Eye movement patterns during the recognition of three-dimensional objects: Preferential fixation of concave surface curvature minima

E. Charles Leek

Wales Institute for Cognitive Neuroscience, School of Psychology, Bangor University, Bangor, UK

Filipe Cristino

Wales Institute for Cognitive Neuroscience, School of Psychology, Bangor University, Bangor, UK

Lina I. Conlan

Wales Institute for Cognitive Neuroscience, School of Psychology, Bangor University, Bangor, UK

Candy Patterson

Wales Institute for Cognitive Neuroscience, School of Psychology, Bangor University, Bangor, UK

Elly Rodriguez

Wales Institute for Cognitive Neuroscience, School of Psychology, Bangor University, Bangor, UK

Stephen J. Johnston

Department of Psychology, School of Social Sciences, Brunel University, Uxbridge, UK

This study used eye movement patterns to examine how high-level shape information is used during 3D object recognition. Eye movements were recorded while observers either actively memorized or passively viewed sets of novel objects, and then during a subsequent recognition memory task. Fixation data were contrasted against different algorithmically generated models of shape analysis based on: (1) regions of internal concave or (2) convex surface curvature discontinuity or (3) external bounding contour. The results showed a preference for fixation at regions of internal local features during both active memorization and passive viewing but also for regions of concave surface curvature during the recognition task. These findings provide new evidence supporting the special functional status of local concave discontinuities in recognition and show how studies of eye movement patterns can elucidate shape information processing in human vision.

Keywords: eye movements, object shape perception, object recognition, surface curvature extrema, concavity, convexity, visual saliency

Introduction

Fixational eye movement patterns have been widely studied in a variety of domains including reading, scene and face perception, object localization, and visual search (Henderson, Brockmole, Castelhano, & Mack, 2007; Land, Mennie, & Rusted, 1999; Liversedge & Findlay, 2000; Mannan, Ruddock, & Wooding, 1997; Rayner, 1995; Renninger, Verghese, & Coughlan, 2007; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006). Surprisingly, beyond two-dimensional (2D) pattern recognition (e.g., Renninger, Coughlan, & Verghese, 2005; Renninger et al., 2007), there have been no detailed analyses of eye movement patterns during three-dimensional (3D)

visual object recognition. The ability of the human visual system to rapidly categorize 3D objects across variation in viewing conditions caused by changes in scale, lighting and viewpoint, is a truly remarkable accomplishment for a biological system that far surpasses, in adaptability and robustness, the most advanced computer vision systems. Although everyday object recognition can be accomplished quickly, and often within a single fixation for a distal stimulus, previous studies, using 2D stimuli, have shown that fixation patterns can be highly informative about shape processing during perception (e.g., Melcher & Kowler, 1999; Renninger et al., 2005, 2007; Vergilino-Perez & Findlay, 2004). For example, Melcher and Kowler (1999) have shown that initial landing position during saccadic localization is driven by a representation of target shape that

determines center-of-gravity (COG) landing sites. Recent evidence also suggests that the perception of information about object presence and identity in a scene may be largely restricted to a relatively small region around the current fixation point (Henderson, Williams, Castelhano, & Falk, 2003), although the nature of the shape information processed during fixations and the role of this information in object recognition remain unclear.

In this context, a variety of different object recognition theories have been proposed, which make different claims about how shape is represented. For example, some accounts propose that shape classification is based on class-specific appearance or image-based feature hierarchies computed across multiple spatial scales (e.g., Ullman, 2006; Ullman & Bart, 2004; Ullman, Vidal-Naquet, & Sali, 2002). Other image-based models have hypothesized the use of 2D views, or aspects, that conjointly encode information about shape and the spatial locations of image features (e.g., Edelman & Weinshall, 1991; Ullman & Basri, 1991). In contrast, geometric accounts like structural description theories propose that shape perception depends on the decomposition of object shape into generic primitives (e.g., generalized cylinders, geons, or surfaces) and that recognition is mediated by representations that independently encode information about these primitives and their spatial configuration (Biederman, 1987; Hummel & Stankiewicz, 1996; Leek, Reppa, & Arguin, 2005; Leek, Reppa, Rodriguez, & Arguin, 2009; Marr & Nishihara, 1978). These approaches are not mutually exclusive. Some recent hybrid models have suggested that both image-based and structural description approaches can be accommodated within the same framework (Foster & Gilson, 2002; Hummel & Stankiewicz, 1996).

However, regardless of whether an image-based, structural description or some other form of representation is proposed, there remains considerable debate about the specific kinds of shape information, and shape analysis algorithms, that underlie object recognition. In principle, there are several different kinds of information from low-level image contrasts (e.g., simple edges derived from luminance boundaries) to intermediate or higher level features derived from combinations of lower level image properties (e.g., vertices, curvature discontinuities, volumetric parts) that may be used during shape perception. Furthermore, the availability of specific kinds of shape information is dependent on the spatial scale of perceptual analysis. For example, some kinds of image features that may be useful in recognition are likely to be detected only at a relatively coarse spatial scale. These include edge co-linearity (parallelism), elongation, symmetry, aspect ratio, and global outline (e.g., Biederman, 1987; Hayward, 1998; Hayward, Tarr, & Corderoy, 1999). A good case in point is shape elongation. Determining elongation requires access to a relatively complete perceptual representation of object shape, but it can be computed from relatively low spatial frequency information. In contrast, other potentially useful shape features may (and in some cases,

must) be computed locally at a relatively finer spatial scale. These include the presence of edge boundaries, corners, vertices, surface depth, and curvature. Other object properties including, for example, color and texture can also be computed locally. In some situations, relatively coarse global image features may be sufficient for shape classification in specific contexts—such as distinguishing between a banana (curved axis) and a cucumber (straight axis) on a kitchen table. However, real-world scenes are often cluttered, containing objects that partially occlude each other, making it difficult to reliably recover global shape descriptions all of the time. This is one reason why many current approaches to pattern classification in computer vision use algorithms based on the detection and matching (or indexing) of local image features, appearance-based feature hierarchies, or interest point operators (e.g., Lowe, 2004; Mikolajczyk & Schmid, 2005; Ullman, 2006; Ullman et al., 2002).

The aim of the current study was not to characterize eye movement patterns during everyday object recognition. Rather, like numerous studies of eye movements in other domains, our goal was to use fixation patterns as an index of information processing during shape perception—where the assumption is that observers fixate locations of high information content (e.g., Renninger et al., 2005). More specifically, our goal was to examine whether fixation patterns can be used to elucidate *local* shape analysis processes beyond those driven by low-level image statistics (e.g., simple contrasts in luminance, orientation, and color) in order to provide insights into the kinds of higher level shape information that supports shape perception. Some previous evidence using other tasks suggests that eye movement patterns are sensitive to 3D shape. In one study, initial localization saccades were compared when viewing 3D targets rendered with lighting and shadows or simple flat unicolor silhouettes (Vishwanath & Kowler, 2004). The results showed that saccades are sensitive to the 3D structure of an object: Although the 2D projection of the target to the retina in both conditions was the same, participants showed a bias toward the 2D COG when viewing silhouettes and the 3D COG when perceiving the target as a volume. In another study, Wexler and Ouarti (2008) have shown that saccadic eye movements during the spontaneous exploration of visual images follow surface depth gradients. Here, a key finding was that surface orientation alone has a large effect on eye movements independent of the task when looking at stimuli in 3D.

In the present study, eye movement patterns were recorded while observers either actively memorized subsets of 3D novel objects or passively viewed them in a pre-test phase, and then performed a recognition memory task. The observed fixation patterns were compared to the predicted distributions derived from different models of shape information content. Since our interest was to examine whether fixation patterns can be driven by

higher level shape features, beyond low-level image statistics alone, we used visual saliency as a baseline contrast (Itti, Koch, & Niebur, 1998; Koch & Ullman, 1985; Walther & Koch, 2006). The visual saliency model generates saliency maps based on weighted contrasts in luminance, orientation, and color. This model has been widely applied to eye movement studies of scene perception although its efficiency in predicting fixation patterns remains the subject of on-going debate (e.g., Baddeley & Tatler, 2006; Cristino & Baddeley, 2009; Henderson et al., 2007). The question of interest was whether specific models of shape analysis could account for fixation patterns beyond that explicable by visual saliency. Three different models were evaluated. *Model 1* was based on external global shape features defined by bounding contour. This hypothesis derives from recent work showing that outline shape influences object recognition (e.g., Hayward, 1998; Hayward et al., 1999; Lloyd-Jones & Luckhurst, 2002). *Model 2* and *Model 3* were derived from the large body of work highlighting the importance of curvature in shape perception (e.g., Attneave, 1954; Barenholtz, Cohen, Feldman, & Singh, 2003; Bertamini, 2008; Biederman, 1987; Cate & Behrmann, 2010; Cohen, Barenholtz, Singh, & Feldman, 2005; Cohen & Singh, 2007; De Winter & Wagemans, 2006; Feldman & Singh, 2005; Hoffman & Richards, 1984; Hoffman & Singh, 1997; Lim & Leek, in press). This work has shown highly robust perceptual sensitivity to curvature extrema, where negative minima define concave image regions (relative to the figure) and positive maxima define convexities—a phenomenon that has also recently been demonstrated in infants as young as 5 months old (Bhatt, Hayden, Reed, Bertin, & Joseph, 2006). Empirically, previous studies have largely examined curvature in the context of contour-defined 2D images such as polygons and line drawings (Cohen et al., 2005; Cohen & Singh, 2007; De Winter & Wagemans, 2006) in which curvature minima and maxima are defined along the occluding contour boundary. In contrast, there is relatively little data examining the role of curvature discontinuities defined by changes in the surface (rather than contour) curvature polarity of 3D objects. We examined two models of internal surface curvature defined by local internal convex curvature maxima (Model 2) and local internal concave minima (Model 3).

Methods

Participants

Sixty students from Bangor University (36 females, mean age = 20.83 years, *SD* = 4.33, 53 right-handed) participated in the study for course credit. All participants had normal or corrected-to-normal visual acuity. Informed consent was obtained from each participant prior to testing in line with local ethics committee and BPS guidelines and the Declaration of Helsinki.

Stimuli

There were 12 novel objects (see Figure 1a) each consisting of a unique spatial configuration of four volumetric parts. The parts were uniquely defined by variation among non-accidental properties (NAPs) comprising: edges (straight vs. curved), symmetry of the cross section, tapering (co-linearity), and aspect ratio (Biederman, 1987).

The object models were produced using Strata 3D CX software (Strata, USA) and rendered in Matlab using a single light source (top left) with anti-aliasing and scaled to fit within an 800 × 800 pixel frame (normalized in size across objects). All stimuli were uniformly colored in mustard yellow: R = 227, G = 190, B = 43. Stimuli subtended 18 degrees of visual angle horizontally with participants seated 60 cm from the display. This scale was chosen to induce saccadic exploration over the stimuli. Each stimulus was rendered depicting the object from six different viewpoints at successive 60-degree rotations in depth around a vertical axis perpendicular to the line of sight. The zero-degree viewpoint was a "canonical" three-quarter view (see Figure 1b). The 0-, 120-, and 240-degree versions served as familiar (pre-test) viewpoints, and the 60-, 180-, and 300-degree versions as novel viewpoints.

Apparatus

Eye movement data were recorded on a Tobii 1750 (Tobii Technology, AB, Sweden) binocular corneal reflection (CR)-based remote eye tracking system (<0.5-degree accuracy, 0.25-degree spatial resolution, and drift < 1 degree). Stimuli were presented on a TFT monitor running at a resolution of 1280 × 1024 pixels and 60-Hz refresh rate. Mean surround luminance was 114.7 cd/m^2 (*SD* = 0.25 cd/m^2) measured with a Minolta CS-100 photometer. A chin rest was used to stabilize the participants' head at a 60-cm viewing distance and a standard USB keyboard was used for response collection.

Design and procedure

Each participant initially completed a nine-point eye tracking calibration procedure. This required the participants to view a static blue dot that appeared, randomly, in each of 9 possible screen locations. Noisy calibration points were resampled algorithmically to ensure accuracy and validity.

The study comprised two phases: pre-test and test phases. All subjects completed both phases. There were
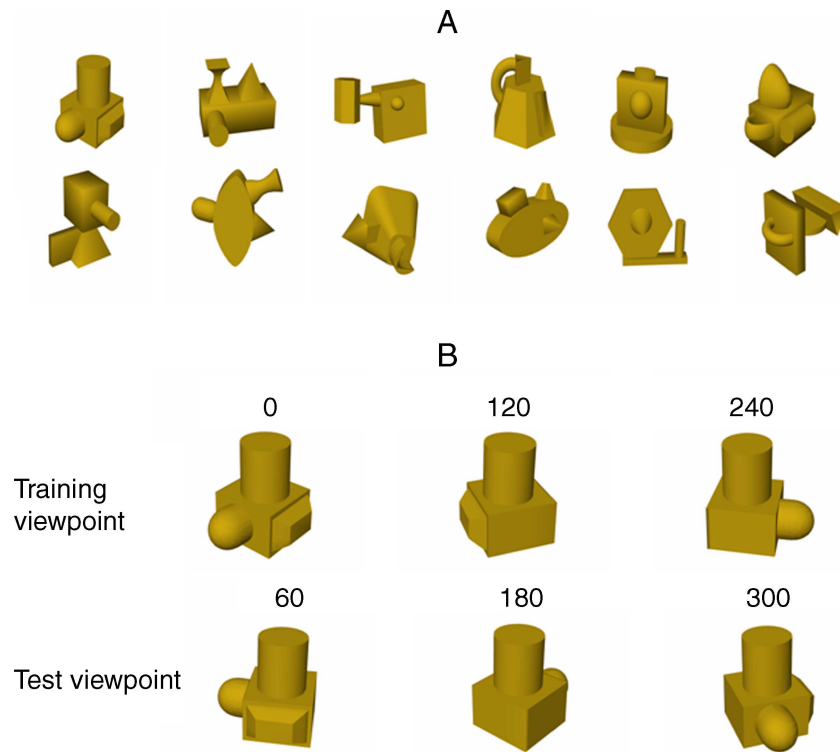
Figure 1. (a) The 12 novel object stimuli used in the current study. (b) An illustration of the three trained and three novel viewpoints used.

two versions of the pre-test: Active learning and passive viewing with participants assigned randomly to one of the two groups. For both pre-test groups, the trial structure was the same comprising 18 trials (6 targets × 3 viewpoints). On each trial, participants initially fixated a square (1° × 1° visual angle) for 2000 ms presented in the center of the display vertically and 9 degrees to either the left or to the right of the object. In the pre-test phase, following a 2000-ms blank ISI, a single stimulus was presented in the center of the monitor for 10 s. In the active learning group, participants were instructed to study the shape of each stimulus and to try to memorize it for a subsequent recognition memory task. They were told that they would see three different views of six objects. In the passive viewing group, participants were instructed only to visually inspect each stimulus. They were not told to memorize the objects, nor forewarned about the subsequent recognition memory task. For each pre-test group, half of the participants viewed Objects 1–6, and half viewed Objects 7–12. The objects viewed in the pre-test phase were assigned as targets. Thus, all 12 stimuli were used both as targets and distracters across groups. In the test phase (N trials = 72), targets (N = 6, depending on the set shown in pre-test) and distracters (N = 6) were presented in random order each at six viewpoints (3 familiar and 3 novel). Across groups there were 12 targets and 12 distracters (each shown from six viewpoints). The trial structure was the same as in the pre-test phases, except that the stimuli were presented until the participants made a keyboard response. Both pre-test groups were given the same instructions in the test phase. They were asked to determine and respond via a key press

(k—"yes"/d—"no") whether the presented stimulus was one of the objects viewed during the pre-test phase regardless of the viewpoint shown. Eye movement data, response time (RT), and accuracy were recorded as dependent measures. The experiment lasted approximately 30 min (including calibration).

Analyses of eye movement data

Fixations were defined as eye movements that remain within the same circular region of diameter 60 pixels (2° visual angle given a viewing distance of 60 cm, a screen resolution of 1280 × 1024 pixels, and a horizontal screen size of 34 cm) for at least 100 ms (e.g., Manor & Gordon, 2003). In addition, for each trial, the first fixation following stimulus onset was discarded in order to eliminate early object localization fixations associated with COG effects that have been shown to be sensitive to global object shape (e.g., Denisova, Singh, & Kowler, 2006; He & Kowler, 1989; Melcher & Kowler, 1999).

The aims of the analysis of eye movement data were twofold: First, rather than attempting to pre-define local areas of interest (AOIs) as is often done in other domains of eye movement research (e.g., van Gompel, Fischer, Murray, & Hill, 2007), our goal was to use the observed distributions of fixations themselves to empirically define AOIs for each stimulus (Johnston & Leek, 2009). These empirically defined AOIs naturally incorporate measures of error (e.g., noise arising from eye tracker accuracy, drift, and spatial resolution), as well as within- and between-subject

variability. Second, the empirically defined AOIs were then subjected to an analysis for shape information content by assessing the degree of fit between the observed distributions of AOIs and the distributions predicted by different algorithmically generated models of local shape information (see below).

The empirical derivation of the AOIs (or AOI region maps) and their quantitative comparison to the predicted distributions were achieved using a modified version of the Fixation Region Overlap Analysis (FROA) methodology (see Johnston & Leek, 2009, for a full description and Matlab implementation of the FROA method). In brief, pre-processed filtered gaze data were used to compute global fixation frequency maps for each target and trained viewpoint in the pre-test phases ($N = 36$, 12 stimuli $\times$ 3 viewpoints) and for both targets and non-targets at familiar and test viewpoints in the test phase ($N = 144$, 12 targets $\times$ 6 viewpoints + 12 non-targets $\times$ 6 viewpoints). The maps for each stimulus were created by summing the convolution of each fixation map (summed across subjects) with a 2D Gaussian kernel ($SD = 0.5$ deg). Since fixation frequency varies across subjects and conditions, the maps were normalized using z scores. The AOI region maps were derived by binary thresholding the fixation frequency distributions using a fixed parameter across all conditions. Here, the threshold was set to $z = 1.2$ in order to ensure that thresholded region maps for the fixation data were approximately equivalent in size to those derived from the models.[1] These binary AOI region maps formed the basis for the subsequent analysis of the pre-test and test phase fixation data. The primary dependent measure in FROA is spatial (i.e., area) overlap percentage (e.g., the amount of area overlap in the binary region maps for each stimulus and the predicted distribution of AOI regions for each theoretical model of shape information normalized by the size of the binary region maps for each stimulus—see Johnston & Leek, 2009). Overlap is determined by calculating the number of suprathreshold pixels that occur at the same spatial locations in the binary fixation region maps of each contrasted (observed versus modeled) image set. The statistical significance of the observed overlap percentage between data sets is then determined with reference to bootstrapped probability distributions derived from Monte Carlo simulations. These are used to generate the expected random frequency distribution of area overlap percentage for a given observed, and modeled, fixation region. This technique provides a method for estimating the random distribution of overlap that would be expected for fixation regions of the observed shape and size (area) and that is constrained to fall within locations bounded by the perimeter (occluding contour) of the original stimulus. It is important to note that this method thus controls for differences in the area of the respective region maps (and specific threshold parameters) in any set of contrasted images. Statistical analyses were conducted across objects (items) and across subjects. The statistical significance of the fit between the observed fixation data and each model prediction was calculated as follows:

Step 1: We compute the "Actual Overlap Percentage" (AOP) between the binary images of the observed thresholded region maps and a given model is calculated for each stimulus. This is computed as a percentage of the total region area in observed thresholded region map (i.e., 0% if the model did not overlap at all with the observed fixation map or 100% if the model overlaps completely with the observed fixation map).

Step 2: For each stimulus, we calculate the "Chance Overlap Percentage" (COP) that corresponds to the percentage overlap we would expect at the 95% CI of a random distribution of observed fixation data–model overlap. The random distribution is computed using a Monte Carlo procedure that is run separately for each stimulus and data–model contrast. This is done by taking the thresholded AOIs corresponding to the observed eye movement data and randomly relocating them within the boundary of the stimulus. Following each random relocation, overlap between the (random) fixation AOIs and model (predicted) regions is computed. This process is repeated 1000 times per item to generate individual distributions of random overlap that are specific to each stimulus and data–model contrast.

Step 3: In order to then compare the degree of observed fixation data–model correspondence, we compute a measure called "Model Matching Correspondence" (MMC): $MMC_{Mx} = AOP_{Mx} - COP_{Mx}$ (where Mx is a given model). As COP and AOP are expressed in percentages of the total region area per item, the distance measure is normalized for variation in thresholded region size across items. Higher values of MMC_{Mx} indicate better correspondence between the tested model and the observed fixation data.

As noted earlier, an important goal in assessing the relative differences in fixation data–model correspondences was to factor out fixations that could potentially be driven solely by low-level image statistics. To this end, we computed a model based on visual saliency that served as a baseline contrast. To generate the visual saliency maps for the baseline contrast, we used the visual saliency algorithm (Itti et al., 1998) to compute a baseline model. The visual saliency baseline was computed using the Matlab Saliency Toolbox implementation (Walther & Koch, 2006). The model was run on each of the 72 stimulus images (12 objects $\times$ 6 viewpoints) used in the recognition task to generate a saliency map for each stimulus. The output of the toolbox is a list of saliency values for each pixel, which are grouped into a saliency region map using a shape estimation function (see Walther & Koch, 2006). The default model parameters were used. The number of saliency regions generated was constrained to approximate the area and number of thresholded regions generated for the other models. The saliency maps were thresholded and binarized using FROA (see Figure 2). These maps represent the thresholded distributions of fixation regions we would expect if eye movements were determined solely by low-level image statistics, that is, by
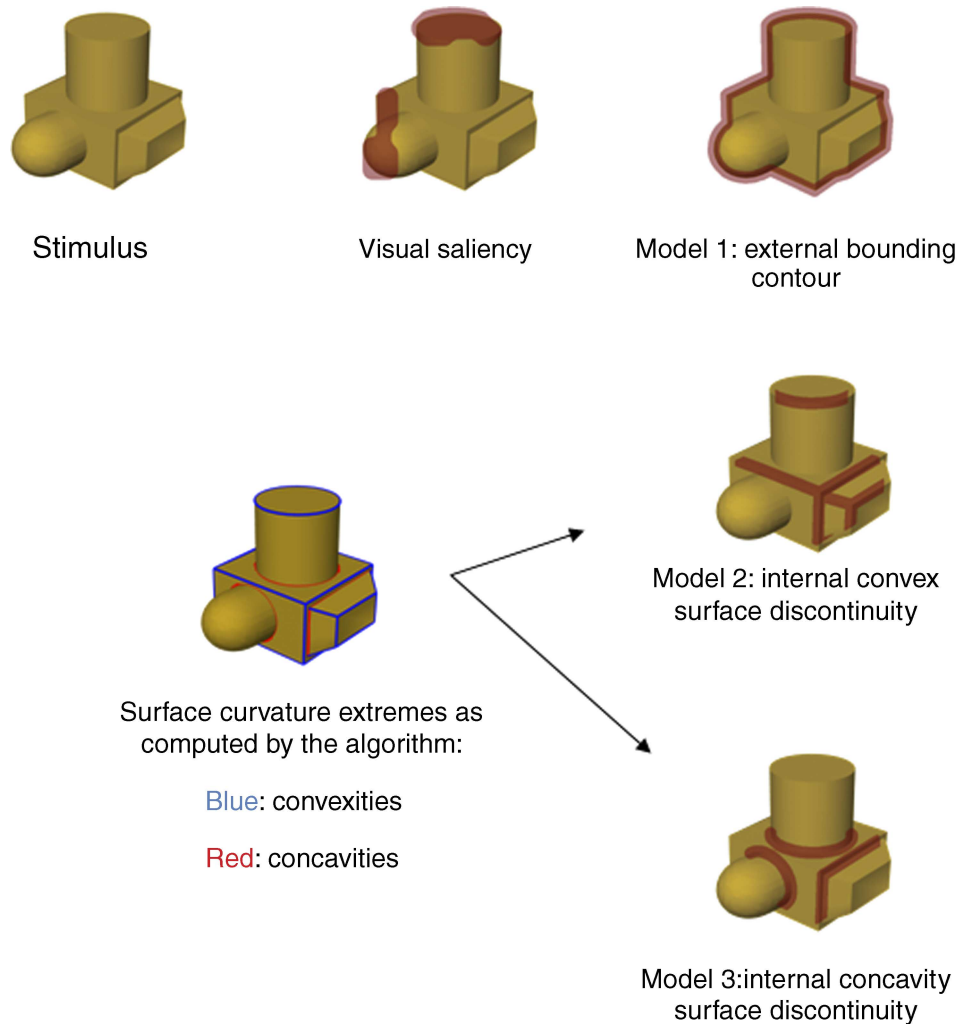
Figure 2. An illustration of the predicted thresholded fixation region maps for the tested models. All of these predicted distributions were generated algorithmically.

the most visually salient image regions defined by intensity contrast, orientation, and color.

Step 4: These saliency maps are used to compute the baseline measure MMC_{VS} for each item: $MMC_{VS} = AOP_{VS} − COP_{VS}$ (where VS = visual saliency). COP_{VS} is estimated using the same Monte Carlo procedure described above. This shows the extent to which observed overlap is greater or less than the 95% CI of the random distribution of the saliency algorithm.

Step 5: Finally, for each model we subtract out the baseline overlap attributable to visual saliency as follows: $MMC_{Mx} − MMC_{VS}$. This shows the difference in overlap between the fixation data and a given model relative to the visual saliency baseline. A positive value here indicates a higher fixation data–model correspondence than that accounted for by visual saliency. In contrast, a negative value would indicate a lower fixation data–model correspondence than that accounted for by visual saliency.

The resulting MMC statistics are then subjected to analyses of variance (ANOVA) across models. In addition, we also examined the generality and robustness of the observed patterns across subjects. The subject analysis was done by contrasting the mean normalized fixation frequency (fm) for thresholded and subthresholded object regions across subjects. Here, thresholded regions corresponded to the AOIs defined by FROA. Subthresholded regions were defined by subtracting the thresholded AOIs from the remaining area of each stimulus image (within the bounding contour). We refer to these regions as subthreshold AOIs. The fixation frequency distributions per subject were normalized for mean region area (across items) and converted to units of visual angle (30 pixels is equal to 1 degree of visual angle, given a viewing distance of 60 cm, a screen resolution of 1280 × 1024 pixels, and screen size of 34 cm). Thus, this measure takes account of differences in pixel area between thresholded and nonthresholded regions. Subject analyses of mean fixation

durations for thresholded versus subthreshold AOIs using the same normalized measures are also reported.

Statistical significance is assessed relative to the two-tailed a priori alpha level (p = 0.05), unless where otherwise stated. Exact probability values are reported ($p = x$) except where $p < 0.0001$.

Generating model predictions

The predicted distributions for each model of image information content were algorithmically computed from the 3D object models using Matlab.

Model 1: External (bounding) contour

Model 1 examined the extent to which fixation patterns focus on external global shape features defined by bounding contour. This hypothesis derives from previous work showing that outline shape influences object recognition (e.g., Hayward, 1998; Hayward et al., 1999; Lloyd-Jones & Luckhurst, 2002). The bounding contour was computed using an edge detector on the image silhouette of the stimuli. It was then replotted using lines of 0.66-degree width (see Figure 2). This value was used as it produced models of a similar size as the binarized eye movement data.

Model 2: Internal convex surface discontinuity

Model 2 generated predicted fixation regions based on the locations of local features defined by *convex* surface curvature maxima. These were generated by applying a curvature estimation algorithm derived from Taubin (1995) to the object mesh models using the Peyre Matlab toolbox. From this, we extracted edges along convex curvature maxima (see Figure 2). The convex features were replotted using lines of 0.66-degree width. Edges on the exterior bounding contour were deleted. Due to the nature of our stimuli, convexities can occur both inside and on the bounding contour of an object, but concavities are more likely to occur on the internal contour (see Figure 2). By keeping internal features only, we are able to compute a bias-free measure of the preference for convex or concave image features.

Model 3: Internal concave surface discontinuity

Model 3 generated predicted fixation regions based on the locations of local features defined by *concave* surface curvature minima (see Figure 2). The same curvature estimation method was used as for Model 2, except that here we extracted edges along concave curvature minima. As with Model 2, edges falling on external bounding contour were removed.

Results

Analyses of behavioral data (test phase)

Analyses were conducted on the mean median test phase RTs (correct responses only) and accuracy data. Only RTs for correct responses were included. Mean median RTs and accuracy rates are shown in Table 1 (targets only) for both the active learning and passive viewing pre-test groups.

RTs

A 2 (pre-test task: active learning vs. passive view) × 2 (viewpoint: familiar vs. novel) × 2 (stimulus type: target versus non-target) mixed factor ANOVA showed a significant main effect of pre-test task, $F(1, 35) = 13.65$, $p = 0.001$, $\eta^2 = 0.281$, and a significant interaction between viewpoint and stimulus type, $F(1, 35) = 6.86$, $p = 0.013$, $\eta^2 = 0.164$. There were no interactions involving the factor of pre-test group. As seen in Table 1, these results indicate that RTs were faster overall in the test phase for the active learning than the passive viewing pre-test group. Post hoc planned comparisons showed that target RTs were faster for familiar than novel viewpoints in both the active learning, $t(71) = -2.485$, $p = 0.018$, and passive viewing pre-test groups, $t(71) = -3.119$, $p = 0.003$. In contrast, there was no difference in mean median RTs between familiar ($M = 1653.56$ ms; $SE = 66.47$ ms) and novel views ($M = 1570.56$ ms; $SE = 60.15$ ms) for non-targets ($t(71) = -1.37$, $p = 0.17$, ns).

	Pre-test group			
	Active learning		Passive viewing	
	RTs (ms)	% Correct	RTs (ms)	% Correct
Familiar views	1323.76 (51.34)	93 (1.4)	1499.08 (63.66)	74 (0.3)
Novel views	1501.22 (102.78)	87 (1.8)	1695.28 (112.98)	69 (0.3)

Table 1. The mean median RTs and accuracy rates (targets) for familiar and novel viewpoints in the test phase. Standard error of the mean is shown in parentheses.

| | Pre-test phase | Test phase | |
	Targets	Targets	Non-targets
Thresholded AOIs	0.52 (0.03)	0.38 (0.04)	0.39 (0.03)
Subthreshold AOIs	0.001 (0.0001)	0.03 (0.002)	0.03 (0.003)

Table 2. The mean normalized fixation frequencies (mean fixation per degree of visual angle) for thresholded and subthreshold AOIs for the active learning group in both the pre-test and test phases. Standard error of the mean is shown in parentheses.

Accuracy rates

Accuracy data were analyzed using non-parametric significance tests for either related groups (Wilcoxon) unless otherwise stated. For the active learning pre-test group, there was no significant difference between accuracy rates for targets ($M = 90\%$; $SE = 0.012$) versus non-targets ($M = 90\%$; $SE = 0.013$), $z = 0.226$, $p = 0.821$, ns. For test phase target trials, accuracy was significantly higher for trained views ($M = 93\%$; $SE = 0.014$) than for novel views ($M = 87\%$; $SE = 0.018$), $z = -2.93$, $p = 0.003$. There was no significant difference in accuracy for non-targets across viewpoints ($z = -1.47$, $p = 0.141$, ns). For the passive viewing pre-test group, overall response accuracy was higher for non-targets ($M = 81.5\%$; $SE = 0.015$) than targets ($M = 72\%$; $SE = 0.021$), $z = 3.21$, $p = 0.001$. In the test phase, there was no difference in accuracy for targets between familiar ($M = 74\%$; $SE = 0.031$) and novel viewpoints ($M = 69\%$; $SE = 0.029$), $z = -1.77$, $p = 0.77$, ns, or for non-targets: familiar, $M = 84\%$; $SE = 0.016$; novel, $M = 87\%$; $SE = 0.014$ ($z = -1.75$, $p = 0.79$, ns). Overall accuracy rates for the active learning group ($M = 90\%$, $SE = 0.01$) and passive viewing group ($M = 82\%$, $SE = 0.01$) were significantly different (Mann–Whitney: $z = -4.37$, $p < 0.0001$).

Analyses of eye movement data
Pre-test phase: Active learning group

A subject analysis was first performed to test the generality and reliability of the thresholded fixation region distributions across participants. This was done by contrasting the frequency of fixations between thresholded and subthreshold AOIs (see Methods section). Separate subject analyses were performed on the pre-test (targets) and test phase (targets and non-targets) data. Table 2 shows the mean normalized frequencies for the thresholded and subthreshold AOIs across participants. These data show that the mean normalized fixation frequency for thresholded AOIs is higher than for subthreshold AOIs in both the active learning and test phases for targets and in the test phase for targets and non-targets.

For the active learning phase, there was a significant difference between the mean normalized fixation frequencies across participants for the thresholded vs. subthreshold

AOIs, $t(29) = 16.95$, $p < 0.0001$. For the test phase, a 2 (AOI: thresholded vs. subthreshold) × 2 (stimulus: target vs. non-target) repeated measures ANOVA showed a significant main effect of AOI, $F(1, 29) = 109.22$, $p < 0.0001$, $\eta^2 = 0.790$, but no other main effects or interactions. These analyses show that the fixation regions identified using FROA were robust across subjects for the active learning group.

Analyses of the local shape feature analysis patterns (pre-test, active learning task)

The remaining analyses of the fixation data for the active learning group were computed across items. For the pre-test phase, the distributions of fixation regions to targets presented at trained viewpoints ($N = 36$) were initially analyzed across 3 epochs allowing us to compare the spatial distributions of fixations occurring at different time periods following stimulus onset. To do this, fixations were divided subject by subject and trial by trial into bins containing the first third, middle third, and final third (e.g., for a particular subject making 9 fixations on a given item, fixations 1–3 would be allocated to the first bin, 4–6 to the second bin, and 7–9 to the third bin). The respective bins were then pooled across subjects for each stimulus. A 3 (epoch) × 4 (Models 1–3, plus the baseline saliency model) repeated measures ANOVA on the MMC distance measure across targets showed a significant main effect of model, $F(3, 105) = 13.33$, $p < 0.0001$, $\eta^2 = 0.276$, but no main effect of epoch ($F = (6, 210) = 0.737$, $p = 0.621$) and no significant interaction. In the absence of an interaction, the MMC distance statistics were collapsed across epoch. A one-way ANOVA across models on the MMC measure was significant, $F(3, 140) = 10.08$, $p < 0.0001$. Subsequent post hoc analyses using the Bonferroni test showed that the pairwise contrasts between models were significantly different for internal features concave vs. visual saliency, $p < 0.0001$; internal features convex vs. visual saliency, $p < 0.0001$; external features vs. visual saliency, $p = 0.029$. There were no other significant differences. These analyses show that the fixation data–model correspondence is greater for all three models of shape analysis than the baseline saliency model. However, there were no differences in the degree of the data–model
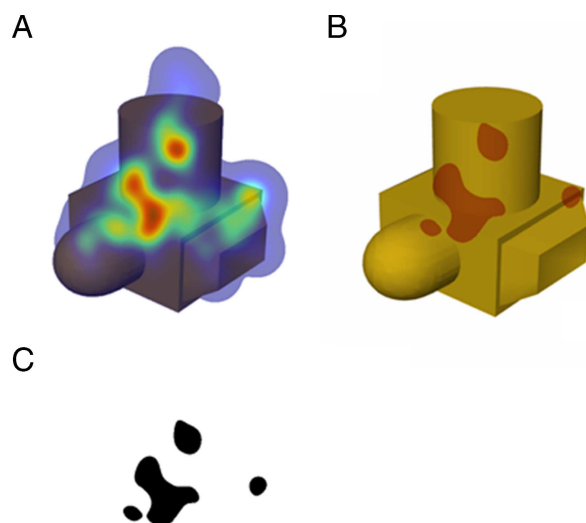
Figure 3. Illustrative visualization of the primary steps used to derive the binary region maps underlying FROA. (a) *Z*-scored heat map made with a Gaussian kernel of 4 degrees from the fixation frequencies overlaid on the object. (b) Thresholded map (*z* = 1.2) overlaid on the object. (c) Binary thresholded region map computed by FROA.

correspondence between the models of shape analysis in the pre-test phase (Figure 3).

Pre-test phase: Passive viewing group

An initial analysis by subjects was undertaken contrasting the frequency of fixations between thresholded and subthreshold AOIs using normalized frequency statistics. Table 3 shows the mean normalized frequencies for the thresholded and subthreshold AOIs across participants. These data show that the mean normalized fixation frequency for thresholded AOIs is higher than for subthreshold AOIs in both the passive viewing and test phases for targets and for targets and non-targets in the test phase.

There was a significant difference between the mean normalized fixation frequencies across participants for the thresholded versus subthreshold AOIs, $t(29) = 15.63$, $p < 0.0001$. For the test phase, a 2 (AOI: thresholded vs. subthreshold) $\times$ 2 (stimulus: target vs. non-target) repeated measures ANOVA showed a significant main effect of AOI, $F(1, 29) = 103.89$, $p < 0.0001$, $\eta^2 = 0.782$, but no other main effects or interactions.

Analyses of the local shape feature patterns (pre-test, passive viewing task)

As previously, the distributions of fixation regions to targets presented at trained viewpoints ($N = 36$) were analyzed across 3 epochs following stimulus onset. At each epoch, FROA was used to compute the observed overlap between the gaze data and each model prediction relative to the random Monte Carlo distribution. A 3 (epoch) $\times$ 4 (Models 1–3, plus the baseline saliency model) repeated measures ANOVA on the MMC distance measure across targets showed a significant main effect of model, $F(3, 105) = 25.66$, $p < 0.0001$, $\eta^2 = 0.423$, but no effect of epoch ($F = (2, 70) = 1.47$, $p = 0.235$ ns) and no interaction. In the absence of any interaction, normalized distance was collapsed across epoch. A one-way ANOVA on mean MMC values across models (visual saliency, internal features convex, internal features concave, external features) was significant, $F(3, 140) = 15.72$, $p < 0.0001$. Post hoc analyses showed that the pairwise contrasts between models were significantly different for internal features concave vs. visual saliency, $p < 0.0001$; internal features convex vs. visual saliency $p < 0.0001$; external features vs. visual saliency $p = 0.007$. In addition, unlike for the active learning group, there was also a significant difference between internal features concave vs. external features, $p = 0.020$. There were no other significant contrasts (Figure 4).

Test phase: Active learning and passive viewing groups

These analyses were run on the MMC data from the test phase. Figure 5 shows the mean $MMC_{Mx} - MMC_{VS}$ contrasts across models. A 2 (group: active learning vs. passive viewing) $\times$ 2 (stimulus type: target vs. non-target) $\times$ 4 (model: Models 1–3, plus the baseline saliency model) mixed design ANOVA showed a significant main effect of model, $F(3, 138) = 39.08$, $p < 0.0001$, $\eta^2 = 0.459$. There were no other significant main effects or interactions. Post hoc analyses showed that the pairwise contrasts (Bonferroni) between models were significantly different for all model–visual saliency baseline contrasts: external features vs. visual saliency, $p < 0.0001$; internal features concave vs. visual saliency, $p < 0.0001$; internal features convex vs. visual saliency, $p < 0.0001$. In addition, mean fixation data–model correspondence was higher for the internal concave
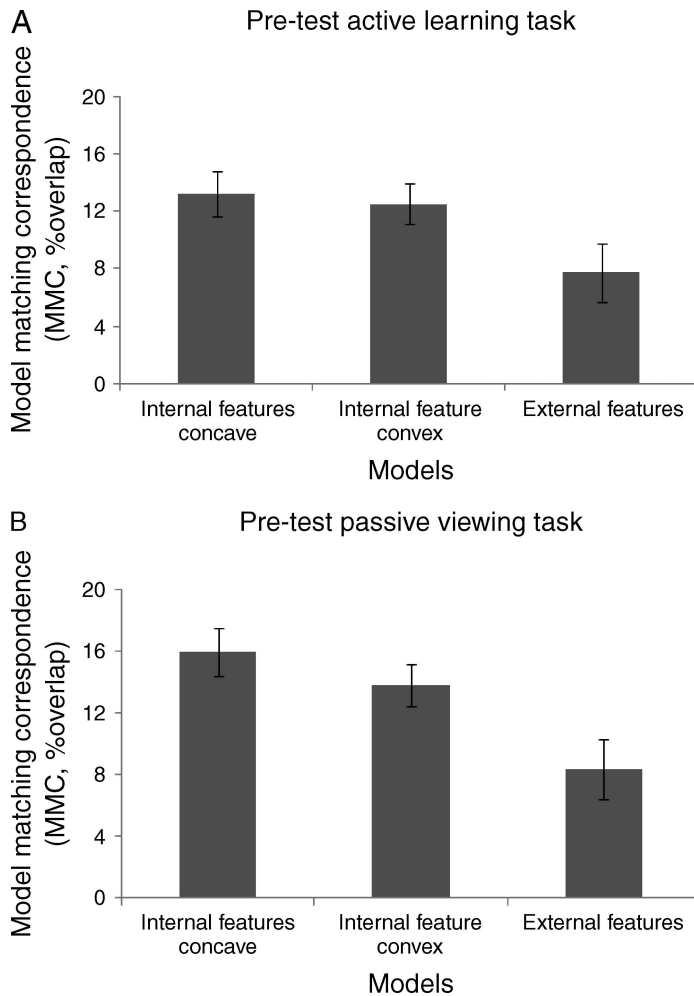
| | Pre-test phase | Test phase | |
	Targets	Targets	Non-targets
Thresholded AOIs	0.49 (0.03)	0.42 (0.05)	0.43 (0.04)
Subthreshold AOIs	0.0014 (.001)	0.03 (0.002)	0.03 (0.002)

Table 3. The mean normalized fixation frequencies (mean fixation per degree of visual angle) for thresholded and subthreshold AOIs for the passive viewing group in both the pre-test and test phases. Standard error of the mean is shown in parentheses.

A

Pre-test active learning task

B

Pre-test passive viewing task

Figure 4. Mean MMC (MMC_{Mx} − MMC_{VS}) measure of data–model correspondences between models (relative to the visual saliency baseline) for (a) pre-test active learning group and (b) pre-test passive viewing group. Bars show standard error of the mean (% overlap).

vs. internal convex contrast, $p = 0.024$; internal features concave vs. external features, $p = 0.001$; and internal features convex vs. external features, $p = 0.030$.

Analysis of fixation duration

We also conducted analyses of fixation duration, contrasting mean durations for fixations falling within the FROA-defined thresholded regions versus duration for fixations falling outside of the thresholded regions (subthreshold AOI). Separate analyses were conducted for the active learning and passive viewing groups and for the pre-test and test phases.

Active learning group

For the pre-test, mean fixation durations were longer for fixations within thresholded AOIs ($M = 295.59$ ms, $SE = 19.76$ ms) than for those within subthreshold AOIs ($M = 266.58$ ms, $SE = 16.16$ ms). This difference was

significant, $t(29) = 3.35$, $p = 0.002$. For the test phase target, mean fixation duration was also longer for thresholded ($M = 259.19$ ms, $SE = 15.88$ ms) than subthreshold AOI fixations ($M = 223.25$ ms, $SE = 10.77$ ms). The same pattern was found for non-targets: thresholded AOI fixations ($M = 252.95$ ms, $SE = 14.30$ ms) and subthreshold AOI fixations ($M = 226.26$ ms, $SE = 10.86$ ms). A 2 (AOI: threshold vs. subthreshold) × 2 (stimulus: target vs. non-target) repeated measures ANOVA on the test phase mean fixation duration data showed a significant main effect of AOI, $F(1, 29) = 22.52$, $p < 0.0001$, $\eta^2 = 0.437$, but no other main effects or interactions.

Passive viewing group

Analyses of the pre-test phase fixation duration data showed longer mean durations for the thresholded AOI fixations ($M = 313.76$ ms, $SE = 31.67$ ms) than for subthreshold AOI fixations ($M = 275.87$ ms, $SE = 23.03$ ms). This difference was significant, $t(29) = 3.17$, $p = 0.004$. For the test phase target, mean fixation duration was also longer for thresholded AOI fixations ($M = 266.39$ ms, $SE = 23.30$ ms) than for subthreshold AOI fixations ($M = 239.75$ ms, $SE = 16.48$ ms). The same pattern was found for non-targets: thresholded ($M = 265.01$ ms, $SE = 22.97$ ms) and subthreshold fixations ($M = 236.03$ ms, $SE = 16.54$ ms). A 2 (AOI: threshold vs. subthreshold) × 2 (stimulus: target vs. non-target) repeated measures ANOVA on the test phase duration data showed a significant main effect of AOI, $F(1, 29) = 10.77$, $p = 0.003$, $\eta^2 = 0.271$, but no other main effects or interactions. These analyses show that for both the active learning and passive viewing groups mean fixation durations were longer for fixations falling within thresholded regions than for those outside of those regions in both the pre-test and test phases of the study.
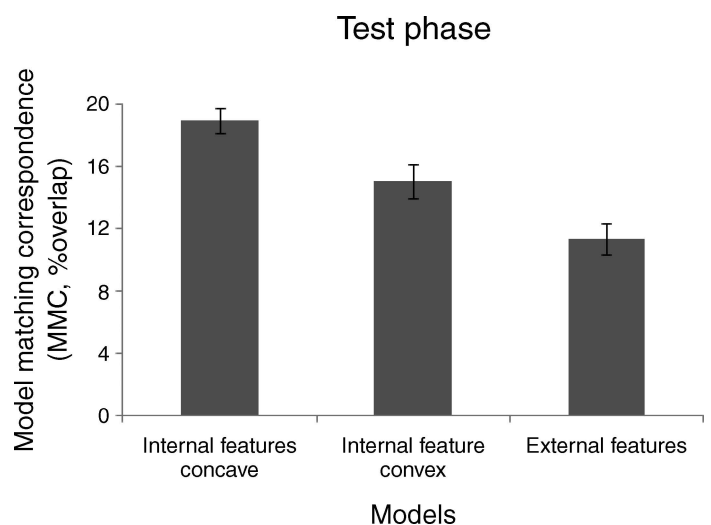
Test phase

Figure 5. Mean MMC (MMC_{Mx} − MMC_{VS}) measure of data–model correspondences between models (relative to the visual saliency baseline) for the recognition memory test phase (collapsed across pre-test groups). Bars show standard error of the mean (% overlap).

Active viewing vs. passive learning

Finally, we also contrasted mean fixation durations for thresholded and subthreshold fixations across task groups on the pre-test phase data. A 2 (task: active vs. passive) × 2 (AOI: threshold vs. subthreshold) mixed ANOVA showed a significant main effect of AOI, $F(1, 58) = 20.54$, $p < 0.0001$, $\eta^2 = 0.262$, but no other main effects or interactions. This suggests that mean durations were not significantly different between pre-test task groups.

General discussion

This study provides some of the first evidence from measures of fixation patterns regarding the acquisition of higher level shape information during the perception and recognition of 3D objects. In a pre-test phase, observers either actively memorized or passively viewed sets of visually similar novel objects prior to performing a recognition memory test. The main empirical results were given as follows: First, the analyses of the RT and accuracy data showed that while observers performed the recognition memory task more accurately following the active learning than passive viewing pre-test, the patterns of test phase RTs for both groups showed faster responses for targets presented at familiar (pre-test) viewpoints than at novel viewpoints. This suggests that: (1) participants in the active learning and passive viewing pre-test groups performed the recognition memory task in a similar way and (2) that recognition in both groups was viewpoint-dependent—consistent with other reports in the literature that recognition is mediated by viewpoint-dependent representations of object shape (e.g., Bülthoff & Edelman, 1992; Edelman & Weinshall, 1991; Riesenhuber & Poggio, 1999; Tarr & Bülthoff, 1998; Ullman, 1998). Second, analyses of the fixation data showed a strikingly consistent pattern of data–model correspondences across tasks. In particular, during both active learning and passive viewing pre-test phases, and during the recognition memory task, we found evidence that fixation patterns are not driven solely by regions of visual saliency defined by contrast in low-level image properties but rather that observers fixate regions containing higher level shape information defined either by external bounding contour or by internal regions of convex or concave surface discontinuity. Third, despite the similarity of the patterns of data–model correspondences between the active learning and passive viewing groups, the distributions of fixations across object shape features differed between the study and test phases: Notably, during the recognition task, we found a preference for fixation at internal regions of surface concavity.

These findings are consistent with previous studies demonstrating the importance of curvature singularities in the visual perception of shape (e.g., Attneave, 1954; Barenholtz et al., 2003; De Winter & Wagemans, 2006; Feldman & Singh, 2005; Hoffman & Richards, 1984; Hoffman & Singh, 1997)—although they are, to our knowledge, the first data from 3D object perception and recognition showing a preference for fixation at these regions in both active and passive viewing tasks. Further, the finding of a preference for fixation at regions of concave surface discontinuity during the recognition task provides new evidence for a direct link between the encoding of information about surface concavity and object recognition. There are two important issues raised here. The first concerns the apparent preference for fixation at regions of surface concavity during the recognition task. The second concerns our observation of similar fixation distributions and, by hypothesis, similar perceptual strategies for the acquisition of shape information, across active and passive viewing tasks. We discuss both of these issues in turn.

Eye movements, surface curvature, and recognition

In other domains, such as scene perception, there is ongoing debate about the relative influence of bottom-up, stimulus-driven factors and top-down, conceptually driven factors in determining eye movement behavior (e.g., Foulsham & Underwood, 2007; Henderson et al., 2007; Itti et al., 1998). Our data show that fixation patterns during the perception and recognition of object shapes cannot be solely accounted for by low-level visual saliency.[2] Moreover, the data further showed a fixation preference for concave regions over convex regions during recognition. This concave preference did not interact with pre-test group. That is, regardless of whether observers actively memorized or passively viewed objects in the pre-test, they showed a preference for fixation at regions of internal concave minima in the recognition task. How can this pattern of results be accounted for?

One possibility is that observers specifically fixate those particular internal regions because they are the optimal locations for extracting global (e.g., outline) shape properties rather than because of their status as regions containing perceptually relevant shape curvature. However, such an account would not provide an obvious explanation for the apparent preference for fixation at regions of concave surface discontinuity in the recognition task but not in the pre-test phase. Additionally, it is more likely that the optimum location for extracting global shape attributes (e.g., elongation, orientation, or symmetry) would be close to the center of mass—but this is clearly not the case as early COG fixations were removed from the data.

Rather, our finding of a preference for fixation at regions of concavity during the recognition task is consistent with hypotheses that outline a special functional status for concave minima in shape recognition (e.g., Feldman &

Singh, 2005; Hoffman & Richards, 1984; Lim & Leek, in press). One influential hypothesis is that concave regions play an important role as segmentation points allowing for the computation of parts-based structural descriptions (e.g., Hoffman & Richards, 1984; Marr & Nishihara, 1978). In this context, one interesting aspect of the data stems from the concurrent observation of a fixation preference for concave surface minima along with viewpoint-dependent performance in the recognition task. The former finding is consistent with the claim that negative curvature minima play a functional role in part segmentation during the derivation of a structural description representation (e.g., Biederman, 1987; Hoffman & Richards, 1984; Marr & Nishihara, 1978), while the latter finding, according to some interpretations of viewpoint-dependent effects, is consistent with image-based view interpolation models (e.g., Bülthoff & Edelman, 1992; Edelman & Weinshall, 1991; Riesenhuber & Poggio, 1999; Tarr & Bülthoff, 1998; Ullman, 1998).

How might these two findings be reconciled? One possibility is that they reflect different stages of object processing within the context of more recent hybrid models of object recognition, which propose the use of both structural description and image-based representations (e.g., Foster & Gilson, 2002; Hummel & Stankiewicz, 1996). Alternatively, within an exclusively image-based approach, one could suppose that the apparent preference for fixation at regions of surface curvature concavity reflects the encoding of local depth information in image-based object representations. Some supporting evidence comes from the recent demonstration by Wexler and Ouarti (2008) showing that saccadic eye movements during the spontaneous exploration of visual images follow surface depth gradients. Thus, these findings present a challenge to image-based models that are based solely on the use of 2D image properties (e.g., Bülthoff & Edelman, 1992) and appear to necessitate, within this theoretical framework, the encoding and use of image features that specify local surface depth information.

Task generality of shape analysis patterns

A further aspect of the results that is of theoretical interest is the consistency of the patterns of data–model correspondences across the active learning and passive viewing tasks. This is perhaps surprising given that one might expect task requirements to affect the perceptual analysis of shape. Here, despite the fact that one group of observers were explicitly told to memorize shape for a subsequent recognition task, the perceptual analysis strategies of the two groups, as evidenced by the patterns of data–model correspondences, were similar. One implication of this finding is that local shape analysis strategies during perception are "hard-wired" in the sense of being invariant to task requirements—at least across the range of tasks tested here. This hypothesis is intuitively appealing

in that during everyday recognition observers cannot entirely predict when unfamiliar objects might become relevant to their immediate or future goals and intentions. However, it remains to be determined whether the observed patterns of shape analyses found here will generalize across other tasks, including, for example, those related to the computation of shape representations for reaching and grasping (e.g., Land et al., 1999).

Caveats and future directions

The findings reported in this study demonstrate the considerable potential for quantitative analyses of fixation patterns to elucidate local shape feature processing during object shape perception and recognition. It is also important to note that the results are limited in several ways. We do not claim that these fixation patterns necessarily characterize the role of eye movements during object recognition in natural environments (e.g., Land et al., 1999; Tatler, Gilchrist, & Rusted, 2003). Indeed, objects may often be identified rapidly within a single fixation depending on their scale (e.g., Schendan & Kutas, 2003), and this most certainly involves sensory input from both foveal and parafoveal vision (Henderson et al., 2003; van Gompel et al., 2007). Rather, our goal was to use fixation patterns as a way to infer information acquisition during shape perception and to see whether eye movements can be used to elucidate object recognition mechanisms in this way. In future studies, it will be important to establish the robustness of these patterns under different conditions, through variation in stimuli (e.g., similarity, complexity, configuration, and part structure), display (e.g., stereo versus mono presentation), and tasks (subordinate vs. basic-level classification).

Conclusion

In summary, the current study used fixational eye movement patterns to examine local shape analysis processes during the perception and recognition of object shape. The key finding was a preference for fixation at regions of concavity surface intersection during the recognition task. This pattern was found regardless of whether observers had actively learned or passively viewed stimuli in the pre-test phase. This finding is consistent with other evidence highlighting the special functional status of concave regions in shape recognition.

Acknowledgments

Society, Houston, November 2006; Experimental Psychology Society Meeting, Plymouth, UK, July, 2006; and Vision Science Society Meeting, Sarasota, May 2008. The authors would like to thank Pamela Arnold, Shelly Kemp, Richard Carvey, and Neil Barker for their contribution to the study.

Footnotes

[1]While the chosen threshold level will determine region size, this is also accounted for in the Monte Carlo procedure that is used in FROA to compute the random distribution of fixation region overlap (see below). It should be noted further that while the specific threshold level used in FROA cannot be rigorously defined a priori, it is set at a level that allows both for spatial precision and reduces the likelihood of missing theoretically relevant "subthreshold" image regions—similar to the issues surrounding the adoption of a given alpha level in statistical testing (see also Johnston & Leek, 2009).

[2]It is important to note that the goal of this study was *not* to evaluate the saliency algorithm but rather to examine to extent to which fixation patterns are predicted by a restricted set of shape feature models over and above what can be accounted for by low-level image saliency. Hence, we used saliency as the baseline contrast only. One might argue that the parameters of the saliency algorithm ought to be tuned to a given task. However, a key issue is to what parameter settings? In addition, how can this be determined a priori? It is an important issue for future work to consider whether saliency models need to incorporate other kinds of image features (such as depth or curvature) beyond the low-level image features that are implemented in Itti et al.'s (1998) model.

References

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61,* 183–193. [PubMed]

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research, 46,* 2824–2833. [PubMed]

Barenholtz, E., Cohen, E. H., Feldman, J., & Singh, M. (2003). Detection of change in shape: An advantage for concavities. *Cognition, 89,* 1–9. [PubMed]

Bertamini, M. (2008). Detection of convexity and concavity in context. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 775–789. [PubMed]

Bhatt, R. S., Hayden, A., Reed, A., Bertin, E. & Joseph, J. (2006). Infants' perception of information along object boundaries: Concavities versus convexities. *Journal of Experimental Child Psychology, 94,* 91–113. [PubMed]

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

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America, 89,* 60–64. [PubMed] [Article]

Cate, D., & Behrmann, M. (2010). Perceiving parts and shapes from concave surfaces. *Attention, Perception and Psychophysics, 72,* 153–167. [PubMed] [Article]

Cohen, E. H., Barenholtz, E., Singh, M., & Feldman, J. (2005). What change detection tells us about the visual representation of shape. *Journal of Vision, 5*(4):3, 313–321, http://www.journalofvision.org/content/5/4/3, doi:10.1167/5.4.3. [PubMed] [Article]

Cohen, E. H., & Singh, M. (2007). Geometric determinants of shape segmentation: Tests using segment identification. *Vision Research, 47,* 2825–2840. [PubMed]

Cristino, F., & Baddeley, R. (2009). The nature of the visual representations involved in eye movements when walking down the street. *Visual Cognition, 17,* 880–903.

Denisova, K., Singh, M., & Kowler, E. (2006). The role of part structure in the perceptual localization of shape. *Perception, 35,* 1073–1087. [PubMed]

De Winter, J., & Wagemans, J. (2006). Segmentation of object outlines into parts: A large-scale integrative study. *Cognition, 99,* 275–325. [PubMed]

Edelman, S., & Weinshall, D. (1991). A self-organizing multiple-view representation of 3-D objects. *Biological Cybernetics, 64,* 209–219. [PubMed]

Feldman, J., & Singh, M. (2005). Information along contours and object boundaries. *Psychological Review, 112,* 243–252. [PubMed]

Foster, D. H., & Gilson, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proceedings of the Royal Society of*

London B: Biological Sciences, 269, 1939–1947. [PubMed] [Article]

Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual saliency in scene perception? *Perception, 36,* 1123–1138. [PubMed]

Hayward, W. G. (1998). Effects of outline shape in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 427–440.

Hayward, W. G., Tarr, M. J., & Corderoy, A. K. (1999). Recognizing silhouettes and shaded images across depth rotation. *Perception, 28,* 1197–1215. [PubMed]

He, P., & Kowler, E. (1989). The role of location probability in the programming of saccades: Implications for "center-of-gravity" tendencies. *Vision Research, 29,* 1165–1181. [PubMed]

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford, UK: Elsevier.

Henderson, J. M., Williams, C. C., Castelhano, M. S., & Falk, R. J. (2003). Eye movements and picture processing during recognition. *Perception & Psychophysics, 65,* 725–734. [PubMed]

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition, 18,* 65–96.

Hoffman, D. D., & Singh, M. (1997). Salience of visual parts. *Cognition, 63,* 29–78.

Hummel, J. E., & Stankiewicz, B. J. (1996). An architecture for rapid, hierarchical structural description. In T. Inui & J. McCelland (Eds.), *Attention and performance XVI: On information integration in perception and communication* (pp. 93–121). Cambridge, MA: MIT Press.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254–1259.

Johnston, S., & Leek, E. C. (2009). Fixation region overlap: A quantitative method for the analysis of fixational eye movement patterns. *Journal of Eye Movement Research, 1,* 1–12.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227. [PubMed]

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28,* 1311–1328. [PubMed]

Leek, E. C., Reppa, I., & Arguin, M. (2005). The structure of three-dimensional object shape representations: Evidence from part-whole matching. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 668–684. [PubMed] [Article]

Leek, E. C., Reppa, I., Rodriguez, E., & Arguin, M. (2009). Surface but not volumetric part structure mediates three-dimensional shape representation. *Quarterly Journal of Experimental Psychology, 62,* 814–829. [PubMed]

Lim, I.-S., & Leek, E. C. (in press). Curvature and the visual perception of shape: Theory on information along object boundaries and the minima rule revisited. *Psychological Review.* [PubMed]

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences, 4,* 6–14.

Lloyd-Jones, T. J., & Luckhurst, L. (2002). Outline shape is a mediator of object recognition that is particularly important for living things. *Memory & Cognition, 30,* 489–498. [PubMed]

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60,* 91–110.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation patterns made during brief examination of two-dimensional images. *Perception, 26,* 1059–1072. [PubMed]

Manor, B. R., & Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuo-cognitive tasks. *Journal of Neuroscience Methods, 128,* 85–93. [PubMed]

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

Melcher, D., & Kowler, E. (1999). Shapes, surfaces and saccades. *Vision Research, 39,* 2929–2946. [PubMed]

Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 10,* 1615–1630. [PubMed]

Rayner, K. (1995). Eye movements and cognitive processes in reading, visual search and scene perception. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications* (pp. 3–22). Amsterdam, Netherlands: Elsevier.

Renninger, L. W., Coughlan, J., & Verghese, P. (2005). An information maximization model of eye movements. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems* (vol. 17). Cambridge, MA: MIT Press. [PubMed]

Renninger, L. W., Verghese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local

uncertainty. *Journal of Vision, 7*(3):6, 1–17, http://www.journalofvision.org/content/7/3/6, doi:10.1167/7.3.6. [PubMed] [Article]

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025. [PubMed]

Schendan, H. E., & Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *Journal of Cognitive Neuroscience, 15,* 111–135. [PubMed]

Tarr, M. J., & Bülthoff, H. H. (Eds.) (1998). *Object recognition in man, monkey and machine*. Cambridge, MA: MIT Press.

Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception, 32,* 579–592. [PubMed]

Taubin, G. (1995). Estimating the tensor of curvature of a surface from a polyhedral approximation. *IEEE Computer Society, 902.*

Ullman, S. (1998) Three-dimensional object recognition based on the combination of views. In M. J. Tarr & H. H. Bülthoff (Eds.), *Object recognition in man, monkey and machine* (pp. 21–44). Cambridge, MA: MIT Press. [PubMed]

Ullman, S. (2006). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Science, 11,* 58–64. [PubMed]

Ullman, S., & Bart, E. (2004). Recognition invariance obtained by extended and invariant features. *Neural Networks, 17,* 833–848. [PubMed]

Ullman, S., & Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13,* 992–1006.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5,* 682–687. [PubMed]

Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency hypothesis. *European Journal of Cognitive Psychology, 18,* 321–342.

van Gompel, R. P. G., Fischer, M. H., Murray, W. S., & Hill, R. L. (2007). *Eye movements: A window on mind and brain*. Oxford, UK: Elsevier.

Vergilino-Perez, D., & Findlay, J. (2004). Object structure and saccade planning. *Cognitive Brain Research, 20,* 525–528. [PubMed]

Vishwanath, D., & Kowler, E. (2004). Saccadic localization in the presence of cues to three-dimensional shape. *Journal of Vision, 4*(6):4, 445–458, http://www.journalofvision.org/content/4/6/4, doi:10.1167/4.6.4. [PubMed] [Article]

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19,* 1395–1407. [PubMed]

Wexler, M., & Ouarti, N. (2008). Depth affects where we look. *Current Biology, 18,* 1872–1876. [PubMed]