Carnegie Mellon University
Research Showcase @ CMU

Department of Psychology | Dietrich College of Humanities and Social Sciences

12-1-1998

Illumination effects in face recognition

W L. Braje

D Kersten

M J. Tarr
michaeltarr@cmu.edu

N F. Troje

Follow this and additional works at: http://repository.cmu.edu/psychology

Recommended Citation
.

This Article is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Psychology by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Braje, W.L., Kersten, D., Tarr, M.J. and Troje, N.F.
Illumination effects in face recognition. (1998)
Psychobiology. 26 (4), 371-380.

Illumination Effects in Face Recognition

Wendy L. Braje[1], Daniel Kersten[2], Michael J. Tarr[3],
& Nikolaus F. Troje[4]

Psychobiology
In press
August 1998
Running head: Face recognition

[1]St. Cloud State University, St. Cloud, MN 56301
[2]University of Minnesota, Minneapolis, MN 55455
[3]Brown University, Providence, RI 02912
[4]Queen's University, Kingston, Ontario, Canada, K7L3N6

Please address correspondence to:

Wendy L. Braje
Department of Psychology, 304 Whitney House
St. Cloud State University
St. Cloud, MN, 56301
(320) 255-4070
wbraje@stcloudstate.edu

ABSTRACT

How do observers recognize faces despite dramatic image variations that arise from changes in illumination? This paper examines 1) whether face recognition is sensitive to illumination direction, and 2) whether cast shadows improve performance by providing information about illumination, or hinder performance by introducing spurious edges. In Experiment 1, observers judged whether 2 sequentially-presented faces, illuminated from the same or different directions, were the same or different individuals. Cast shadows were present for half of the observers. Performance was impaired by a change in the illumination direction and by the presence of shadows. In Experiment 2, observers learned to name 8 faces under one illumination direction (left/right) and one cast-shadow condition (present/absent); they were later tested under novel illumination and shadow conditions. Performance declined for unfamiliar illumination directions, but not for unfamiliar shadow conditions. The finding that face recognition is illumination dependent is consistent with the use of image-based representations. The results indicate that face recognition processes are sensitive to either the direction of lighting or the resultant pattern of shading, and that cast shadows can hinder recognition, possibly by masking informative features or leading to spurious contours.

INTRODUCTION

Illumination variation has enormously complex effects on the image of an object. In the image of a familiar face, changing the direction of illumination leads to shifts in the location and shape of shadows, changes in highlights, and reversal of contrast gradients. Yet every-day experience shows that we are remarkably good at recognizing faces despite such variations in lighting. Here we examine how humans recognize faces, given image variations caused by changes in lighting direction and by cast shadows. One issue is whether faces are represented in an illumination-invariant or illumination-dependent manner. A second issue is whether cast shadows improve face recognition by providing information about surface shape and illumination direction, or hinder performance by introducing spurious edges that must be discounted prior to recognition. The influences of illumination direction and cast shadows are examined using both short-term and long-term memory paradigms.

The large image variations that result from changing the illumination direction have been demonstrated by Adini, Moses, and Ullman (1995). They compared images of several faces rendered with the same or different lighting direction. Several representations of these images were considered: gray-scale images, images filtered with Gabor functions, edge maps, and 1st and 2nd derivatives of gray-scale images. For all of these representations, they found that varying the illumination direction resulted in larger image differences than did varying the identity of the face.

How can we recognize faces given these dramatic image variations that arise from changes in illumination? One class of models proposes that, in the early stages of processing, the visual system extracts features that are invariant over changes in illumination (e.g. edges defined by surface material, orientation discontinuities, or occlusion), and discounts spurious features (e.g. shadows and specularities) (Biederman 1987; Biederman & Ju, 1988; Marr & Nishihara, 1978).

These *edge-based* models presume that the early image representation is much like an artist's line-drawing, with lines marking key surface shapes and material features.

According to such models, recognition should be unaffected by changes in lighting conditions. Consistent with this view, Moses, Ullman, and Edelman (1996) demonstrated that observers were highly accurate at naming pictures of faces under novel illumination conditions. Additionally, neurophysiological studies have shown that some neurons responsive to faces are invariant to changes in the viewpoint, position, spatial frequency, and/or size of the object (Ito *et al.*, 1995; Rolls, 1992; Rolls, Baylis, & Leonard, 1985; Tovee, Rolls, & Azzopardi 1994). It might be expected that neurons of this type would also be insensitive to changes in illumination conditions. From a computational perspective, however, no algorithm has yet been devised that can produce a clean line-drawing of natural objects and smooth objects (such as faces) under all the image variations that occur naturally (cf. Adini *et al.*, 1995).

Image-based models, on the other hand, propose that object representations are more closely tied to the 2-D image (Bülthoff, Edelman, & Tarr, 1995; Gauthier & Tarr, 1996; Poggio & Edelman, 1990). According to these models, introducing a new illumination direction results in a large change in the stimulus representation, which should lead to an impairment of recognition. Cavanagh (1991) has pointed out that cast shadows are difficult to identify using early mechanisms. This suggests that it is easier to encode illumination effects than to discount them. Similarly, Adini *et al.* (1995) and Ullman (1996, p. 324) have argued that illumination must be processed using higher-level mechanisms. Moreover, illumination information may be useful for computing 3-D shape and identifying shadows (Tarr, Kersten, & Bülthoff, 1998). Thus, illumination information might be retained in face representations because it is too difficult to discount at an early level and/or because it is useful at higher levels.

Support for image-based models comes from recent studies showing that changing the direction of illumination can influence recognition of geometric objects (Tarr *et al.*, 1998) and facial surfaces (i.e. 3-D renderings of faces without color or texture) (Hill & Bruce, 1996; Troje & Bülthoff, 1998). It has also been shown that faces are more difficult to recognize when presented in photographic negative, implying that edge information, without contrast polarity information, is not sufficient to account for face recognition (Hayes, Morrone, & Burr, 1986). Other studies show that the direction of illumination can influence 3-D shape perception (Belhumeur, 1997; Berbaum, Bever, & Chung, 1983; Christou & Koenderink, 1996). Furthermore, although Moses *et al.* (1996) found high accuracy for naming pictures of faces under novel illumination conditions, their observers did show a cost in reaction time.

Neurophysiological studies also suggest that some effects of illumination may be encoded. Warrington (1982) described patients with right posterior lesions who had difficulty recognizing objects under different illumination conditions. Weiskrantz (1990) discussed similar findings in monkeys with lesions in inferotemporal and prestriate areas. Hietanen *et al.* (1992) discovered face-selective cells in the temporal cortex of the macaque that were sensitive to changes in illumination. Such findings suggest that information about illumination (or image effects resulting from illumination) is not discarded, but is instead retained in the object representation.

One consequence of changing illumination conditions is a change in the characteristics (e.g. shape and location) of shadows. How do shadows affect face recognition? Edge-based models of object recognition propose that the visual system discounts spurious features such as shadows. In this scheme, shadows should not affect recognition. Support for this view was found by Braje, Legge, and Kersten (submitted). Their observers named pictures of familiar natural objects (fruits and vegetables), shown with or without shadows. They found that shadows had no effect on observers' accuracy or reaction time.

According to image-based models, however, object representations retain information present in the original image, including shadows. Cavanagh (1991) proposed that early processing of an image involves a crude match of the image to a memory representation, in which all image contours (including shadow contours) are used. Only once a candidate object is selected are the contours labeled as belonging to objects or shadows. If shadows are encoded as part of the object representation, they can be problematic for recognition in that they introduce spurious luminance edges that can be confused with object contours. Consistent with this prediction, Warrington (1982) demonstrated that patients with right posterior lesions had difficulty recognizing photographs of common objects containing shadows, and Moore and Cavanagh (1998) showed that two-tone images of novel objects with shadows are difficult to recognize. Alternatively, encoding shadows might improve recognition by providing useful information about object shape (e.g. surface orientation and curvature) and about illumination conditions in a scene (e.g. light source direction). Tarr *et al*. (1998) demonstrated that cast shadows can improve recognition of novel geometric objects, suggesting that shadows provide useful information about shape or lighting conditions.

Shadows can be classified into two types: attached shadows and cast shadows. Each type places different requirements on a recognition system attempting to discount shadows at any stage of processing. An *attached shadow* occurs when a surface gradually turns away from the lighting direction. If a Lambertian shading model is assumed, then the intensity at a given pixel in an attached shadow depends only on the local surface orientation with respect to the illumination direction. Thus any algorithm (artificial or biological) attempting to discount attached shadows can work locally. A *cast shadow* occurs when an object is interposed between a light source and a surface, blocking the illumination from reaching the surface (Beck 1972). Cast shadows may be particularly difficult to discount, in that they are influenced both by local surface characteristics and by surfaces more distant to the shadowed area.

Face recognition may be particularly susceptible to the influence of shadows. Although Braje *et al*. (1998) found no effect of shadows on the ability to name familiar natural objects, familiar stimuli may be more resistant to noise in general and to shadows in particular. Faces, on the other hand, although familiar as a class of objects, are not necessarily familiar on an individual basis. They may therefore be less resistant to the influence of shadows. A second reason shadows should affect face recognition is that rather fine discrimination is required to distinguish between different faces. Different faces are fairly similar in their global shape, hue, and texture. Because of the small differences between faces, recognizing a particular face requires a detailed analysis, and may also rely more heavily on correct labeling of contours. Under such conditions, shadows should have a larger impact on recognition.

The present experiments investigate whether face representations are illumination-dependent, and how cast shadows influence face recognition. The first experiment uses a same/different matching task to examine whether face representations in short-term memory retain illumination and shadow information. The second experiment extends the findings to representations in long-term memory by using a naming task. The key issues are whether recognition performance is influenced by 1) changing the direction of illumination, and 2) the presence of cast shadows.

EXPERIMENT 1: MATCHING

In this experiment, observers viewed two sequentially-presented faces, and decided whether they were the same or different people. The two faces were illuminated from the same or different directions on each trial. If face representations are closely tied to the 2-D image, then performance

should decline when the illumination changes between the two images. However, if a more invariant encoding is used, illumination should have no effect on performance.

Cast shadows were present for half of the observers and absent for the other half. If cast shadows are retained in the face representation, then they should influence performance in one of two ways. Their presence should lead to an overall decline in performance if they introduce confusing contours or mask important features. If, however, cast shadows provide useful information regarding 3-D shape or illumination, performance should improve when they are present.

Methods

Observers

Thirty-two undergraduate psychology students (ages 17 to 34) at the University of Minnesota participated in the experiment for class credit. All had normal or corrected-to-normal visual acuity (Snellen acuity of 20/20 or better) and gave informed consent. The observers were not familiar with the people whose faces were used as stimuli.

Stimuli and Apparatus

Face images were obtained from 3-D models of 80 real human heads (Troje & Bülthoff, 1996). The colored texture and 3-D geometry of the heads were digitized using a Cyberware™ 3-D laser scanner. The scanner samples a regular grid of surface points in cylindrical coordinates. Each sample contains the spatial coordinates of the surface point (azimuth, height, and radius), as well as

its color (RGB) values. The distance between neighboring points is 0.8° (roughly 1 mm) in the azimutal direction, 0.615 mm in the vertical direction, and 0.016 mm along the radius. There were 40 males and 40 females, all Caucasian people between about 20 and 40 years of age. They had no spectacles or facial hair, and they had neutral facial expressions. The hair and back of the head were removed from the 3-D models before the images were created.

The face images were rendered in color in orthographic projection using Wavefront's "The Advanced Visualizer 4.1" (Santa Barbara, CA) on a Silicon Graphics workstation. Each face was rendered from two different viewpoints (7° and 11° with respect to the frontal view), and in two sizes (7.9° by 9.5° and 8.5° by 10.3°), for reasons discussed in the *Procedure*. Faces were illuminated by a point source located 130 cm from the face. The source was positioned 45° above and 45° to the right or left of the viewing axis. The faces were rendered with ray-tracing (producing cast shadows) and without ray-tracing (no cast shadows). A small amount of ambient light was also present, such that the shadowed regions were not entirely dark. **Figure 1** illustrates the different renderings. Note that an image with no cast shadows does contain attached shadows, and that all cast shadows are intrinsic (i.e. cast by parts of the face onto the face itself).

The experiment was run on an Apple Macintosh Quadra 950 using RSVP software (Williams & Tarr). The faces were presented on a black background (<1 cd/m^2) on a 640 x 480-pixel Apple 17-inch color monitor. They were roughly 150 pixels horizontally by 215 pixels vertically, and were viewed from a distance of 45 cm. The luminance of each face ranged from less than 1 cd/m^2 to around 40 cd/m^2.

A 256 by 256-pixel mask was created using Adobe Photoshop. The mask consisted of a collage of scrambled face features—eyes, noses, mouths, and ears—taken from the different renderings of the 80 faces. The luminance of the mask was 10 cd/m^2.

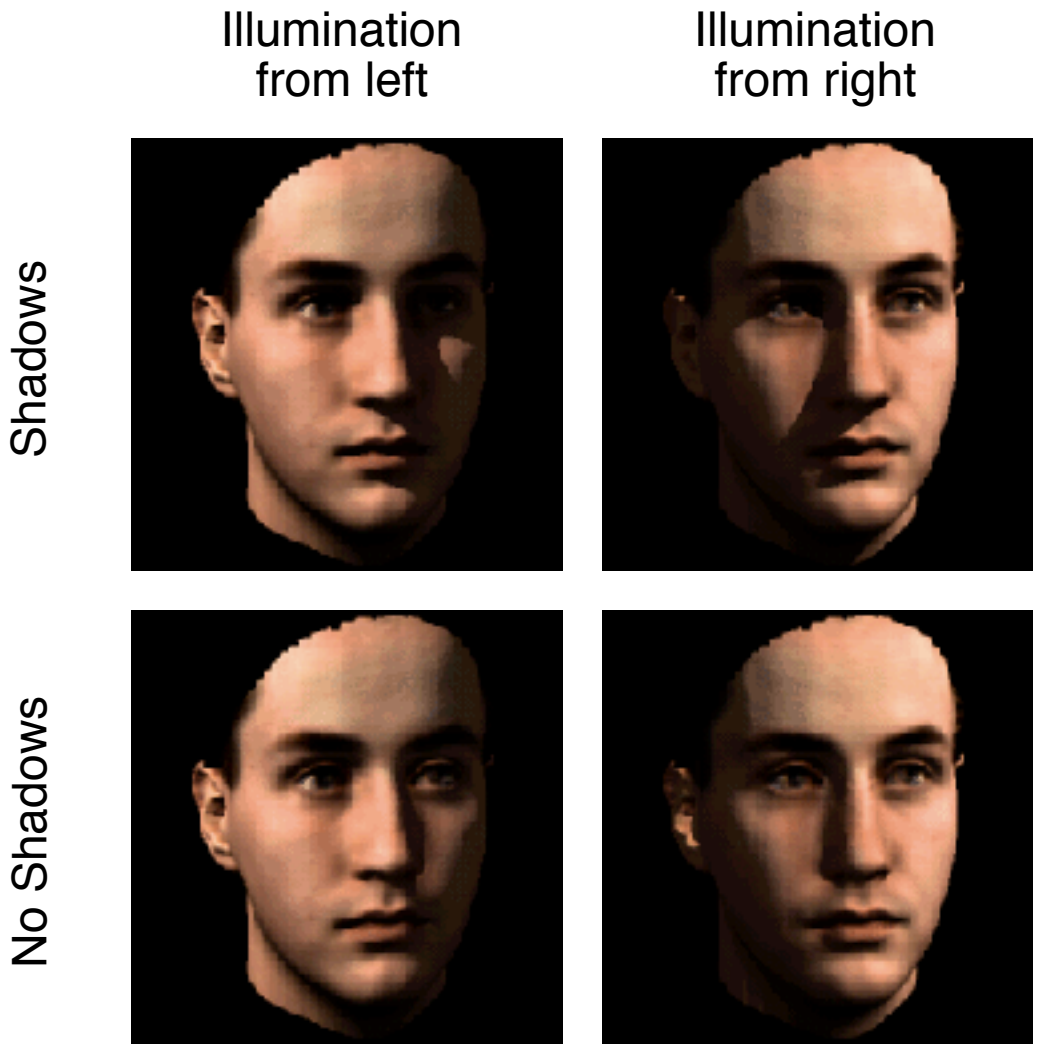

Figure 1. A face in the four rendering conditions (left/right illumination, with/without cast shadows). The viewpoint is 11° with respect to the frontal view.

Procedure

A sequential matching paradigm was used (see **Figure 2**). On each trial, a fixation cross was presented for 500 msec, followed by a face for 200 msec, the mask for 750 msec, a second face for 100 msec, and the mask again for 500 msec.

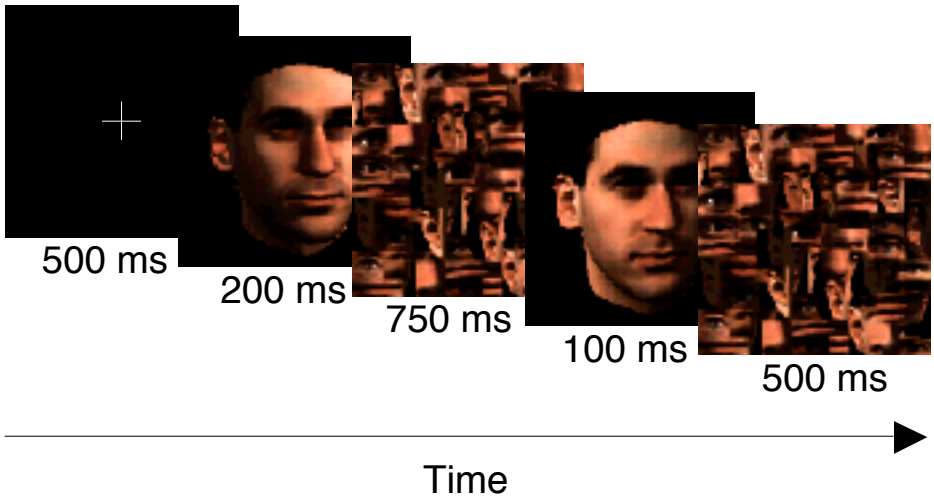

Figure 2. Sequential matching paradigm used in Experiment 1. The observers' task was to decide whether the two faces were the same or different people.

The observer's task was to decide whether the two faces were the same or different people, and to respond by pressing one of two keys on a keyboard. Feedback was not provided. The durations were chosen based on an experiment by Ellis and Allport (1986), who showed that an ISI of 750 msec was long enough to prevent observers from simply using an iconic representation of the first face to perform the match. With the timing used here, observers are likely to encode the first image in short-term memory for subsequent match to the second image (Tarr *et al.*, 1998). The presentation durations were also chosen to prevent observers from obtaining ceiling performance (i.e. 100% accuracy).

Response time was measured as the time between the onset of the second face and the observer's key-press. Observers were allowed up to 5 seconds to respond. Sensitivity (d')was calculated using the z-scores for the correct responses on *same-face* trials (hits) and the incorrect responses on *different-face* trials (false alarms)[1]: $d' = z_{hit} - z_{fa}$.

On each trial, the two faces could be illuminated from either the same or different directions. Half of the observers viewed faces with cast shadows, the other half viewed faces without cast shadows. The presence or absence of cast shadows was run as a between-groups variable in order to prevent observers from relying on a salient shadow when making a matching judgment.

Each observer completed a 20-trial practice block, followed by 4 blocks of 120 experimental trials. There were an equal number of left-illuminated and right-illuminated faces in each block, and an equal number of *same-face* trials and *different-face* trials in each block. Over the 480 total trials, each of the 80 possible *same-face* pairs was presented 3 times, and a random selection of the 6,320 possible *different-face* pairs was presented. Each *different-face* pair was presented only once to a given observer. Both same-sex and different-sex pairs were used on *different-face* trials. An ANOVA run using just the trials containing same-sex pairs revealed the same results as found when all trials (both same- and different-sex) were used. Thus all trials were included in the analyses reported below.

For each presentation of each image, the viewpoint and size were randomly chosen from among those described in the *Stimuli* section. This prevented observers from simply matching images for size, silhouette, or local image features. Non-frontal views were used in order to avoid symmetrical image pairs, which would occur in a frontal view when the light source was moved from one side to the other. Symmetrical images are perceptually very similar (Troje & Bülthoff, 1998) and would produce an accidental situation in which changing the illuminant's position would not impair recognition performance.

Results

Performance declined when there was a change in illumination direction. **Figure 3** shows that, in *same-face* trials, response time was 44 msec slower when there was a change in illumination direction than when there was no change. In *different-face* trials (data not shown), changing the illumination direction had no effect on response time. An ANOVA (run on the median response times for the trials in which a correct response was made) confirmed a main effect of illumination direction ($F(1,30)=18.24$, $p<.001$) and an interaction between illumination direction and face (same *vs*. different) ($F(1,30)=17.98$, $p<.001$). A Tukey HSD test ($\alpha=.05$) confirmed that the illumination effect occurred only for *same-face* trials. Sensitivity was also lower when there was a change in illumination direction ($d' = 2.3$) than when there was no change ($d' = 2.6$) ($F(1,30)=17.81$, $p<.001$). Observers scored 85% correct when there was a change in illumination, and 92% when there was no change.

The presence of cast shadows impaired response time ($F(1,30)=5.54$, $p<.05$), but not sensitivity ($F(1,30)=1.81$, $p=.189$). Response time was 127 msec slower when cast shadows were present. A Tukey HSD test ($\alpha=.01$) confirmed that observers were always slower when cast shadows were present, for all illumination and face conditions.

This decline in performance in images containing cast shadows is not simply a result of lower mean luminance or higher contrast in the cast shadow images. There was no correlation between response time and mean pixel value of each face ($R^2=0.002$), or between response time and RMS contrast ($R^2=0.01$)[2].

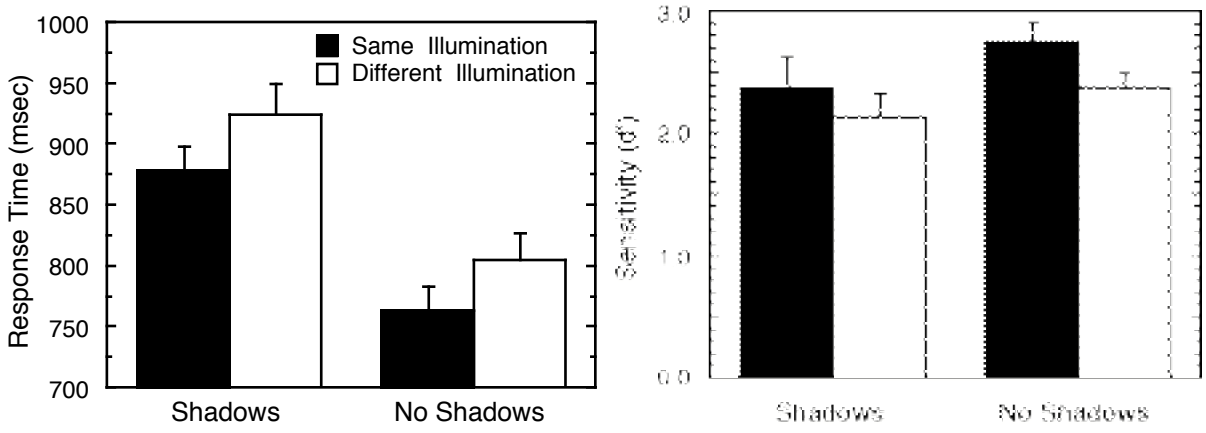

Figure 3. Response time and sensitivity for the trials in which the faces were the same person. Response times shown are the average of the median response times for trials in which a correct response was made. Error bars show standard errors.

Finally, the effect of illumination was not influenced by the presence of cast shadows—there was no significant interaction between illumination (same *vs*. different) and cast shadows (present *vs*. absent), for either response time ($F(1,30)=3.32$, $p=.078$) or sensitivity ($F(1,30)=1.62$, $p=.213$).

Discussion

Performance declined when the illumination changed and when cast shadows were present. This suggests that short-term face representations, lasting at least several hundred milliseconds, preserve image information such as illumination and cast shadows, and it supports image-based models of recognition. The findings are consistent with the results of Tarr *et al*. (1998) for recognition of novel geometric objects. In a matching task like the one used here, they found that recognition was 21 msec faster when there was no change in illumination direction than when there

was a change. In the present study, the difference was 44 msec, even larger than that found by Tarr *et al.*

One possible reason that changing the illumination impairs performance is that doing so introduces uncertainty into the light position. If the location of the illuminant is ambiguous, it could be more difficult for observers to discount the illuminant. Cast shadows can provide information about the location of the light source. Therefore, when cast shadows are present, changing the illumination direction might be expected to have a smaller effect on recognition. This interaction was not found—the effect of changing the illumination direction was the same whether cast shadows were present or absent. If the lighting was indeed ambiguous, then the observers did not make use of cast shadows to disambiguate it.

The decline in performance in images containing cast shadows could be explained in two ways. One possibility is that time is required for locating and discounting cast shadows (equivalently, higher-level processing may be needed). However, the finding that changing the illumination direction impaired performance suggests that image information is preserved, and thus it seems unlikely that cast shadows in the image are discounted. A second possibility is that cast shadows mask informative features, forcing the visual system to rely on a smaller (less informative) set of features for recognition. This possibility requires further investigation.

In summary, the results are consistent with the assumption that image information is preserved in short-term face representations. The second experiment examines whether these results extend to longer-term representations.

EXPERIMENT 2: NAMING

In Experiment 1, observers needed to hold the first face in memory for only about one second in order to make a judgment. It is possible that these short-term representations retain image information, while longer-term representations are more invariant. Experiment 2 therefore examined the effects of cast shadows and illumination on these long-term representations of faces.

It may also be argued that the matching task used in Experiment 1 is not an accurate measure of recognition ability. Observers were asked to detect differences between the stimuli, rather than to explicitly recognize them. In this case, it is possible that illumination-dependent performance arose from within a system that looks for image differences and is sensitive to any changes between stimuli, rather than from within the face-recognition system itself. Thus, a second purpose of Experiment 2 was to examine illumination effects using a more natural face-recognition task.

Observers were trained for two days to recognize eight faces under one illumination direction and one cast-shadow condition. The faces were later presented under new illumination and cast-shadow conditions for observers to recognize. If long-term representations preserve image information, the results should be similar to those in Experiment 1—performance should be best when the faces are presented under familiar illumination and cast-shadow conditions. Additionally, the overall impairment caused by cast shadows in Experiment 1 might disappear in Experiment 2 — observers should have sufficient time to either build representations based on visible features (if cast shadows mask informative features), or to use higher-level mechanisms to process cast shadows.

Methods

Observers

Twenty University of Minnesota undergraduate psychology students (ages 18 to 38) participated in the experiment for class credit. None had participated in Experiment 1. All were Caucasian, had normal or corrected-to-normal visual acuity (Snellen acuity of 20/20 or better), and gave informed consent. The observers were not familiar with the people whose faces were used as stimuli.

Stimuli and Apparatus

The stimuli consisted of 8 full-color faces (4 males and 4 females) presented on a black background. The 8 faces were randomly selected from the 80 used in Experiment 1. Only the 11° view and the larger-sized images were used. The lighting and cast-shadow conditions were identical to those used in Experiment 1.

The experiment was run on a Power Macintosh 7100 using RSVP software (Williams & Tarr). The images were presented on a black background (<1 cd/m^2) on a 640 by 480-pixel Apple 17-inch color monitor.

Voice-activated response time was recorded using the CMU Button Box (accurate within 1 msec). The observers' verbal responses (the names of the faces) were compiled by the experimenter using an IBM XT personal computer.

Procedure

The experiment was conducted over two consecutive days for each observer. There were three phases to the experiment: *training*, *practice*, and *testing*.

During *training*, observers learned to associate a name with each face. For a given observer, all faces were illuminated from the same direction (right or left) and were presented in the same cast-shadow condition (present or absent). On each trial, a face was presented for 2 seconds with its name written below it, and observers were instructed to simply learn the name of each face; no response was required. There were 3 blocks of 16 trials (each face was presented twice per block), and the order of trials was randomized for each observer.

The training was followed by a *practice* phase, in which observers practiced naming the faces they had just learned. The faces were shown in the familiar illumination and shadow conditions (i.e. the same as those seen during training). Each trial consisted of a fixation cross for 500 msec, followed by a warning tone, and then one of the 8 faces. The purpose of the tone was to signal the beginning of a trial to the experimenter, who could not see the images being presented. The task was to name the face aloud as quickly and accurately as possible. The face remained on the screen until the observer responded or 10 seconds had passed. After the observer responded, the correct name and the observer's response time were displayed for 500 msec. On the first day of the experiment, observers completed 4 blocks of 40 practice trials. Each of the 8 faces was presented 5 times per block. On the second day of the experiment, observers completed 2 additional 40-trial blocks of practice. Observers were required to reach at least 90% accuracy by the end of the final practice session[3].

The final phase was the *testing* phase. The procedure was the same as for the practice trials, except that feedback was not provided. The stimuli consisted of the 8 faces in each of the 4 possible rendering conditions: left/right illumination and with/without cast shadows. Eight of these images

were familiar (those the observers had trained and practiced on), and the other 24 were unfamiliar (the same 8 faces but rendered under the other 3 illumination/cast-shadow conditions). Observers completed 8 blocks of 32 trials, with each of the 32 images presented once per block. Response time was measured as the time between the onset of the face and the onset of the observer's verbal response[4].

Results

Accuracy was high (>90%) in all conditions, confirming that observers had learned the faces well. Except for an improvement across blocks (see below), no main effects or interactions were significant for the accuracy data. The response times for only the testing phase will be considered here. Response times are compared for familiar *vs*. unfamiliar illumination directions and cast-shadow conditions.

Recognition was impaired when faces were displayed under unfamiliar illumination. Median response times (see **Figure 4**) were 134 msec slower when the illumination at testing was different from the illumination during training ($F(1,15)=34.53$, $p<.001$)[5]. This difference (a 7% change) is of similar magnitude to that in Experiment 1 (a 5% change).

Cast shadows did not influence performance. Response times were equally fast for recognizing faces under familiar and unfamiliar cast-shadow conditions ($F(1,15)=.90$, $p=.36$ for the interaction between trained and tested cast-shadow conditions). Observers trained with no cast shadows tended to respond more quickly overall during the test phase, but this effect did not reach significance ($F(1,15)=2.69$, $p=.12$). The presence of cast shadows during the test phase did not affect response time ($F(1,15)=0.21$, $p=.65$).

During the testing phase, response time decreased across blocks (F(7,105)=5.51, *p*<.001), from an average of 1329 msec in the first block to 885 msec in the final block. Accuracy also improved (F(7, 105)=2.76, *p*<.02) from 92% to 97% correct. However, the effects of illumination and cast shadows did not change across blocks: none of the cast-shadow or illumination variables interacted with block number.

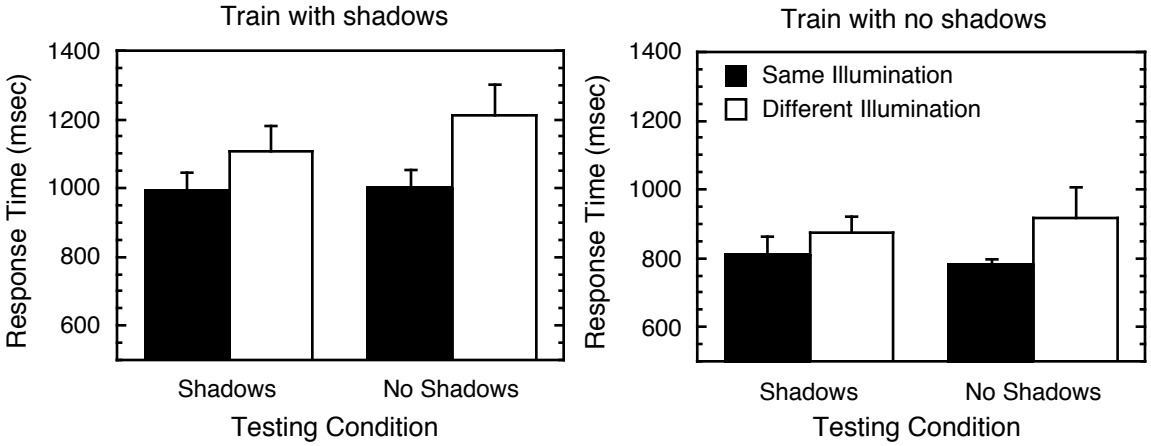

Figure 4. Response times during testing for naming faces under familiar or unfamiliar illumination and cast-shadow conditions. Response times shown are averages of the median response times for trials in which a correct response was made. Error bars show standard errors.

Discussion

Experiment 2 partially replicated the results of Experiment 1. The effect of illumination direction was still present when a naming task was used, demonstrating that illumination information is retained even in long-term representations. This finding also suggests that the

illumination effects observed in Experiment 1 arose from within a face-recognition system, rather than a system that simply detects image differences. The results of Experiment 2, like those of Experiment 1, are thus consistent with an image-based representation of illumination.

It should be noted that naming tasks have sometimes failed to reveal invariance. For example, Biederman and colleagues have found invariance for translation, reflection, and size using a priming paradigm (Biederman & Cooper, 1991, 1992; Cooper, Biederman, & Hummel, 1992; Fiser & Biederman, 1995). The present results are thus significant in that they demonstrate the existence of image-based processing using a naming task.

Unlike Experiment 1, Experiment 2 revealed no effect of cast shadows on performance. Observers performed equally well in the testing phase whether they learned the faces with cast shadows present or absent. This demonstrates that a certain amount of invariance with respect to cast shadows is acquired as observers learn to recognize faces. Observers may have had sufficient time to use higher-level mechanisms for dealing with cast shadows. Alternatively, if cast shadows were masking important features, observers may have learned to recognize the faces using other features.

GENERAL DISCUSSION

Face recognition was found to be sensitive to the presence of cast shadows and to changes in illumination. Observers were slower and less accurate at matching and naming faces when there was a change in illumination direction. There was also a cost in speed and accuracy for matching (but not naming) faces that contained cast shadows.

The finding that illumination changes impair recognition is in agreement with the recent findings of Tarr *et al*. (1998), and it is consistent with image-based models of face recognition. The

findings do contradict some neurophysiological studies, which show that some face-selective cells respond over a wide range of viewpoints, positions, spatial frequencies, and sizes (Ito *et al.*, 1995; Rolls, 1992; Rolls, Baylis, & Leonard, 1985; Tovee, Rolls, & Azzopardi 1994). On the other hand, our findings are in agreement with studies that have shown face-selective cells to be sensitive to the viewpoint of faces (Desimone *et al.*, 1984; Perrett *et al.*, 1991) and, more importantly, to changes in illumination (Hietanen *et al.*, 1992).

Why should the visual system encode illumination information? For one thing, knowledge of illumination conditions is useful for computing 3-D shape. This may be particularly important in face recognition, a task for which the stimuli are geometrically similar. Several studies have demonstrated that viewpoint dependence is stronger when objects are most similar (Edelman, 1995; Tarr & Bülthoff, 1995), and it is likely that illumination dependence would behave in the same manner. Tarr *et al.* (1998) have argued that discounting illumination effects increases the similarity between objects, thus making them harder to distinguish. Preserving illumination information may therefore be useful for discriminating between different objects.

Cast shadows were found to impair performance on the matching task (Experiment 1). This is consistent with an image-based model of face recognition, and it suggests that shadows introduce spurious contours that are confused with surface contours. It is also possible that some of the shadows were cast onto features important for recognition, thus making such features difficult to use. A third possibility is that the shadows that are cast onto a particular surface disrupt shape-from-shading processes, e.g. by making a curved surface (which would otherwise have a shading gradient) uniformly dark. Further research is needed to distinguish between these possibilities.

The finding that cast shadows impaired performance contradicts Tarr *et al.*'s (1998) results, which showed that cast shadows improved recognition of geometric objects. One reason for this discrepancy may be that the stimuli used by Tarr *et al.* were relatively simple compared to faces, and

cast shadows may be more likely to obscure details in the more complex face images. Tarr *et al.* have also argued that the 3-D shape of faces is well-known, and so cast shadows are unlikely to offer much additional information. This argument is supported by the finding that cast shadows did not influence performance on the face-naming task.

In the naming task (Experiment 2), observers were given plenty of opportunity to "learn" the faces. Even if cast shadows normally cause problems in building early face representations, observers had ample time in the naming experiment to either "solve" the problem of labeling shadow edges, or resort to an alternative recognition strategy that is more invariant with respect to shadows.

Further research is necessary in order to test the generality of these results. In particular, the present studies used a slightly oblique viewpoint (7°–11°) for the faces, which may not provide adequate information for generalization across illumination. Troje and Bülthoff (1996) showed that recognition was more viewpoint invariant for faces initially learned from a viewpoint between 22.5° and 45° than those learned from a frontal view (0°). Thus it may be the case that the viewpoints used in the present studies (7° and 11°) were not optimal for face encoding. However, Tarr *et al.* (1998) found that illumination dependency in the recognition of geometric objects was not influenced by the familiarity of the viewpoint in which the object was presented. This suggests that the current findings would not change if different viewpoints were used.

Recognition models that rely on edge detection and/or early spatial filtering cannot account for the visual system's ability to allow for illumination variation. One alternative approach is to compute the direction of illumination, and then use this to determine 3-D shape in face images (Kersten *et al.* 1987; Pentland, 1990). However, it is unlikely that human vision uses a global estimate of light source direction to compute shape-from-shading or depth-from-shadows (see Tarr *et al.* for a discussion). Another approach comes from "appearance-based" models. Here, low-

dimensional representations that do not explicitly distinguish shape, material, and lighting are extracted (e.g. through principal components analysis). These models do not explicitly represent illumination, but instead store illumination-specific models (Murase & Nayar, 1995). It is unclear, however, how well these models can deal with novel combinations of illumination and material changes. A third approach is to use a linear model of illumination to derive illumination basis images for a specific view of an object (Epstein, Hallinan, & Yuille, 1995; Hallinan, 1994, 1995; Shashua, 1992, 1996). These "eigenimages" define the space over which the model can generalize to new illumination conditions (Belhumeur & Kriegman, 1996). In this model, cast shadows can be treated as residuals, in keeping with Cavanagh's suggestion (1991) that cast shadows are identified by a secondary top-down process.

In summary, the results demonstrated that both short- and long-term face representations are closely tied to information contained in the original image. Face recognition was found to be illumination-dependent, indicating that face-recognition processes are sensitive to either the direction of lighting or the resultant pattern of shading. Cast shadows impaired recognition on a sequential matching task, perhaps because they introduce confusing contours. On a long-term task, however, observers are able to overcome the harmful effects of shadows, indicating that some degree of invariance can be achieved.

FOOTNOTES

[1]Two observers had a hit rate of 1.0, and one observer had a false alarm rate of 0.0. For these observers, *d'* scores were calculated from their percent correct scores, using the equation given in Macmillan & Creelman (1991) for same-different tasks.

[2]Although the faces were very similar in mean luminance and RMS contrast, they were not identical.

[3] Three observers failed to reach this criterion, even after given an extra practice block on the second day of the experiment. The data from these observers were excluded from the analysis. Of the remaining 17 observers, there were 5 in the cast-shadow/right-illumination condition and 4 in each of the other 3 conditions.

[4] Trials in which the timer was inccorectly triggered were excluded from the analysis. Such trials were rare.

[5] An ANOVA was performed on the median reaction times raised to the -3/2 power. This transformation was necessary in order to satisfy the assumption of normality.

REFERENCES

Adini, Y., Moses, Y., & Ullman, S. (1995). *Face recognition: the problem of compensating for changes in illumination direction* (Report No. CS93-21). The Weizmann Institute of Science.

Beck, J. (1972). *Surface Color Perception*. Ithaca, NY: Cornell University Press.

Belhumeur, P. N. (1997). The bas-relief ambiguity. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1060-1066.

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

Berbaum, K., Bever, T., & Chung, C. S. (1983). Light source position in the perception of object shape. *Perception*, *12*, 411-416.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115-147.

Biederman, I. & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, *20*, 38-64.

Biederman, I. & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance, 18(1)*, 121-133.

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

Braje, W. L., Legge, G. E., & Kersten, D. (1998). Invariant recognition of natural objects in the presence of shadows. Submitted to *Perception*.

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, *5*(3), 247-260.

Cavanagh, P. (1991). What's up in top-down processing? In A. Gorea (Eds.), *Representations of Vision: Trends and tacit assumptions in vision research* (pp. 295-304).

Christou, C. G. & Koenderink, J. J. (1996). Shape from shading and shape constancy under variations in illumination. *Investigative Ophthalmology and Visual Science, Supplement*, *37* (3), S934.

Cooper, E. E., Biederman, I., & Hummel, J. E. (1992). Metric invariance in object recognition: a review and further evidence. *Canadian Journal of Psychology, 46(2)*, 191-214.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience, 4*, 2051-2062.

Edelman, S. (1995). Class similarity and viewpoint invariance in the recognition of 3D objects. *Biological Cybernetics, 72*, 207-220.

Ellis, R. & Allport, D. A. (1986). Multiple levels of representation for visual objects: a behavioural study. In A. G. Cohn & J. R. Thomas (Eds.), *Artificial Intelligence and its Applications* (pp. 245-257). John Wiley & Sons Ltd.

Epstein, R., Hallinan, P. W., & Yuille, A. L. (1995). 5 ± eigenimages suffice: An empirical investigation of low-dimensional lighting models. In *IEEE Workshop on Physics-based modeling in Computer Vision* (p. 108-116). Boston, MA.

Fiser, J. & Biederman, I. (1995). Size invariance in visual object priming of gray-scale images. *Perception, 24*, 741-748.

Gauthier, I. & Tarr, M. J. (1997). Becoming a "greeble" expert: exploring mechanisms for face recognition. *Vision research, 37(12)*, 1673-1682.

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

Hallinan, P. W. (1995). *A deformable model for face recognition under arbitrary lighting conditions*. Ph.D. Thesis, Division of Applied Sciences, Harvard University.

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

Hietanen, J. K., Perrett, D. I., Oram, M. W., Benson, P. J., & Dittrich, W. H. (1992). The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex. *Experimental Brain Research*, *89*, 157-71.

Hill, H. & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 986-1004.

Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology, 73(1)*, 218-226.

Kersten, D., O'Toole, A. J., Sereno, M. E., Knill, D. C., & Anderson, J. A. (1987). Associative learning of scene parameters from images. *Applied Optics, 26,* 4999-5006.

Macmillan, A. & Creelman, C. D. (1991). *Detection Theory: A User's Guide*. New York, NY: Cambridge University Press.

Marr, D. & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Philosophical Transactions of the Royal Society of London, B*, *200*, 269-294.

Moore, C. & Cavanagh, P. (1998). Recovery of 3D volume from 2-tone images of novel objects. *Submitted to Cognition (special issue on object recognition).*

Moses, Y., Ullman, S., & Edelman, S. (1996). Generalization to novel images in upright and inverted faces. *Perception*, 25, 443-462.

Murase, H. & Nayar, S. (1995). Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision, 14,* 5-24.

Pentland, A. P. (1990). Linear shape from shading. *International Journal of Computer Vision, 4*, 153-162.

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., & Benson, P. J. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Experimental Brain Research, 86*, 159-173.

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

Rolls, E. T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philosophical Transactions of the Royal Society of London B*, *335*, 11-21.

Rolls, E. T., Baylis, G. C., & Leonard, C. M. (1985). Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus in the monkey. *Vision Research, 25(8)*, 1021-1035.

Shashua, A. (1992). Illumination and View Position in 3D Visual Recognition. In J.E. Moody, S. J. Hanson, & R. P. Lippmann (Eds.), *Advances in neural information processing systems 4* (p. 404-411). CA: Morgan Kaufmann Publishers.

Shashua, A. (1996). On photometric issues in 3D visual recognition from a single 2D image. *International Journal of Computer Vision, 21*, 99-122.

Tarr, M. J., Kersten, D., & Bülthoff, H. H. (1998). Why the visual recognition system might encode the effects of illumination. *Vision Research, 38(15/16)*, 2259-2275.

Tarr, M. J. & Bülthoff, H. H. (1995). Is human object recognition better described by geon-structural-descriptions or by multiple-views? *Journal of Experimental Psychology: Human Perception and Performance, 21(6)*, 1494-1505.

Tovee, M. J., Rolls, E. T., & Azzopardi, P. (1994). Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. *Journal of Neurophysiology*, *72*(3), 1049-1060.

Troje, N. F. & Bülthoff, H. H. (1998) How is bilateral symmetry of human faces used for recognition of novel views? Vision *Research*, 38(1), 79-89.

Troje, N. F. & Bülthoff, H. H. (1996). Face recognition under varying poses: the role of texture and shape. *Vision Research*, *36*(12), 1761-1771.

Ullman, S. (1996). *High-level Vision: Object Recognition and Visual Cognition*. Cambridge, MA: M.I.T. Press.

Warrington, E. K. (1982). Neuropsychological studies of object recognition. *Philosophical Transactions of the Royal Society of London B*, *298*, 15-33.

Weiskrantz, L. (1990). Visual prototypes, memory, and the inferotemporal lobe. In E. I. M. Mishkin (Eds.), *Vision, Memory, and the Temporal Lobe* (pp. 13-28). New York: Elsevier.

Williams, P. & Tarr, M. J. (No date). *RSVP: Experimental control software for MacOS* [Online]. Available: http://psych.umb.edu/rsvp/ [1998, May 27].