992; Tarr & Pinker, 1990). However, there is only one recent study—at least to our knowledge—investigating systematically the ability to detect mirror symmetry between a memorized and an actual view. Quinlan (1995) measured the effect of symmetry in two different Same/Different paradigms. In the first experiment, he presented the two stimulus items simultaneously side by side. If the two items were mirror symmetric versions of each other they combined into a single mirror symmetric image. In these cases the subjects' performance was significantly improved compared with conditions in which the two items were not presented in a symmetrical arrangement. In the second experiment, he presented the two items one after the other with an interstimulus interval of up to 500 msec. In this experiment, he found the same effects of symmetry as in the previous one, although in this case, symmetry had to be detected not in a single image but between the stored training image and the current test image.

The ability to identify two images that are mirror symmetric to each other could be used for viewpoint generalization within classes of bilaterally symmetric objects by exploiting the fact that the views taken from symmetric viewpoints with respect to the symmetry plane of the object often result in mirror symmetric images (Vetter, Poggio, & Bülthoff, 1994; Vetter & Poggio, 1994). Bilateral symmetry of an object is expressed by the fact that the spatial arrangements of the corresponding features in the two symmetric views are mirror symmetric to each other. Usually the images themselves are also mirror symmetric to each other; however, there are situations in which this is not the case. If, for instance, the object is illuminated by a strong point light source that is located outside the plane defined by the rotation axis of the object and the observer's viewpoint, the grey-level patterns of the images can deviate significantly from mirror symmetry, although the mirror symmetry between the spatial arrangements of the features is still retained (Fig. 1). A mechanism that does not evaluate the mirror symmetry between the entire images but takes into account only the mirror symmetry between the spatial arrangements of salient features would thus be more robust to illumination changes. On the other hand, it would require more computational effort, because higher order features would have to be detected, indexed and compared.

In this paper, we investigate how bilateral symmetry is used in the recognition process. Our hypothesis is that the ability to identify mirror symmetric images is used for viewpoint generalization by approximating the symmetric view of a learned view using its mirror symmetric image. The hypothesis leads to the prediction that the mirror reversed image of a learned view should be recognized better than the realistic symmetric view, even in cases in which mirror reversal results in an unrealistic and impossible image of the target face. Furthermore, we want to find out whether we have direct access to the bilateral symmetry of the 3D object by extracting features and using the mirror symmetry of their arrangement or whether we are restricted to the mirror symmetry of the image.

The paper is organized as follows: we first present two psychophysical experiments. In Experiment 1, we measure the increase in generalization performance that can be achieved by exploiting the ability to identify mirror symmetric images.

In Experiment 2, we focus on decoupling the bilateral symmetry of the 3D faces from the mirror symmetry between the images yielded by symmetric views by using different lighting conditions.* Finally, the experiments and their results will be discussed and related to existing approaches to object recognition.

GENERAL METHODS

Stimuli

The images were created using 98 surface models from a data base of 3D head models (for details, see Troje & Bülthoff, 1996a). Each model consists of a range data map providing the geometry of the head and a texture map accounting for its local reflectance. The head models did not contain distinctive features such as glasses, beards or earrings. The hair had been removed digitally, because the shape of the hair might also provide easy viewpoint independent features, and because our scanning technique had problems digitizing the hair.

---

*The effects of different light conditions on the image of a 3D object have been studied extensively by several authors (e.g., Belhumeur & Kriegman, 1996; Braje, Kersten, Tarr, & Troje, 1996; Hallinan, 1994; Johnston, Hill, & Carman, 1992). A detailed discussion of their work is beyond the scope of this paper. Here, changing illumination just serves as a tool to change the image of a face without changing the spatial layout of its features.

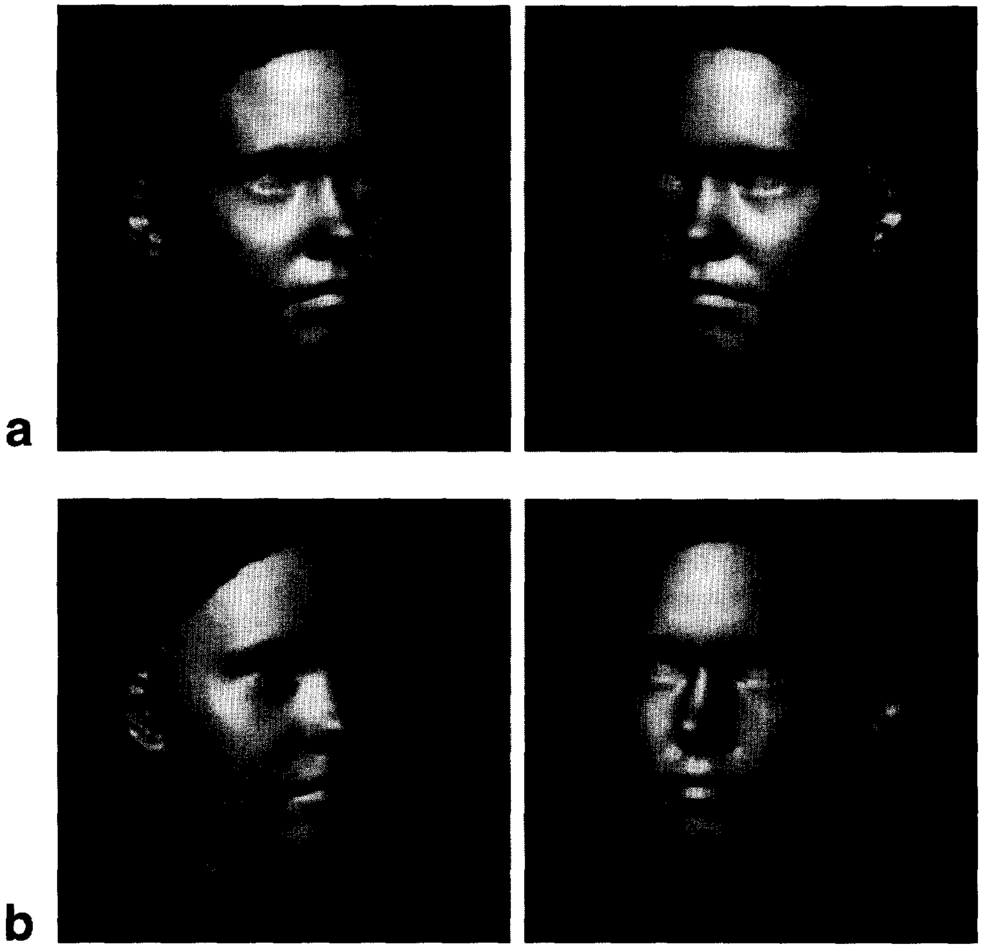

FIGURE 1. Symmetric views of the same face. If the face is illuminated by a light source positioned in the plane defined by the viewing position and the rotation axis of the face, the resulting images are approximately mirror symmetric to each other (a). If the light source is not within that plane the two images are no longer mirror symmetric to each other. Only the symmetry between the spatial arrangements of the features is retained (b).

Images had a size of 256 × 256 pixels. The height of the faces on the screen was approximately 6 cm, subtending a region of approximately 4.5 × 4.5 deg of visual angle at the position of the observer.

The faces were rendered without using the original texture information. Instead, they were rendered by assuming homogeneous reflectance and by applying an illumination model. We did this for two reasons. First, we had observed in previous experiments (Troje & Bülthoff, 1996a) that most effects concerning the generalization to novel views of faces are qualitatively the same for naturally textured faces and for faces deprived of texture, but quantitatively much more pronounced for faces without texture. The reason for this is certainly the presence of more inherently viewpoint invariant features in the texture of a face, resulting in better generalization performance. Second, we wanted the changes in the images owing to changing illumination in Experiment 2 to be as pronounced as possible.

The illumination model assumed Lambertian reflection and had a single light source 2 m away from the face. A small amount of ambient light (12% with respect to the illuminance of the point light source) was also added. The vertical position of the light source was always 20 deg above the equator. The distance between camera and face was 1.30 m. For the stimuli used in Experiment 1 the azimuthal position of the light source was always identical to the azimuthal position of the simulated camera. Each face was rendered in eight different orientations. Their angles with respect to the frontal view of the face were −90, −67.5, −45, −22.5, 22.5, 45, 67.5 and 90 deg. For Experiment 2, each face was rendered with the four different combinations of two orientations and two light source positions: the head was oriented either 15 deg to the left or 15 deg to the right with respect to the frontal view. The light source was positioned either 35 deg to the left of the camera or 35 deg to the right of the camera.

Subjects

A total of 28 subjects participated in this study. There were 14 subjects for each of the two experiments. They were mainly undergraduate students from Tübingen University and were paid DM 15 per hour. They were not familiar with the presented faces.

Design and procedure

We used a Same/Different recognition task. Subjects were sequentially presented with two images of faces. The task was to judge whether the images showed the same person or not, regardless of any change in viewing conditions. The answer was required to be given "as accurately and as quickly as possible".

Each trial was initiated by hitting the SPACE bar on a computer keyboard. A fixation cross appeared for 1000 msec on the screen. Then the learning view was briefly shown, immediately followed by a random mask. The presentation time for the learning view differed in the two experiments. In Experiment 1, learning views were shown for 700 msec. In Experiment 2, they were shown for 300 msec. The task of the second experiment was somewhat easier and we used a shorter presentation time to achieve the same overall error level as in Experiment 1. The mask was shown for 1100 msec (in both experiments). After that, the fixation cross appeared again for 1000 msec and finally the testing view was shown. The testing view remained on the screen until the subject responded. For each trial the subject's response and the response time were recorded.

There were four within-subject conditions in each experiment, corresponding to the combination of viewing conditions in the learning and testing images. The conditions themselves varied between the two experiments and are described in more detail for each experiment. Each subject performed 256 trials, 64 in each condition. The 96 face models appeared exactly four times each, once in each condition. Except for this constraint, the assignment of the faces to the different trials was randomized for each subject.

One half of the trials in each condition paired a face with itself ("Same" response expected) and one half of the trials paired a face with another face of the same gender ("Different" response expected). The presentation order of the trials within the experiment was randomized for each subject.

EXPERIMENT 1

Purpose

In the first experiment we tried to document and to quantify the viewpoint generalization advantage that can be achieved by exploiting the ability to identify mirror symmetric images. Real human faces are never perfectly bilaterally symmetric and therefore images taken from symmetric viewpoints are not perfectly mirror symmetric. In this experiment, we tested both symmetric views and perfectly mirror symmetric images of the learned views. Since real faces always have slight asymmetries, the perfectly mirror symmetric image is, in fact, an impossible and unrealistic view of the learned face. However, if generalization is based on a virtual view derived by flipping the learned image, the mirror symmetric image should be identified with the learned view better than the actual symmetric view is.

Methods

In this experiment, we used images rendered with a light source positioned above the camera location. Thus, images taken from symmetric viewpoints resulted in roughly mirror symmetric images. The only sources of slight asymmetries between the images are the deviations from a perfect bilateral symmetry of the faces. The four conditions were the following [Fig. 2(a)]:

Condition A: The learning and testing images showed the faces from the same orientation.

Condition B: The learning and testing images showed the faces from symmetric orientations.

Condition C: The learning and testing images showed the faces from otherwise different orientations according to the following table.

Learning	Testing	Learning	Testing
+ 22.5	− 45	+ 67.5	− 22.5
− 22.5	+ 45	− 67.5	+ 22.5
+ 45	− 90	+ 90	− 67.5
− 45	+ 90	− 90	+ 67.5

This scheme yields a mean orientation change (i.e. the angle between learning and testing views) of 112.5 degrees. This is the same mean orientation change as in conditions B and D.

Condition D: Same as condition B but instead of the symmetric orientation, the mirror symmetric image of the learning view was shown as the testing view.

In each of the conditions, faces were shown equally often from one of the eight possible orientations. In half of the trials the same face was shown; in the other half, different faces were shown. Note that the distinction between conditions B and D makes sense only for the trials showing the same faces.

Results

We ran ANOVAs on both the error rates and the response times. In addition to the factor coding for the four viewing conditions, we introduced a second factor with two levels indicating whether a trial showed images from the same face or from different faces. The ANOVA for the error rate revealed a reliable main effect for the viewing conditions ($F_{3,39} = 14.45$, $P < 0.01$) and no main effect for the Same/Different conditions ($F_{1,13} < 1$). In addition, there was a significant interaction between the two factors ($F_{3,39} = 13.09$, $P < 0.01$). For the response times there was an effect for the viewing condition ($F_{3,39} = 11.40$, $P < 0.01$), an effect for the Same/Different condition ($F_{1,13} = 8.271$, $P < 0.05$), but only a marginal effect for their interaction ($F_{3,39} = 2.655$, $P = 0.06$).

In order to be able to make *post hoc* comparisons, we also ran separate ANOVAs for the miss rate and the false alarm rate by using either only the trials showing the same faces or only the trials showing different faces. The effect of the viewing conditions on the miss rate is

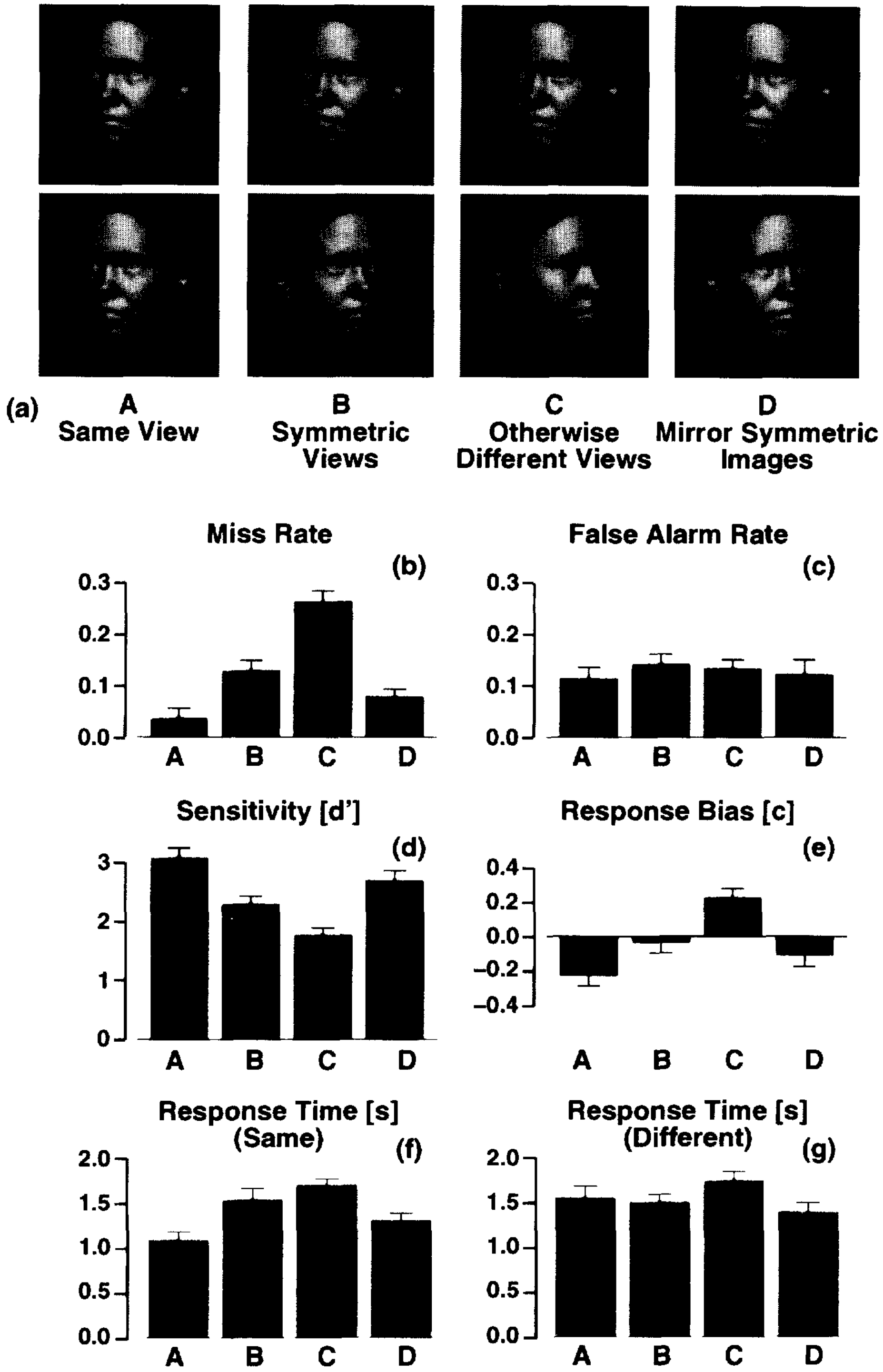

FIGURE 2. Conditions (a) and results (b)–(g) of Experiment 1. Error bars indicate the normalized standard error of the mean.

significant ($F_{3,39} = 34.9$, $P < 0.01$), the effect on the false alarm rate is not significant ($F_{3,39} = 0.4$). Similarly, we ran separate ANOVAs for the response times using either only the trials showing the same faces or only the trials showing different faces. If the same faces were shown, the effect of the viewing conditions was significant ($F_{3,39} = 11.3$). If different faces were shown, there was no significant effect ($F_{3,39} = 2.5$, $P > 0.05$). For *post hoc* comparisons, we used Tukey's Honestly Significant Difference (HSD). The critical difference for the miss rates was $d_T(5\%) = 0.0632$ and the critical distance for the corresponding response times was $d_T(5\%) =$ 297 msec.

The results are presented in Fig. 2. Error rates and response times are plotted separately for the trials showing the same faces and the trials showing different faces. Figure 2(b, c) show how miss rates and false alarm rates depend on the four conditions. Miss rates were more strongly affected than false alarm rates, which were more or less constant. Identification of symmetric views was worse than identification of the same views ($d_{A,B} = 0.0915$, $P < 0.05$), but still much better than of otherwise different views ($d_{B,C} = 0.1340$, $P < 0.05$). Performance in the condition showing mirror symmetric images was better than in the condition showing symmetric views ($d_{B,D} = 0.058$). The difference is a little bit smaller than the critical value of the HSD. Since 12 of the 14 subjects had a smaller error rate in condition D than in condition B, we nevertheless take this difference as reliable. A paired *t*-test on the miss rates of conditions B and D yielded a value of $t = 2.49$ ($P < 0.05$). Figure 2(d, e) show sensitivity (d′) and response biases (c) using the measures provided by signal detection theory (Macmillan & Creelman, 1991). Because of the constant false alarm rate, the d′ values show a similar pattern as the miss rates. The bias values indicate a tendency to respond "Different" in the different view condition (condition C) and a bias to respond "Same" in the other conditions.

In Fig. 2(f, g) response times are presented. The mean response time was slightly longer for the trials showing different faces than for trials with the same faces (1571 vs 1431 msec). The pattern of the response times for trials with the same faces was very similar to that for the miss rates. The response times for the trials showing different faces differed slightly from the corresponding false alarm rates. Response times were longest in the condition showing completely different views.

Discussion

Generalization performance to the symmetric view of a face is much better than generalization to otherwise different views. There is still a difference, however, between the performance in conditions A (same orientation) and B (symmetric orientation). This difference decreases if we use mirror reversed images instead of symmetric views. We conclude that mirror reversal is perceptually "inexpensive" and causes few additional errors even if it results in an unrealistic view of the learned face. The difference in performance between conditions A and B is most likely due to deviations from perfect bilateral symmetry in the faces. This is surprising since these asymmetries are usually not very pronounced, especially in the stimuli used in this experiment, because they were deprived of any features from which strong asymmetries could emerge. They did not contain hair or texture. Hair could cause asymmetries in shape. Texture contains asymmetries due to scars or blemishes and allows one to see the asymmetry that emerges when the eyes are not looking straight ahead. The fact that the remaining asymmetries still cause a reliable difference in generalization performance demonstrates an amazing sensitivity to asymmetries in the shape of faces.

Note that the elimination of asymmetries in condition D was not performed by eliminating the asymmetries in the 3D head model but by flipping the images. Thus, the heads seen in the learning and the testing image were no longer identical (even if the "Same" faces were used) but were mirror symmetric copies of each other. Nevertheless, a view of this modified head is treated as being more similar to the learned view than a symmetric view of the identical head. The stimulus seems to be supernormal in the sense of Tinbergen (1951). The similarity between the images seems to be more important than a correct match between the underlying 3D structures.

Another aspect of the results points in the same direction. The diagrams in Fig. 2 show that the differences in performance between the four conditions are mainly due to differences in the miss rates. The false alarm rates are very similar for all conditions. This can be explained if one assumes that subjects match the images rather than the 3D shapes. Images of different people are always different, even if they are shown from the same viewpoint. If discrimination were based on higher order features (e.g., the shape of the nose or the distance between the eyes), then discrimination between different faces shown from the same or from symmetric viewpoints would be expected to be better than if the faces were shown from otherwise different views. Seen from similar viewpoints, differences between the features and their metric relation could be detected more reliably.

The response biases plotted in Fig. 2(e) show that there is a tendency to respond "Same" in conditions with same or symmetric views and "Different" if the views were otherwise different. Even if the learning and testing images show different persons, the similarity between the images taken from the same viewpoint is higher than between images taken from different viewpoints. This image similarity seems to influence the response behaviour. Assuming equivalence between mirror symmetric images, the same argument can explain the good performance in the conditions using symmetric images.

EXPERIMENT 2

Purpose

Matching the learning and testing images requires some kind of mental transformation between them. What is the nature of this transformation? We consider two possibilities. The transformation could be based on the extraction of parameters describing the 3D scene (including information about the 3D shape of the object and illumination conditions) and a subsequent transformation of these parameters. Alternatively, it could be a simple image transformation.

A recognition system based on the extraction of scene attributes would yield a more flexible recognition system. A scene-based description is inherently viewpoint invariant. However, much computational effort is needed to extract this information. Image transformations, on the other hand, might be performed much faster and with only very basic or even without any knowledge about the content and significance of the image. Some basic knowledge (e.g., about the object class) might be needed to restrict the operation to appropriate situations. The symmetry operation, for instance, might help to generalize to the symmetric view of a bilaterally symmetric object, but it is otherwise of limited value.

With Experiment 2 we want to find out on what perceptual level symmetry information is processed. Do we only use the mirror symmetry between two images or can we extract the bilateral symmetry of the 3D face from the image and use it for recognition? In Experiment 1, stimuli were generated simulating a light source above the location of the camera. Consequently, the symmetric views always resulted in more or less mirror symmetric images. In Experiment 2, however, we used images that were rendered by simulating a light source that was no longer positioned above the camera, but was instead 35 deg either to the right or to the left of it. This allowed us to dissociate the bilateral symmetry of the 3D face from the mirror symmetry between the images taken from symmetric viewpoints. Images taken from symmetric viewpoints but with a fixed position of the light source are no longer mirror symmetric to each other on the level of the pixel intensities. The symmetry between the spatial arrangements of the features in the face, however, is still retained.

The faces as shown in this experiment were always oriented either 15 deg to the left or 15 deg to the right of the frontal view. We did that to ensure that both halves of the face were visible in all of the images. The images thus provided enough information to reconstruct the realistic symmetric view.

Methods

For the learning view the four possible combinations of orientation (left or right) and light source position (left or right) were used equally often. The orientation and the illumination in the testing view were determined by the following four conditions [Fig. 3(a)]:

Condition A: The learning and testing images showed the faces in the same orientation and with the same illumination.

Condition B: The learning and testing images showed the faces in symmetric orientations but with the position of the light source fixed.

Condition C: Both the orientation of the face and the position of the light source were changed to their symmetric positions.

Condition D: Instead of the symmetric viewing conditions used in Condition C, the mirror symmetric image of the learning view was shown as the testing view.

Results

As in Experiment 1, we calculated 4×2 ANOVAs modelling the error rates and the response times. The first factor accounted for the four symmetry conditions and the second was introduced to indicate whether a trial showed images from the same face or from different faces. The symmetry conditions had a significant effect on error rates ($F_{3,39} = 16.23$, $P < 0.01$) and response times ($F_{3,39} = 4.49$, $P < 0.01$). The Same/Different condition had a marginal effect on error rates ($F_{1,13} = 6.58$, $P < 0.05$). For the "Same" trials, we recorded a slightly higher error rate than for the "Different" trials. Response times were not affected by this factor ($F_{1,13} = 6.117$, $P > 0.05$). The interaction between the two factors significantly affected both the error rate ($F_{3,39} = 17.76$, $P < 0.01$) and the response time ($F_{3,39} = 4.30$, $P < 0.05$). As a basis for *post hoc* comparison, we also ran separate ANOVAs on the miss rate and on the false alarm rate as well as on the response times for the trials using the same faces and the trials using different faces. Significant effects were measured only for the miss rates ($F_{3,39} = 30.36$, $P < 0.01$) and for the response times using the same faces ($F_{3,39} = 9.29$, $P < 0.01$) but not for the false alarm rates and for the response times in the "Different" trials.

For *post hoc* comparisons Tukey's Honestly Significant Difference (HSD) was calculated, revealing a value of $d_T(5\%) = 0.056$ for the miss rates and a value of $d_T(5\%) = 133$ msec for the response times when using same faces.

Figure 3(b, c) show miss and false alarm rates. As in Experiment 1, the miss rate varied to a much greater degree across the four symmetry conditions than did the false alarm rate. The miss rate was much higher when only the orientation of the face changed (condition B) than when both orientation and illumination changed (condition C) ($d_{B,C} = 0.109$, $P < 0.01$). Showing mirror symmetric images (condition D) instead of symmetric viewing conditions (condition C) further lowered the miss rate ($d_{C,D} = 0.058$, $P < 0.05$) to a value that was statistically indistinguishable from the miss rate yielded when using identical viewing conditions (condition A) ($d_{A,D} = 0.031$). The false alarm rates were almost constant.

Figure 3(d, e) show the values for sensitivity and response bias. As a consequence of the constant false alarm rates, both the sensitivities and the response biases reflect the pattern of the miss rates. Higher miss rates correspond to smaller d' values but also shift the tendency to respond "Same" in condition A towards a tendency to respond "Different" in conditions B and C.

Figure 3(f, g) present the response times. As in Experiment 1, response times for the trials in which the same face was shown during learning and testing followed the same pattern as the miss rates. However, except for the difference of response times in conditions A and B, the differences were below the corresponding HSD.

Discussion

The approximate mirror symmetry in the arrangement of the features in the face was retained between the learning and the testing view in condition B. The grey-level patterns, however, were no longer mirror symmetric, corresponding to a strong reduction in recognition performance with respect to condition C. Subjects

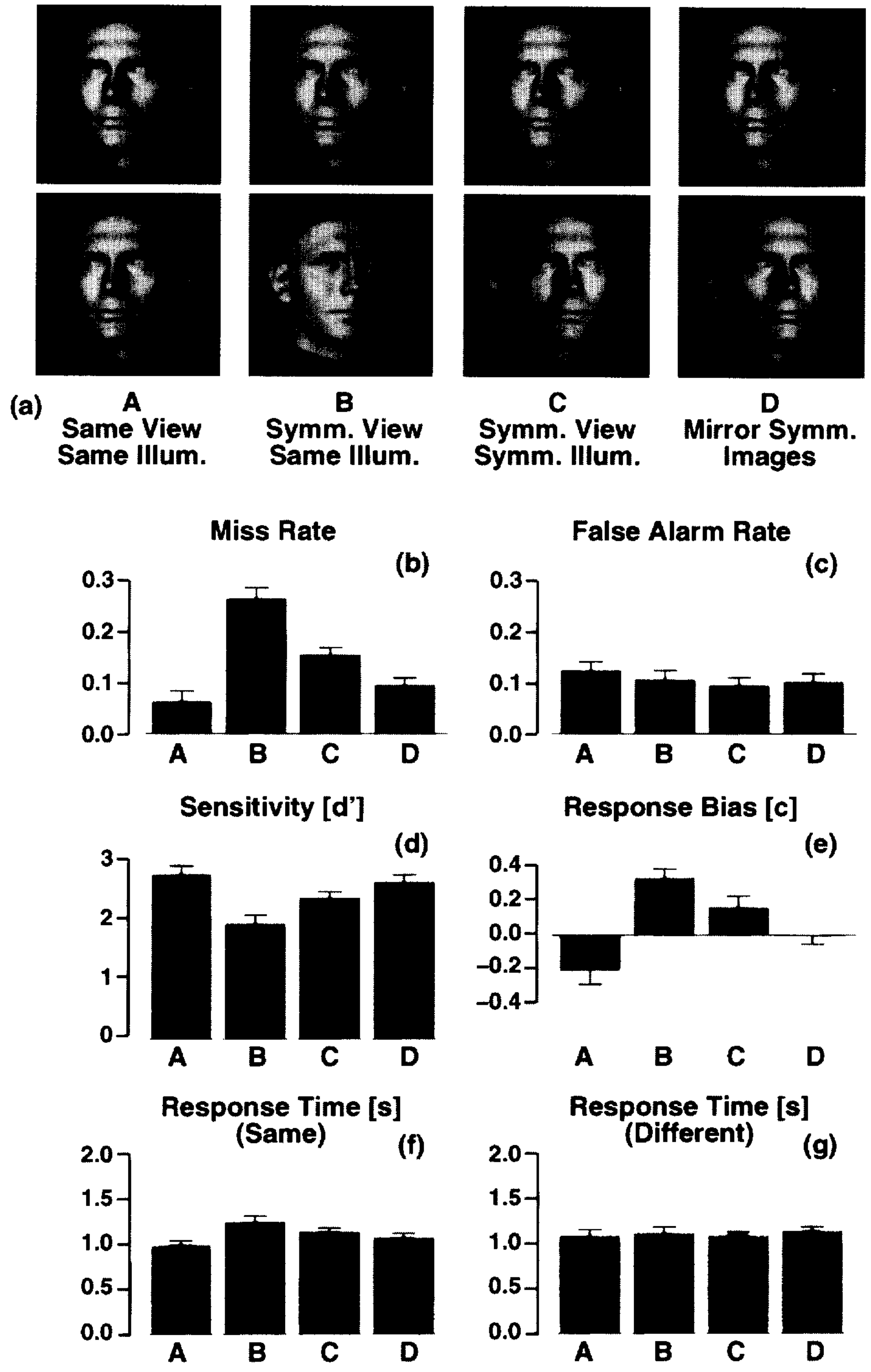

FIGURE 3. Conditions (a) and results (b)–(g) of Experiment 2. Error bars indicate the normalized standard error of the mean.

seemed not to be able to use the mirror symmetry in the spatial arrangement of the features. They exploited mirror symmetry between the images but did not seem to have direct access to the bilateral symmetry of the 3D face.

Most of the other issues discussed concerning Experiment 1 are also relevant for Experiment 2. The difference in the performance between conditions A and C reflects a large sensitivity to the minor asymmetries that are still present. The symmetry operation itself accounts only for a minor part of the performance drop. Mirror symmetry between the images seems to be more important for matching two objects than does the consistency of the 3D interpretation of the images.

The experiment and its results can be described from a slightly different and more general viewpoint by considering the nature of the mental representation of the faces to be compared. The distance between learning and testing views in the four conditions can be described either in terms of higher order attributes of the scene depicted in the image, or in terms of an image distance. In condition A, in which the images were presented with identical viewing conditions, the distance is smallest in both cases. In terms of a scene based description, the distance is larger in condition B with one attribute changing (the orientation of the face), but still smaller than in condition C in which two attributes change (the orientation of the face and the location of the light source). The changes between learning and testing view in condition D can also be described in terms of scene parameters. Here are not only the orientation of the head and the position of the illumination different, but also the face itself. The face presented in the testing image is no longer identical to the one seen in the learning image, but it is the mirror reversed version of it.

The order of the distances in conditions B, C and D is reversed if we use an image based distance measure that is insensitive to mirror reversal. The distance between the learning and testing image in condition D, in which the images are perfectly mirror symmetric, is as small as in condition A. In condition C, in which the images deviate only slightly from mirror symmetry, the distance is still small, and in condition B, it is greatest.

The data clearly show a pattern that is consistent with an image based distance rather than a distance based on scene attributes. The difference in performance between condition A (identical viewing conditions) and condition C (symmetrical orientation and symmetrical illumination) is most likely due to slight asymmetries between the resulting images. The symmetry operation itself causes only a very small increase in error rates.

GENERAL DISCUSSION

The results of Experiment 1 showed that the performance to generalize to a new viewpoint does not simply depend on the angle between the learning and the testing view. Images of faces taken from symmetric views and thus resulting in more or less mirror symmetric images are much better recognized than images taken from otherwise different views. With Experiment 2, we could show that the mirror reversed image of a learned view of a face is in fact used as an approximation of the symmetric view of the face. Two views are treated as if they are taken from the same face when the images are mirror symmetric to each other. Bilateral symmetry of the 3D face as expressed by the mirror symmetry between the spatial arrangements of the features in the images does not appear to be exploited. Information processing as expressed in these experiments seems to be mainly image based.

This makes the whole notion of "generalization" to a novel orientation or a novel illumination questionable. This notion suggests that generalization to a new instance of a scene attribute is accomplished and can be measured independently of other attributes. Recognition seems, however, not to be based on the extraction of scene attributes but rather on an image based comparison between the learned and the tested instance of an object. The attempt to measure the dependence of recognition performance on such attributes can yield inconsistent and unclear results. For an example, see the discussion of the results of Bruce *et al.* (1987) in Troje & Bülthoff (1996a).

If we try to describe recognition performance as being affected by different scene attributes, we have to be aware of very prominent interactions between these attributes. The situation in Experiment 2 provides an example. Generalization to a new view causes a decrease in recognition performance (compare conditions A and B). The same is true if subjects have to generalize to a new illumination (Braje, Kersten, Tarr, & Troje, 1996; Troje & Bülthoff, 1996b). If both attributes are changed, the effects of the changes in both attributes do not add to but partly cancel each other. If the orientation of the face had already been changed, then a related change in illumination can lead to an increase in performance (compare conditions B and C). Subjects do not generalize to new instances of scene attributes. They compare images. The way this comparison is accomplished, however, reflects an adaptation to the requirements of recognizing objects under changing viewing conditions.

Do we treat mirror reversal as being so inexpensive only when we know that we are dealing with a bilaterally symmetric object? Is the cost associated with this operation only so small because we already know that we are confronted with faces? Or do we take into account the false identification of two asymmetric objects that are mirror symmetric to each other? The results of our experiments do not provide an answer to these questions. The likelihood that we are faced with a situation in which such a false identification could occur is, however, so small that we probably could easily afford the assumption that two mirror symmetric images show the same object from symmetric views. The only case in which this becomes a problem is a very modern one compared with the time range relevant for the evolution of our cognitive system: the Latin alphabet has some letters such as b and d or p and q that would be confused. In fact, young children confuse these letters more frequently than others, when learning to read and write. They obviously have to learn not to identify two mirror symmetric images.

Note that the same, somewhat artificial situation occurs in the early experiments on mental rotation by Shepard & Metzler (1971). These authors investigated the generalization performance to new views of simple three-dimensional objects using a Same/Different paradigm. In the Different trials, they always used two objects that were mirror symmetric to one another. As for letter recognition, subjects could come into the situation where two mirror symmetric images would not show the same, but different objects. In this experiment subjects were thus explicitly required to distinguish between two mirror symmetric objects. The difficulty of this task is reflected in long response times of up to several seconds.

Different models of object recognition have been developed in the past. How do our findings relate to these approaches? Ullman (1989) classified current and past models of object recognition into three major groups: (1)

invariant properties methods; (2) parts decomposition methods; and (3) alignment methods. This classification scheme focuses on the way objects are represented and how these representations are matched. Invariant properties methods are based on a representation of the object in terms of higher order features. The features should fulfil the following criteria: (a) they can be derived from the image; (b) they are to a large extent independent of the viewing conditions; and (c) they are diagnostic, that is, they are shared by all views of the object but not by views of other objects. Use of such features would be ideal for solving the recognition task but in practice they are not easy to find. Parts decomposition methods cope with this problem by decomposing objects into generic parts that are so simple that it is easier to find invariants for each of them. Alignment methods, finally, are based on pictorial descriptions. The basic idea is to compensate for the transformations separating the viewed object and the corresponding stored model and then compare them.

Our data can best be interpreted within the framework of an image based alignment approach (Ullman, 1989; see also Poggio & Edelman, 1990). They are not consistent with invariant properties methods (e.g. Pitts & McCulloch, 1947) or with parts decomposition methods (e.g. Biederman, 1985). These descriptions assume the extraction of features (or parts) such as eyes, nose and mouth. An easily derivable property that is invariant with respect to symmetric views (even with nonsymmetric illumination) would be metric information about the relationship of the locations of such features. However, such information appears not to be used. Descriptions based on either invariant properties or on parts decomposition should not change when lighting changes. The distance between the learning and the testing image in conditions B and C in Experiment 2 should be about the same. The subjects' responses, however, indicated that this was not the case.

We are aware that the present results and the conclusions drawn from them might be restricted to the Same/Different paradigm that we used in these experiments. Learning and testing views were shown immediately one after the other with only a 2 sec interval between them. Only short-term episodic memory is needed to perform this task. The visual representations used to perform other tasks might be organized in a completely different way. The face of a well known friend might well be represented using invariant properties or models of the entire 3D structure, and it might be worthwhile to run experiments similar to the ones presented here but with different recognition paradigms that address different kinds of memory.

REFERENCES

Barlow, H. B. & Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research, 19*, 783–793.

Belhumeur, P. & Kriegman, D. (1996). What is the set of images of an object under all possible lighting conditions? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 270–277.

Biederman, I. (1985). Human image understanding: recent research and a theory. *Computer Vision, Graphics and Image Processing, 32*, 29–73.

Biederman, I. & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception, 20*, 585–593.

Braje, W. L., Kersten, D., Tarr, M. J. & Troje, N. F. (1996). Illumination and shadows influence face recognition. *Investigative Ophthalmology & Visual Science, 37*, S176.

Brookes, M. & Pomiankowski, A. (1994). Symmetry in the eye of the beholder. *Trends in Ecology & Evolution, 9*, 201–202.

Bruce, V., Valentine, T. & Baddeley, A. (1987). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology, 1*, 109–120.

Cooper, L. A., Schacter, D. L., Ballesteros, S. & Moore, C. (1992). Priming and recognition of transformed three-dimensional objects: effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 43–57.

Corballis, M. C. & Roldan, C. E. (1975). Detection of symmetry as a function of angular orientation. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 221–230.

Delius, J. D. & Novak, B. (1982). Visual symmetry recognition in pigeons. *Psychological Research, 44*, 199–212.

Enquist, M. & Arak, A. (1994). Symmetry, beauty and evolution. *Nature, 372*, 169–172.

Fersen, L. von, Manos, C. S., Galdowski, B. & Roitblat H. (1992) Dolphin detection and conceptualization of symmetry. In J. Thomas (Ed.), *Marine mammal sensory systems* (pp. 753–762). New York: Plenum Press.

Gardner, M. (1964). *The ambidextrous universe; left, right and the fall of parity*. New York: Mentor.

Giurfa, M., Eichmann, B. & Menzel, R. (1996). Symmetry perception in an insect. *Nature, 382*, 458–461.

Hallinan, P. W. (1994). A low-dimensional representation of human faces for arbitrary lighting conditions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 995–999.

Johnston, A., Hill, H. & Carman, N. (1992). Recognizing faces: effects of lighting direction, inversion and brightness reversal. *Perception, 21*, 365–375.

Julez, B. (1971). *Foundations of cyclopean perception*. Chicago: University of Chicago Press.

Lehrer, M., Horridge, G. A., Zhang, S. W. & Gadagkar, R. (1994). Shape vision in bees: innate preference for flower-like patterns. *Philosophical Transactions of the Royal Society of London B, 347*, 123–137.

Mach, E. (1903) *Analyse der Empfindungen*. Jena: Fischer.

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

Møller, A. P. (1993). Female preference for apparently symmetrical male sexual ornaments in the barn swallow *Hirundo rustica*. *Behavioural Ecology and Sociobiology, 32*, 371–376.

Møller, A. P. (1995). Bumblebee preference for symmetrical flowers. *Proceedings of the National Academy of Sciences USA, 92*, 2288–2292.

Moses, Y. (1993). Face recognition. generalization to novel images. Ph.D. Thesis, Applied Math. and Computer Science, The Weizmann Institute of Science, Israel.

Osorio, D. (1996). Symmetry detection by categorization of spatial phase, a model. *Proceedings of the Royal Society of London B, 263*, 105–110.

Pitts, W. & McCulloch, W. S. (1947). How we know universals: the perception of auditory and visual forms. *Bulletin of Mathematical Biophysics, 9*, 127–147.

Poggio, T. & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343*, 263–266.

Quinlan, P. T. (1995). Evidence for the use of scene-based frames of reference in two-dimensional shape recognition. *Spatial Vision, 9*, 101–125.

Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.

Swaddle, J. P. & Cuthill, I. C. (1994). Preference for symmetric males by female zebra finches. *Nature, 367*, 165–166.

Tarr, M. & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame?. *Psychological Science, 1*, 253–256.

Tinbergen, N. (1951). *The study of instinct*. Oxford: Clarendon Press.

Troji, N. F. & Bülthoff, H. H. (1996a). Face recognition under varying pose: The role of texture and shape. *Vision Research, 36*, 1761–1771.

Troji, N. F. & Bülthoff, H. H. (1996b). What is the basis for good performance to symmetric views of faces? *Investigative Ophthalmology & Visual Science, 37*, S194.

Tyler, C. W. (Ed.) (1994). The perception of symmetry. Part I: Theoretical aspects [Special Issue]. *Spatial Vision 8(4).*

Tyler, C. W. (Ed.) (1995). The perception of symmetry. Part II: Empirical aspects [Special Issue]. *Spatial Vision 9(1).*

Ullman, S. (1989). Aligning pictorial descriptions: an approach to object recognition. *Cognition, 32*, 193–254.

Vetter, T. & Poggio, T. (1994). Symmetric 3D objects are an easy case for 2D object recognition. *Spatial Vision, 8*, 443–453.

Vetter, T., Poggio, T. & Bülthoff, H. H. (1994). The importance of symmetry and virtual views in three-dimensional object recognition. *Current Biology, 4*, 18–23.

Wagemans, J. (1995). Detection of visual symmetries. *Spatial Vision, 9*, 9–32.

Watson, P.L, Thornhill, R. (1994). Fluctuating asymmetry and sexual selection. *Trends in Ecology & Evolution, 9*, 21–25.

Acknowledgements—We would like to thank Jeffrey Liter, Michael Tarr and Martin Giurfa for their helpful comments on earlier versions of this paper. We also appreciate the very constructive comments of two anonymous reviewers.