Invariant Recognition of Natural Objects in the Presence of Shadows

Wendy L. Braje, Gordon E. Legge, & Daniel Kersten

ABSTRACT

Shadows are frequently present when we recognize natural objects, but it is unclear whether they help or hinder recognition. Shadows could improve recognition by providing information about illumination and 3-D surface shape, or impair recognition by introducing spurious contours that are confused with object boundaries. In three experiments, we explored the effect of shadows on recognition of natural objects. The stimuli were digitized photographs of fruits and vegetables displayed with or without shadows. Experiment 1 evaluated the effects of shadows, color, and image resolution on naming latency and accuracy. Performance was not affected by the presence of shadows, even for grayscale, blurry images, where shadows are difficult to identify. Experiment 2 explored recognition of two-tone images of the same objects. In these images, shadow edges are difficult to distinguish from object and surface edges because all edges are defined by a luminance boundary. Shadows impaired performance, but only in the early trials. Experiment 3 examined whether shadows have a stronger impact when exposure time is limited, allowing little time for processing shadows; no effect of shadows was found. These studies show that recognition of natural objects is highly invariant to the complex luminance patterns caused by shadows.

INTRODUCTION

The human visual system has a remarkable ability to recognize objects. As viewpoint and lighting change in a scene, the retinal image of an object can vary dramatically. One particular effect of changing lighting conditions is a change in the characteristics (e.g. shape and location) of shadows. How is object recognition affected by the presence of shadows? Image-based theories of object recognition propose that object representations retain information present in the original image, including shadows (Bülthoff, Edelman, & Tarr, 1995; Edelman & Bülthoff, 1992; Edelman, 1995; Gauthier & Tarr, 1997; Poggio & Edelman, 1990). Cavanagh (1991) proposed that the 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. Alternatively, encoding shadows might improve recognition by providing useful information about local object shape (e.g. surface orientation and curvature) and/or global scene properties (e.g. light source direction). Other theories of object recognition propose that the visual system extracts invariant features, such as object edges, while discounting spurious features, such as shadows (Biederman & Ju, 1988; Marr & Nishihara, 1978). In this scheme, shadows should not affect recognition, provided they can be easily discounted.

The present studies examine whether shadows improve or impair object recognition. As discussed below, performance on many tasks has been shown to benefit from the presence of

shadows, while other studies conclude that shadows impair performance. However, little research has addressed the specific issue of how shadows influence performance on the task of recognizing objects.

Shadows can be classified into two types (**Figure 1**). An *attached shadow* occurs when a surface turns away from the lighting direction, causing that region to become darker. 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 can be *extrinsic*, i.e. one object casts a shadow onto another; or they can be *intrinsic*, i.e. an object casts a shadow onto itself. All types of shadows tend to be present in real-world scenes, although intrinsic cast shadows are confined to objects with concavities.
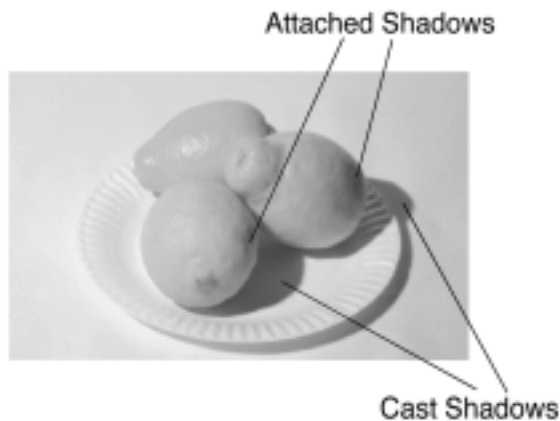
Figure 1. Cast and attached shadows.

The present experiments examined the effects of shadows in general, without distinguishing between these different types. The stimuli were digitized photographs of real objects, in which the different varieties of shadows are difficult if not impossible to isolate. Real images typically contain all types of shadows, and it is nearly impossible to photograph a real object such that only one

type of shadow is present. Therefore, although distinguishing among different types of shadows may be important for natural images, the present experiments focus on the more general effects of shadows.

Shadows have been shown to be useful for performing a variety of tasks. Bülthoff, Kersten, and Bülthoff (1994) demonstrated that shadows can cause flat objects to appear three-dimensional, and non-rigid motion to appear rigid. Cavanagh and Leclerc (1989) showed that shadows can provide information about 3-D shape. Observers can also use cast shadows to disambiguate convex from concave in shaded images (Erens, Kappers, & Koenderink, 1993). Yonas, Goldsmith, and Hallstrom (1978) showed that cast shadows can be used to infer the shape of the casting object, and the height of an object above the ground. Shadows can also influence perception of motion in depth (Kersten, Knill, Mamassian, & Bülthoff, 1996), and they are sufficient to produce stereo depth perception (Puerta, 1989). Waltz (1975) has used shadow geometry to computationally classify 3-D orientations of surfaces. Similarly, Shafer and Kanade (1983) used shadows to generate constraints on the orientations of surfaces in line-drawings of scenes. These studies demonstrate that shadows are useful for making judgments about both local shape and more global scene properties, which may be beneficial when recognizing objects.

Although shadows can be useful, they have also been shown to be ignored or even problematic. Berbaum, Bever, and Chung (1983) showed that, when making judgments about convexity, observers did not make use of information about direction of illumination provided by cast shadows. Shadows also have no effect on judgments of slant and tilt, and can impair judgments of illumination direction by causing the illumination to appear more oblique (Mingolla & Todd, 1986). Shadows are very difficult to identify computationally because cast shadow boundaries are generally unrelated to object boundaries (Cavanagh, 1991). Some models of object recognition argue that extraction of object edges is a crucial stage in recognition,

and object edges may also be important in the determination of shape and depth. Thus, if shadows introduce edges that are confused with object edges, they should impair recognition performance.

Do these findings extend to the task of object recognition? Warrington (1982) demonstrated that patients with right posterior lesions had difficulty recognizing photographs of common objects containing shadows. Moore and Cavanagh (1997) showed that two-tone images of novel objects with shadows are difficult to recognize. These two findings are consistent with the notion that shadows introduce confusing edges into an image. On the other hand, Tarr, Kersten, and Bülthoff (1998) demonstrated that cast shadows can improve recognition of novel geometric objects, suggesting that shadows provide useful information about shape or lighting conditions. The stimuli in the latter two studies were unfamiliar objects. The present experiments explore whether either of these results extends to recognition of more natural stimuli.

In the present experiments, digitized images of fruits and vegetables were used to study the impact of shadows on the recognition of natural objects. There are two reasons for using natural objects, rather than man-made or computer-generated objects. First, humans have evolved to recognize objects in the natural world, and therefore experiments with such stimuli should provide useful information about how human vision operates. Second, most studies of vision in general and object recognition in particular have used more artificial stimuli, and it is unclear how well the results of those studies apply to real-world stimuli. Common lighting models used for computer-rendered objects often ignore real-world effects such as mutual illumination and cast shadows. Although there is some evidence that computer-generated stimuli can be as effective as man-made objects for certain tasks using simple objects (Johnston & Curran, 1996), such

effects may not hold for more complex or natural stimuli.

Fruits and vegetables were chosen because they embody a variety of shapes and colors encountered in the natural world. It is also likely that human vision evolved in part to recognize biologically relevant stimuli such as food. Finally, such stimuli have been used in previous recognition studies (Wurm, Legge, Isenberg, & Luebker, 1993; Ostergaard & Davidoff 1985).

In the present experiments, observers identified food images presented on a computer screen. The images were presented with or without shadows, in order to determine whether shadows help or hinder recognition performance. We also examined whether the impact of shadows on recognition is influenced by the presence of information useful for identifying shadows. Shadows can often be identified by certain characteristics, such as a penumbra or continuity of chromaticity across a shadow boundary. These characteristics were manipulated to make shadow identification easier or more difficult, which should alter the impact shadows have on recognition.

EXPERIMENT 1: SHADOWS, COLOR, AND BLUR

The effect of shadows on recognition may depend on other information available in the image. For example, the presence of color, which has been shown to improve object recognition (Wurm *et al*., 1993; Ostergaard & Davidoff, 1985; Markoff, 1972), should alter the influence shadows have on recognition. In general, the hue and saturation of a surface patch will be very similar on both sides of a shadow boundary, allowing for easier labeling of an edge as a surface boundary *vs*. a shadow boundary. Computational techniques for finding shadows sometimes involve finding regions of low intensity that are similar in hue to a neighboring region (Nagao, Matsuyama, & Ikeda, 1979; Rubin & Richards, 1982). In color images, shadows should be easier to identify. Therefore, if shadows normally impair recognition, their

influence should be smaller in color than grayscale images.

It should be noted that color can only aid identification of shadows if the direct and ambient illumination in the scene are similar in their spectral distributions. In natural scenes, for example, there may be a tendency for shaded regions to contain more short-wavelength ("blue") energy, because such regions are illuminated primarily by the ambient light from the blue sky. However, Hailman (1979) demonstrated that the spectral distributions of light in sunny and shaded areas of several natural scenes (forests) are in fact very similar. Thus, the assumption of similar spectral distribution of direct and ambient light appears to be a reasonable one.

A second factor that may affect the impact shadows have on recognition is the spatial resolution of the image. Resolution can be reduced by blurring an image, which will disrupt the labeling of shadows in at least two ways. First, fine texture will be degraded or even lost. In a high-resolution image, the continuity of texture across a shadow boundary can signal the presence of a shadow, but blurring an image removes this information. Second, shadows often have a penumbra, or "fuzzy contour." In a high-resolution image, this blurry boundary can distinguish a shadow contour from the sharper object contours. In a blurry image, however, all contours are similar to shadow penumbrae. In fact, Hering (1874, 1964) showed that occluding the penumbra of a shadow causes the contour to appear as an albedo change, rather than an illumination change. Therefore, in a blurry image, shadow edges are more easily confused with surface edges. If shadows normally impair performance, then making them more difficult to identify by blurring the image should cause even larger deficits in recognition.

The overall impact of shadows on recognition, as well as these expected interactions between shadows, color, and image resolution, were tested in Experiment 1. Observers identified images that were displayed with or without shadows, color, and blur.

Methods

Observers

Twenty-four undergraduate psychology students (ages 18 to 31) at the University of Minnesota participated for class credit. All were fluent English speakers, had normal or corrected-to-normal visual acuity (Snellen acuity of 20/20 or better) and normal color vision (as measured by the D-15 color test), and gave informed consent.

Stimuli and Apparatus

The stimuli were digitized images of 24 types of fruits and vegetables (**Table 1**). Each food appeared in two different forms, such as slices, groups of the same item, or different varieties of the same food. These will be referred to as different *poses*. Care was taken to select objects and poses that were likely to produce shadows (e.g. groups of foods and foods with concavities).

The stimuli were photographed onto color slides (Kodak Ektachrome 100 film), using a Nikon FM2 camera with a Tamron SP90mm F2.5 lens. Each item sat on a white 8-inch paper plate, against a backdrop of a white sheet. Photographs were taken from an angle of 45° above the stimuli. The distance from the lens to the stimulus was 3 feet. Ambient illumination was provided by three 100-watt soft white bulbs and one 90 watt halogen bulb. Directional illumination was provided by a Vivitar 283 electronic flash, subtending roughly 9.3 x 5.5 degrees. Subjectively, this resulted in fairly sharp shadows (see **Figure 2**).

Table 1. List of stimuli used, and the two poses of each.

STIMULUS	POSE 1	POSE 2
Apples	Group red[#]	Slices red
Bananas	Bunch[#]	Pile (not in a bunch)
Cantaloupe	1/4 slices	1/8 slices
Carrots	Bunch[##]	Sticks
Celery	Bunch[##]	Sticks
Grapes	Bunch green	Bunch red
Kiwi	Group	Slices
Lemons	Group[##]	Slices
Lettuce	Green leaf[#]	Red leaf
Limes	Group	Slices
Mushrooms	Group	Slices
Olives*	Group black	Group green
Onions	Group green	Group yellow
Oranges	Group	Slices
Peanuts*	Group in shells	Group shelled
Peas*	Group of pods	Group not in pods
Peppers	Group green[##]	Group red
Pickles*	Spears	Slices
Pineapple	Chunks	Rings
Potatoes	Group brown	Group red
Radishes	Group with greens[##]	Group without greens
Strawberries	Group[#]	Slices
Tomatoes	Group[#]	Slices
Watermelon	1/4 slices	1/8 slices

* Items used in practice trials only.

[#] Items judged as the more prototypical pose on 100% of the trials.

[##] Items judged as the more prototypical pose on at least 65% of the trials.

The presence of shadows was manipulated by changing the direction of illumination. Each stimulus was photographed with the flash illumination from three different directions: for the *no-shadow* condition, the stimulus was photographed with the illuminant near the camera's lens[1], i.e. on the viewing axis (**Figure 2**, top left). For the *shadow* condition, the illuminant was placed 45° to the right or left of the viewing axis, while maintaining the same distance to the stimulus (**Figure 2**, top right). The types of shadows present included both attached and cast, and both intrinsic and extrinsic.

The slides were scanned into a Power Macintosh 7100 computer using a Polaroid SprintScan slide scanner. The resulting images

were 300 pixels (10.6 cm) horizontally by 187 pixels vertically (6.6 cm), equivalent to 3.03° by 1.89° at a viewing distance of 2 meters. The images were presented on a white background (67 cd/m^2) on a 640 by 480-pixel Apple 17-inch color monitor.
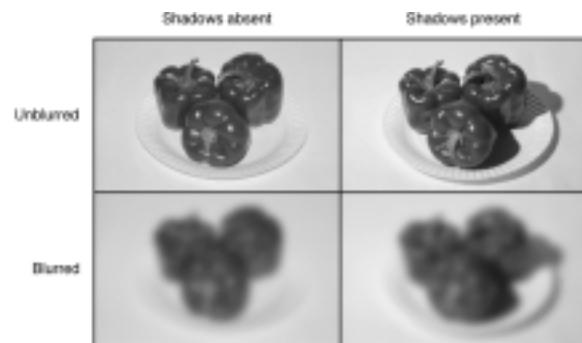
Figure 2. Peppers rendered in grayscale with and without shadows present, and in high-resolution and blurred. Full-color versions of these four conditions were also used.

Eight-bit grayscale images were derived from the color images using Equilibrium Technologies DeBabelizer software. The results of the conversion were checked photometrically to confirm that there were no differences in luminance between the color and grayscale images. These measurements revealed only small differences, almost always less than 5%, and never greater than 15%.

The color and grayscale im^{1}ages were low-pass spatial-frequency filtered using the *convolve* program on a Silicon Graphics Power Indigo 2 computer. A Gaussian filter with a 1/e bandwidth of 6 cy/deg (20.4 cy/object) was used. This is similar to the shape and bandwidth of the filter used by Wurm *et al.* (1993). The blurred images were produced by the convolution of the filter and the input images; for color images, the filter was convolved separately with the red, green, and blue channels.

Examples of grayscale images of a pepper in each *shadow* x *blur* condition is shown in **Figure 2**. The color versions of these images contained color but were otherwise identical to the grayscale images.

The experiment was run using RSVP software (Williams & Tarr, 1998). Naming latencies were recorded using the CMU Button Box (accurate within 1 msec). The observers' verbal responses (the names of the foods) were compiled by the experimenter using an IBM XT personal computer.

Procedure

Observers were provided with a written list of the foods they would be asked to recognize. The list contained the 20 food category names (e.g. *apples*), but not the specific poses (e.g. *sliced apples*). They were told that the listed items were the only ones that would be presented, and that each item could appear in different forms (groups, slices, etc.). Observers were given as long as they needed to review the list of foods, and they were allowed to refer to it between trials during the experiment.

Each trial in the experiment consisted of a fixation cross for 500 msec, followed by a brief tone, and then a food image. The purpose of the tone was to signal the beginning of a trial to the experimenter, who could not see the images being presented to the participants. The observer's task was to name the food aloud as quickly and accurately as possible. Observers were instructed to name the food category, rather than the specific pose. Each image remained on the screen until the observer responded or 10 seconds had passed. Reaction time was measured as the time between the onset of the stimulus and the observer's verbal response. No feedback was provided.

There were 8 blocks of 40 trials. A single *shadow* x *color* x *blur* condition was tested in each block. Within each block, each food was presented once in each of its two poses. For the conditions containing shadows, illumination direction was chosen randomly on each trial.

The 8 *shadow* x *color* x *blur* conditions were presented in a different random order to each observer. The order was constrained such that each condition (*shadow* x *color* x *blur*)was presented in a given block (1st, 2nd, … 8th) to at
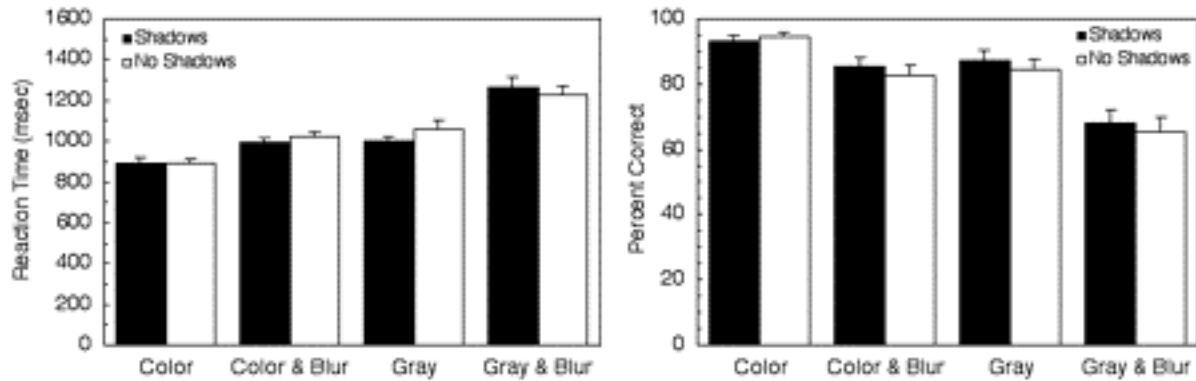
Figure 3 . (**a**) Reaction times for Experiment 1. Each reaction time is the average of the median reaction times for each observer in each condition. Standard errors are plotted. (**b**) Accuracy for Experiment 1. Each percentage is the average of the percentages for each observer in each condition. Standard errors are plotted.

least one observer and no more than four observers throughout the experiment (e.g. between one and four observers ran the *shadow–gray–high-resolution* condition first, between one and four observers ran this condition second, etc.).

Prior to the experiment, observers completed 16 practice trials. These trials contained the 4 food items that pilot subjects demonstrated to be the most difficult (took the longest time) to recognize. The practice items are tagged in **Table 1** with an asterisk.

Results

Median reaction times for trials in which a correct response was made are plotted in **Figure 3a**. The main finding was that shadows did not influence performance, regardless of the color and resolution conditions. Reaction times (averaged across color and resolution conditions) were about the same whether shadows were present or absent (1037 *vs*. 1051 msec). An ANOVA run on the log median reaction times revealed

no main effect of shadows (F(1,23)=0.87, p=0.361) and no interactions between shadows and color (F(1,23)=0.06, p=0.806) or shadows and resolution (F(1,23)=0.01, p=0.508).

Consistent with the findings of Wurm *et al*. (1993), response times were faster when the stimuli were presented in color than in grayscale (951 *vs*. 1137 msec) (F(1,23)=98.16, p<.001), and when they were presented in high resolution than blurred (962 *vs*. 1126 msec) (F(1,23)=88.94, p<.001). There was no significant interaction between color and resolution (F(1,23)=2.47, p=0.130).

Accuracy is plotted in **Figure 3b**. As with reaction times, shadows did not influence performance. Averaged across color and resolution conditions, accuracy was about equal for images with and without shadows (83% *vs*. 82%). An ANOVA run on the arcsine of the proportion of correct responses revealed no main effect of shadows (F(1,23)=1.08, p=.309) and no interactions between shadows and color (F(1,23)=1.08, p=0.309) or shadows and resolution (F(1,23)=0.98, p=0.332).

Accuracy was higher when the stimuli were presented in color than in grayscale (89% *vs*. 76% correct) (F(1,23)=92.85, *p*<.001), and when they were presented in high resolution than blurred (90% *vs*. 75%) (F(1,23)=85.28, *p*<.001). Additionally, there was a significant interaction between image resolution and color (F(1,23)=5.87, *p*<.030)—blurring the images led to a greater decline in performance for grayscale images (86% *vs*. 67%) than for color images (94% *vs*. 84%).

Discussion

Performance was best with color images and with high-resolution images, consistent with the findings of Wurm *et al*. (1993). However, shadows did not influence recognition, even in images where shadows should be more difficult to identify, i.e. grayscale and blurred images. We explored several possible explanations for the lack of an effect of shadows.

Not enough shadows

One possibility is that there simply were not many shadows in the images used here. A measure of the "amount" of shadows in the images would therefore be useful. Although there is currently no satisfactory objective (computational) method for identifying shadows in an image, behavioral measures of whether the shadows are noticeable can be obtained.

An experiment was run to examine whether the shadows were perceptually salient in the images. Eight observers viewed two of the food images (from Experiment 1) presented side-by-side, and indicated with a key-press which one contained more shadows. The two images were both of the same food in the same pose, and were presented in full color at high resolution. One of the images was a *no-shadow* image from the original experiment, and the other was the corresponding *shadow* image. Each pair of

images remained on the screen until the observer responded, and no feedback was provided.

There were 8 blocks of 40 trials. For the *shadow* images, there were an equal number of trials containing left- and right-illuminated stimuli. The *shadow* images were presented on the right and left sides of the screen an equal number of times.

If the shadows were not perceptually salient, then observers should perform this task rather poorly—they should not be able to tell which of the images is the one with more shadows. The results showed, however, that observers had little trouble discriminating the *shadow* images from the *no-shadow* images, achieving 96% accuracy in performing the task. This suggests that most of the shadows were perceptually salient.

Prototypicality

Another factor that may determine whether shadows influence recognition is the food pose used. Specifically, the effect of shadows on recognition might depend on the prototypicality of the food pose. For example, prototypical poses might contain a great deal of information (e.g. about object shape) useful for recognizing an object, and such information may be sufficient to override any confusion caused by shadows. Less prototypical poses, lacking this information, might therefore be more strongly affected by shadows.

An experiment was run to determine which poses appeared more prototypical. Five experienced psychophysical observers were presented with the two poses of each food side-by-side on the computer screen. The images were displayed in full color at high resolution with no shadows. The name of the food was displayed above the images. Observers were asked to decide which of the two poses was more prototypical for that item, and to respond with a key-press. They were given as long as needed to perform this task. Each pair of images was presented 4 times, with each image shown twice on the left and twice on the right. There were 4

blocks of 20 trials, and all pairs were presented twice by the end of block 2.

The results revealed no correlation (R^2=.005) between the proportion of times an image was rated as more prototypical and normalized shadow effect (difference between *shadow* and *no-shadow* reaction times divided by the average reaction time for *shadow* and *no-shadow* conditions). This means that shadows had the same influence (or lack of influence) regardless of how prototypical an image was judged to be. There was a low correlation (R^2=.12) between prototypicality and reaction time, indicating a slight tendency for more prototypical poses to be recognized more quickly.

The lack of correlation between prototypicality and shadow effect cannot be explained by the absence of prototypical images. Five food poses were highly prototypical—they were judged as the more prototypical pose on 100% of the trials. These items are tagged with a "#" in **Table 1**. Five other food poses were judged as more prototypical on at least 65% of the trials. These items are tagged with a "##" in **Table 1**. For the remaining items, neither pose was more likely to be judged as prototypical. The food images thus included both prototypical and non-prototypical poses.

Priming

Throughout the experiment, each food category (e.g. *apple*) was presented several times. It is possible that priming occurred for the category names, i.e., performance on the later presentations of a food item may have improved relative to its first presentation. Such priming could be strong enough to overcome small effects of shadows. However, if priming was present at all in Experiment 1, it was very weak. There was no difference in accuracy for the first *vs.* second presentation of a food, and median reaction times improved by only 8 msec

between the first and second presentations. This reaction time difference is much smaller than was found by Biederman and Gerhardstein (1993), for example, who found improvements of over 100 msec for primed objects. Thus, priming cannot explain the lack of influence of shadows.

Another possibility is that priming occurred between different viewing conditions. For example, recognition in a *shadow* condition may have improved after completing a *no-shadow* condition. This possibility was also unsupported by our data: performance was unaffected by the order of presentation of shadow conditions.

Difficulty of recognition

Shadows might have an effect only on the foods that are most difficult to recognize. An ANOVA was run on the median reaction times from the trials containing the 15 food poses that were incorrectly identified the most often. The findings were the same as those obtained using all images: performance was best with color ($F(1,23)$=48.9, p<.001) and high-resolution ($F(1,23)$=14.0, p<.001) images, but shadows had no significant effect ($F(1,23)$=0.97, p=.34). Thus, difficulty of recognition cannot account for the lack of a shadow effect.

Effects with specific foods

Although shadows had no effect on recognition overall, they may have affected recognition of particular foods. A Tukey HSD test (α=.05) revealed that for most foods, shadows had no significant effect; however, a few foods were significantly affected by the presence of shadows, although the results reveal no clear pattern. Recognition of *shadow* images was slower than for *no-shadow* images by more than 700 msec for red lettuce and red potatoes; recognition of *shadow* images was faster than for *no-shadow* images for lemon slices, orange slices, groups of oranges, and radishes with greens.

Slower recognition with *shadow* images may be a result of confusion between albedo changes

and shadow boundaries, particularly for the red lettuce. However, it is unclear why such confusions would not arise in other food images. Faster performance with *shadow* images of lemon slices was likely a result of enhanced contrast at the edges of the lemon slices, as the cast shadows provided a darker background for the light-colored food. It is not clear why shadows would offer a specific benefit to the recognition of the orange and radish images.

EXPERIMENT 2: TWO-TONE IMAGES

A possible explanation for the lack of a shadow effect in Experiment 1 is that the images, even when degraded by blurring or removing color, still contained substantial information for performing the task—information that could override any shadow effects. Another way to degrade the images in a way that may reveal shadow effects is to threshold them, creating a "two-tone" image. This involves setting all pixel luminances above a certain value to white, and all those below it to black. The result is a black-and-white image in which illuminated areas are white and shadowed areas are black (Mooney, 1957).

In a two-tone image, all edges (object and shadow) are defined in the same way, i.e. by an abrupt luminance change. This can greatly impair object recognition, in that cast shadow areas are often interpreted as part of the object (Cavanagh, 1991). In fact, Moore and Cavanagh (1997) demonstrated that 2-tone images of objects with shadows are difficult to recognize. An effect of shadows may therefore emerge when two-tone images of the food stimuli are used.

Methods

Observers

Eight undergraduate psychology students at the University of Minnesota participated in this experiment for class credit. All had normal or corrected-to-normal visual acuity (Snellen acuity of 20/20 or better), had normal color vision (as measured by the D-15 color test), were native English speakers, and gave informed consent.

Stimuli and Apparatus

The stimuli were thresholded versions of the grayscale high-resolution images used in Experiment 1. Thresholding was done using Equilibrium Technologies DeBabelizer software. The threshold for each image was taken as 30% of the maximum image luminance. This threshold value preserved both the general object shape and the shadows in the images. All pixel values above the threshold level were set to white (73 cd/m^2), and all values below it were set to black (<1 cd/m^2). A sample image is shown in **Figure 4.**


Figure 4. A 2-tone image of peppers with shadows present.

Procedure

The procedure was similar to that described in Experiment 1. The task was to verbally name each food as quickly and accurately as possible. Before the experiment began, observers were presented with a "slide show" of the full-color high-resolution food images (both *shadow* and *no-shadow*) on the computer screen. They were also given a set of 8 practice trials with two-tone images.

There were 40 *no-shadow* images (20 foods, 2 poses each) and 80 *shadow* images (20 foods, 2 poses, 2 illumination directions). Each image was presented once. The experiment contained 2 blocks of 60 trials, and the observer was allowed to rest between blocks. The trials were presented in random order, and no feedback was provided.
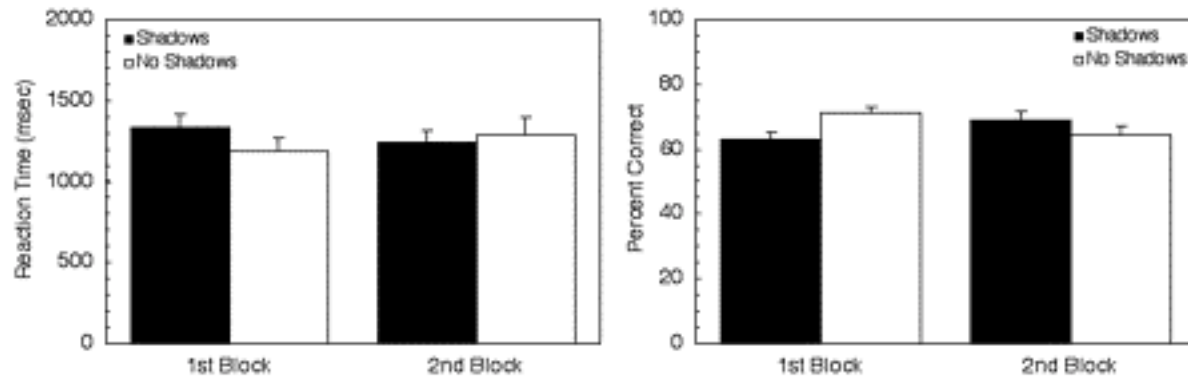
Figure 5 . (**a**) Reaction times for Experiment 2. Each reaction time is the average of the median reaction times for each observer. Standard errors are plotted. (**b**) Accuracy for Experiment 2. Each percentage is the average of the percentages for each observer. Standard errors are plotted.

Results

Median reaction times for trials in which a correct response was made did not differ significantly across shadow condition or block (**Figure 5a**). However, an ANOVA revealed a significant interaction between block and shadow condition (F(1,7)=7.76, *p*<.03). In the first block of trials, observers were faster with *no-shadow* than *shadow* images (1178 *vs*. 1295 msec) (confirmed by a Tukey HSD test, α=.05). In the second block, reaction times were not significantly different for the two shadow conditions (1277 *vs*. 1184).

Similar results were obtained with the accuracy data (**Figure 5b**): there was no main effect of block or shadow condition, but there was a significant *shadow* x *block* interaction (F(1,7)=23.87, *p*<.01). As with reaction times, performance was better with *no-shadow* than *shadow* images in the first block of trials (71% *vs*. 63%) (Tukey HSD test, α=.01), but not in the second block (64% *vs*. 69%).

Discussion

Shadows impaired performance on the early trials (i.e. in the first block) of the experiment. This is consistent with Moore and Cavanagh's (1997) finding that shadows impair recognition of two-tone stimuli. The impairment found here is likely due to the difficulty in distinguishing shadow contours from object contours in the impoverished image, since all contours were defined in the same way. It has been suggested (Moore & Cavanagh, 1997) that top-down processing is required for labeling shadows in two-tone images, and this may explain the deficit associated with *shadow* images in this experiment.

The effect of shadows disappeared for later trials of the experiment. It is possible that

observers were able to "learn" about shadow or illumination conditions during the early trials, and then use this information later on. This is consistent with the argument that higher-level mechanisms are used in processing shadows (Cavanagh 1991). Observers may have also obtained sufficient information about the objects themselves during the early trials (e.g. their shapes or sizes), such that shadows contributed relatively little noise to the later trials.

EXPERIMENT 3: LIMITED EXPOSURE TIME

Another reason shadows may have had no detrimental effect on recognition in Experiment 1 is that observers were given plenty of time (up to 10 seconds) to respond. Even the 1-2 second reaction time that was typical in Experiment 1 may have allowed ample time for processing shadows. Cavanagh (1991) has suggested that the labeling of shadows occurs rather late, requiring top-down processing, and that early bottom-up processes do not distinguish shadow contours from object contours. The procedure used in Experiment 3 attempted to limit processing time, so that there would be little time to identify shadows. The stimuli were presented very briefly, followed by a mask. Such backward masking paradigms have been used to control the amount of time observers spend processing a stimulus (e.g. Reynolds, 1981). If labeling of shadows does not occur until later, then larger shadow effects should occur when a limit is imposed on the amount of time observers have for processing shadows.

Methods

Observers

Three volunteers (ages 28 to 32) at the University of Minnesota for payment, and 14 volunteers (ages 18 to 30) at St. Cloud State University participated for course credit. All observers had normal or corrected-to-normal visual acuity (Snellen acuity of 20/20 or better), had normal color vision (as measured by the D-15 color test), were native English speakers, and gave informed consent.

Stimuli and Apparatus

The apparatus and stimuli were identical to those used in Experiment 1. There was also a mask, consisting of a collage of the food images (see **Figure 6**). The mask included all foods and poses, and both *shadow* and *no-shadow* images. The mask was presented in color for the color conditions, in grayscale for the grayscale conditions, and blurred (using the same filter as in Experiment 1) for the blur conditions. The masks were the same size as the food images.
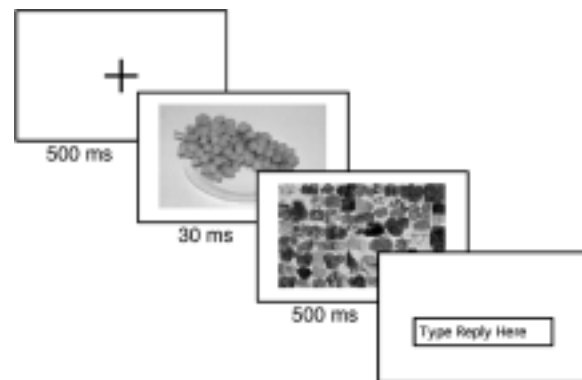
**Figure 6**. Recognition task used in Experiment 3.

Procedure

Each trial consisted of a fixation cross for 500 msec, followed by a food image for 30 milliseconds, the mask for 500 msec, and finally a reply box on the screen, in which the observers were instructed to type the name of the food (**Figure 6**). No feedback was provided. As in Experiment 1, observers were given a list of the foods they would be asked to recognize. They ran 8 practice trials before the experiment began.

There were 8 blocks of 40 trials. Each block contained one *shadow* x *color* x *blur* condition.

In a block, each food was presented once in each of its two poses. The conditions were presented in a different random order to each observer.

Results and Discussion

Shadows did not influence observers' accuracy in recognizing the foods ($F(1,15)=0.52$, $p=.48$), as shown in **Figure 7**. Performance was significantly better with color than grayscale images ($F(1,15)=137.61$, $p<.001$) and with high-resolution than blurred images ($F(1,15)=98.19$, $p<.001$). No interactions were found. In all conditions, accuracy was lower than in Experiment 1 (by 30% on average).
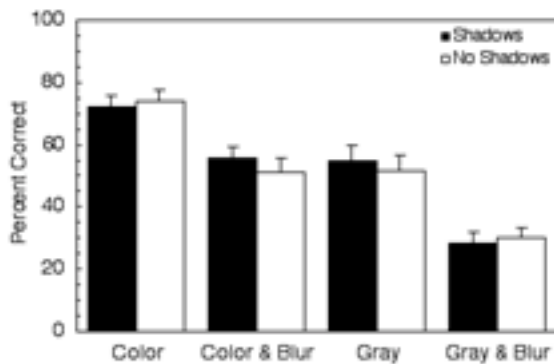
Figure 7. Accuracy for Experiment 3. Each percentage is the average of the percentages for each observer in each condition. Standard errors are plotted.

Shadows did not affect recognition even when observers were given only 30 msec to visually process the objects and shadows. It is possible that the exposure durations were still long enough to allow for labeling of shadows. However, it seems that, in general, no matter how difficult the task is, shadows have little if any effect on recognition.

GENERAL DISCUSSION

The overall finding was that recognition of natural objects was highly invariant to the presence of shadows. Several experimental conditions were explored here, but none revealed any strong effect of shadows. One interpretation of these findings is that the visual system contains a mechanism for processing shadows quickly and early on. If this is the case, then the mechanism is very fast, completing its work within the first 30 msec of processing.

The findings argue against purely image-based approaches to recognition, in which the visual system must take time to detect and label shadows. Instead, the results are consistent with more abstract feature-based object representations, in which objects are represented by features that are invariant to spurious effects like shadows. What might these features be? In the present studies, color, texture, and boundary sharpness were explored. However, even when these cues were degraded or removed, shadows had no impact on recognition. This implies that these features were not necessary for invariant recognition, and that other shadow-invariant cues were still present.

One such "cue" could be the global shape of the stimuli. For example, apples tend to be round, bananas curved and elongated, etc. In blurry grayscale images, and even in 2-tone images, this global shape cue is still present. Since shadows tend to be fairly local, global shape information is not disrupted by their presence, and recognition can proceed using this feature.

The familiar stimuli used here apparently contained plenty of redundant information for recognition, such that shadows contributed relatively little to the recognition process. Cavanagh (1991) has suggested that recognition begins with a simple match between an image and a stored representation. Any information (such as shadows) that is not matched at this early stage is processed later, using higher-level mechanisms. When there are many redundant cues to the identity of an object, this early match might be quite successful, and shadows may provide very little extra information, or add very little extra noise. Although the degraded

renderings used here (blurred, 2-tone, or gray-scale images) did lead to a reduction in recognition performance, these images apparently still contained sufficient information for recognition in the presence of shadows.

It is also possible that shadows had no effect because the task required distinctions between basic-level categories (e.g. *apple vs. orange*), rather than subordinate-level categories (e.g. *Red Delicious apple vs. Granny Smith apple*). In subordinate-level recognition, observers would need to rely on finer distinctions between texture, color, and precise contour information. Moreover, global shape information is typically less useful for making such discriminations. Shadows might impair performance under such conditions. Consistent with this prediction, Braje, Kersten, Tarr, & Troje (1998) have shown that cast shadows do impair face recognition in a same-different matching task.

Shadows did not impair recognition in the present studies, but neither did they improve recognition. This lack of improvement suggests that, at least under the stimulus conditions used here, shadows were not used for the extraction of 3-D surface shape or the location of objects in 3-D space. Tarr *et al.* (1998) have suggested that in order to use shadows for shape extraction, the visual system must do two things. First, it must be able to identify a region as a shadow, rather than as a change in surface orientation or albedo. This could be accomplished using low-level information, such as contrast- or color- invariance or a shadow's penumbra. However, removing such low-level information did not induce shadow-dependent performance in the present experiments. This implies that either such information is not used for identifying shadows, or that any information gained about shadows is not used at higher levels involved in recognition. A second requirement proposed by Tarr *et al.* is that the visual system must relate the shadow to the casting object. Observers appeared to have no difficulty in doing this, in that shadows usually did not impair performance. However, the fact that shadows did not improve performance suggests that the observers did not make use of shadows for extracting shape information (or that such shape information was simply not needed for these tasks). In fact, Cavanagh (1995) has suggested that shadows are not normally used for recovering shape. It may be the case that shadows are only a useful cue when novel shapes are used (as in Tarr *et al.'s* study), or when no other information is available.

It should be noted that the few shadow effects that were obtained in the present experiments could be due not to shadows but to the change in lighting direction that was used to produce the shadow images. It would therefore be useful to perform similar experiments using computer-generated objects, where shadows can be manipulated independently of other scene parameters.

Another important issue is the particular contributions made by the different types of shadows to the recognition process. For example, both attached and cast shadows might aid recognition by providing information about lighting direction or other global scene properties, thus indirectly contributing to shape recovery; or, recognition processes could make use of the precise shapes of shadows to determine local object shape. Likewise, intrinsic and extrinsic shadows may play different roles in recognition. Intrinsic shadows are always cast onto the same "background," i.e. the object itself. They may therefore be well-learned for familiar stimuli (such as used in the present experiments), and may even be helpful for recognition. Extrinsic shadows, on the other hand, might be problematic, since it would be prohibitive to learn all possible background images. However, Braje *et al.* (1998) found that intrinsic cast shadows impaired face recognition, suggesting that they are not well-learned. The present data cannot speak to these issues, since all types of shadows were present in the stimuli. Further research is necessary to examine the

particular contributions of different types of shadows.

The effect of the experimental design must also be examined. In Experiment 2, the shadow effect disappeared with practice, implying that observers can quickly learn about shadows, illumination conditions, or other stimulus conditions, and use this information for later recognition. This result is highly relevant to experiments 1 and 3. In these two experiments, each observer was exposed to every condition over many trials. If observers can learn about stimulus conditions in a small number of trials, any effects of shadows on performance may disappear when data are averaged across blocks and observers. It would therefore be useful to repeat these experiments using a design in which each observer is exposed to only a small number of trials of a single experimental condition.

Finally, our experiments have examined only one type of object (fruits and vegetables). Although this class of objects appears to be broadly representative of the shapes and colors of objects encountered in the natural world, further research is needed to determine whether these findings extend to other types of objects.

This study has shown that recognition of natural objects is remarkably invariant to the presence of shadows over a wide range of viewing conditions. Even with highly degraded stimuli, the visual system can still use the available information to recognize objects regardless of the presence of shadows. This is probably a consequence of strong evolutionary pressure for the development of object representations that can effortlessly cope with shadows under everyday conditions. Our findings argue against image-based models of object recognition in which shadows must somehow be detected, labeled, and discounted. Instead, our findings favor a more abstract representation of object features that are immune to the presence of shadows.

REFERENCES

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

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

Biederman, I. & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(6), 1162-1182.

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

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

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.

Bülthoff, I., Kersten, D., & Bülthoff, H. H. (1994). General lighting can overcome accidental viewing. *Investigative Ophthalmology and Visual Science, Suppl.*, *35*(4), 1741.

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).

Cavanagh, P. (1995). A horse of a different color: shadows have to be darker but shading does not. *Investigative Ophthalmology and Visual Science, Suppl.*, *36*(4), S184.

Cavanagh, P. & Leclerc, Y. G. (1989). Shape from shadows. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(1), 3-27.

Edelman, S. (1995). Representation, similarity, and the chorus of prototypes. *Minds and Machines*, *5*(1), 45-68.

Edelman, S. & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, *32*(12), 2385-2400.

Erens, R. G. F., Kappers, A. M. L., & Koenderink, J. J. (1993). Perception of local shape from shading. *Perception and Psychophysics*, *54*(2), 145-156.

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

Hailman, J. P. (1979). Environmental Light and Conspicuous Colors. In J. Edward H. Burtt (Eds.), *The Behavioral Significance of Color* (pp. 291-354). New York: Garland STPM Press.

Hering, E. (1874, 1964). *Outlines of a theory of the light sense*. Translated from the German by L. M. Hurvich and D. Jameson. Cambridge, Mass.: Harvard University Press.

Johnston, A. & Curran, W. (1996). Investigating shape-from-shading illusions using solid objects. *Vision Research*, *36*(18), 2827-2836.

Kersten, D., Knill, D. C., Mamassian, P., & Bülthoff, I. (1996). Illusory motion from shadows. *Nature*, *379*, 31.

Markoff, J. I. (1972). *Target recognition performance with chromatic and achromatic displays* (Report No. SRM-148). Honeywell.

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.

Mingolla, E. & Todd, J. T. (1986). Perception of solid shape from shading. *Biological Cybernetics*, *53*, 137-151.

Mooney, C. M. (1957). Age in the development of closure ability in children. *Canadian Journal of Psychology*, *11*(4), 219-227.

Moore, C. & Cavanagh, P. (1997). Recovery of 3D volume from 2-tone images of novel objects. *submitted to Cognition (special issue on object recognition).*

Nagao, M., Matsuyama, T., & Ikeda, Y. (1979). Region extraction and shape analysis in aerial photographs. *Computer Graphics and Image Processing*, *10*, 195-223.

Ostergaard, A. L. & Davidoff, J. B. (1985). Some effects of color on naming and recognition of objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(3), 579-587.

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

Puerta, A. M. (1989). The power of shadows: shadow stereopsis. *Journal of the Optical Society of America A*, *6*(2), 309-311.

Reynolds, R. I. (1981). Perception of an illusory contour as a function of processing time. *Perception*, *10*, 107-115.

Rubin, J. M. & Richards, W. A. (1982). Color vision and image intensities: when are changes material? *Biological Cybernetics*, *45*, 215-226.

Shafer, S. A. & Kanade, T. (1983). Using shadows in finding surface orientations. *Computer Vision, Graphics, and Image Processing*, *22*, 145-176.

Tarr, M. J., Kersten, D., & Bülthoff, H. H. (1998). Why the visual recognition system might encode the effects of illumination. *Vision Research38,* 2259-2275.

Waltz, D. (1975). Understanding line drawings of scenes with shadows. In P. H. Winston

(Eds.), *The Psychology of Computer Vision* (pp. 19-91). New York: McGraw-Hill.

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

Williams, P. & Tarr, M.J. (1998). *RSVP: Experimental control software for MacOS* [Online]. Available: http://psych.umb.edu/rsvp/.

Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(4), 899-911.

Yonas, A., Goldsmith, L. T., & Hallstrom, J. L. (1978). Development of sensitivity to information provided by cast shadows in pictures. *Perception*, *7*, 333-341.

ACKNOWLEDGEMENT

FOOTNOTE

1. The illuminant was located slightly above the camera lens, rather than right at the lens. Therefore the resulting images did contain a very small amount of shadow (subjectively there were practically no shadows), and this condition could be more accurately termed the "almost no shadow" condition. For brevity and clarity, however, it will be referred to as the "no shadow" condition.