Eye-position during identification of living and nonliving objects seen in canonical and non-canonical views

Torstein Låg

Bruno Laeng

Department of Psychology, University of Tromsø

Running head: Eye movements in object identification

Address for correspondence: Torstein Låg
Department of Psychology
University of Tromsø
N-9037 Tromsø
Norway

e-mail: tlaag@psyk.uit.no
phone.: +47 77 64 63 61
fax: +47 77 64 52 91

Abstract

Through the experiment described in this article we investigated the eye movements of observers during a simple object-naming task in which objects from living and nonliving categories were presented at different depth rotations. Specifically, we looked at their tendency to fixate the centre of gravity (COG) of the to-be-named figures. Results reveal a clear tendency to focus more on the COG as objects are rotated from canonical to non-canonical views. However, this tendency was reliable only for the nonliving objects. We interpret these results as indicating an increased reliance on a 'global' information extraction strategy with increasingly challenging contour shapes. Findings are consistent with an explanation of normal category-specific effects in object recognition that emphasises structural similarities and variations of visual form across categories.

What do our eyes do while we attempt to identify objects? Despite a rich and interesting research literature on human eye movements in visual perception (see e.g. Findlay & Gilchrist, 2003; Henderson & Ferreira, 2004), there is not much data that will guide an attempt to answer such a question. Object recognition has not been the main focus of this research. Rather, visual search tasks (Findlay, 2004), complex scene perception (Henderson & Hollingworth, 1998; 1999; 2003), and reading (Rayner & Juhasz, 2004), apart from studies of the detailed mechanics of eye movements in general (e.g. Findlay & Brown, 2006a; 2006b), have been the topics of choice among eye movement researchers. Although object recognition has been touched upon incidentally in much of the scene perception literature, there have, to our knowledge, been no studies on gaze during the identification of single objects.

The present work is an attempt to apply eye movement methodology to a commonly used experimental task in object recognition research. In this experiment, we used an unconstrained object-naming task in which participants saw single objects from various common categories. Our main interest was to explore the potential usefulness of eye movement methodology to inform hypotheses about the underlying processes of object recognition. Specifically we looked for systematic changes in eye-movement patterns with changes in the viewing angle of the presented objects, and with objects from two different domains (living and nonliving objects). As a starting point for developing a useful dependent measure, we chose a well-known phenomenon from the eye movement literature: The so-called 'global effect' (Findlay, 1982).

The 'Global effect'

Research using the eye-tracking method has shown that when subjects seek out targets within a visual display, the saccadic landing position often corresponds, at least approximately, to low-level geometric features of the whole image, and not the individual elements or parts. Specifically, saccadic end points frequently land on or near the centre of gravity (COG) of the luminance distribution of all the elements (distractors as well as targets) in the entire display (Coren & Hoenig, 1972; Findlay, 1982). This phenomenon in gaze control has been named the "global effect", to capture the idea that computation of the saccade landing position is based on the integration of visual information from a relatively large spatial area (Findlay, 1982).

The global effect seems to be a reliable and pervasive phenomenon. Early experiments used relatively simple stimulus displays of a few discrete elements (Coren & Hoenig, 1972; Findlay, 1982), but later studies have shown a global or centre of gravity effect for more complex displays. The effect tends to occur when viewers are instructed to look at spatially extended targets (Kowler & Blaser, 1995; Vishnawath & Kowler, 2003), such as dot patterns (McGowan, Kowler, Sharma & Chubb, 1998), or even perspective images of a single 3D shape (Vishnawath & Kowler, 2004). The effect is stable in that it occurs even when subjects are instructed to fixate a specific target not located in the COG or in its neighbourhood (Coren & Hoenig, 1972). This reliability and seeming mechanical nature of the global effect have led some researchers to hypothesise that the end point of a saccade is computed by a rapid, automatic, and involuntarily mechanism that, as such, is unable to disregard task-irrelevant stimulus elements (cf. e.g. Ottes, Van Gisbergen & Eggermont, 1985).

However, some studies show rather unequivocally that the global effect can be influenced to some extent by various contingencies. For instance, Coëffé and O'Regan (1987) have shown that the influence of non-targets on the saccadic landing position was attenuated if the predictability of the target location was increased, or if onset of the initial saccade was delayed. Although Coëffé and O'Regan attributed these effects to increased discriminability of target and non-target, consistent with rapid, automatic end point computation, more recent experiments seem to rule out such an interpretation. He and Kowler (1989), for instance, demonstrated a reduced global effect as a function of the probability of the target appearing in a given location. This effect was independent of stimulus discriminability. In other words, the global effect may be the result of a viewers' uncertainty about target and distractor locations within the stimulus array. In such cases, fixating the COG may simply be an effective default search strategy (He & Kowler, 1989; 1991). Support for this interpretation comes from recent work using a simulated ideal searcher that displayed a tendency to fixate a location near the centroid of a cluster of separate locations (Najemnik & Geisler, 2005).

In another series of experiments He and Kowler (1991) asked participants to move their eyes to either designated (but invisible) target locations inside surrounding contour shapes or to the contour shape as a whole. Fixations under instructions to look at the contour shape as a whole showed the usual global effect, with saccade endpoints landing on or near the symmetric point or the COG of the shape. It was also evident, however, that participants were quite accurate at fixating the predetermined specific locations within the contour form, even when these did not correspond to the COG. This means that decisions on the part of the subjects can influence the computation of saccades in such a manner that irrelevant stimulus elements are disregarded. He and Kowler's (1991) experiments also revealed that saccades to designated but invisible target locations presented *without* a contour shape were considerably less accurate. This implies that the contour shapes served as reference frames for the computation of saccadic endpoints, but that they did not serve to pull saccades toward the COG unless subjects attempted to look at the form as a whole.

The research of He and Kowler (1989; 1991) thus clearly indicates that the global effect may be the consequence of an effective default information gathering or search strategy employed by our visual system. The effect can be modulated by task demands or stimulus presentation parameters. These conclusions may prove a valuable starting point for investigations of object identification by means of eye movements. In particular, changes in the reliance on a global effect strategy may index changes in the perceptual system's demand for visual information. Our focus in the present study was on changes in objects' visual presentation with rotation or viewpoint, and the different demands of recognising objects from different domains (living and nonliving).

Contours or global shape in object identification

One could claim that the central theoretical question in object recognition is how the perceptual system handles the fact that the same object projects a variety of images onto the retina, depending on, firstly, the viewpoint from which it is observed. The literature on object recognition and identification processes has been dominated during the last 25 years by two major types of answers to this question (see Hayward, 2003 and Palmeri & Gauthier, 2004 for reviews). According to one account, object recognition proceeds by reconstruction of an accurate 3D description of the object in terms of i) its constituent primitive parts, or "geons", and ii) their spatial relations, with respect to an object centred reference frame, in short, its structure (e.g. Biederman, 1987; Marr, 1982; Marr & Nishihara, 1978). The rival account proposes that a finite number of 2D views or images of objects are stored in memory and novel views are recognized initially by interpolating the known views (e.g. Bülthoff & Edelman, 1992; Tarr & Bülthoff, 1995; Ullman, 1996). The geon structural theories predict that objects can be recognised equally well from different viewpoints, as long as the same constituent parts remain visible in those various viewpoints (Biederman & Gerhardstein, 1993). In contrast, the image-based, so-called view-dependent models, predict that object recognition can be more or less efficient depending on the similarity of the input image to those in the catalogue of stored images. According to this view, recognition of objects will be viewpoint dependent, even when the same constituent parts are visible in the different images.

Whichever theoretical perspective one might lean toward, it seems clear that both of them place considerable emphasis on the importance of silhouettes or outline shape in object recognition. For instance, in the case of one image based recognition scheme (Huttenlocher & Ullman, 1990), edge detection is first used to establish the contours present in an image, before constrained alignment transformations are applied and matching of image contours to a stored model is attempted. In fact, for some schemes, the problem of recognising one object from different views can be stated in terms of predicting the changes in an object's silhouette which results from a different orthographic projection of its "rim" (see Ullman, 1996, pp. 63-67).

For the geon-structural theories, the emphasis on object contours is equally clear. These theories typically hypothesise a stage in which an object is segmented into its various constituent parts. Evidence has accrued over the years, that features of the object outline is important for this segmentation. Much work has been done, for instance, on the role of local minima of negative curvature (i.e. concave regions) of the outline as cues to part boundaries of an object (Feldman & Singh, 2005; Hoffman & Richards, 1984; Hoffman & Singh, 1997). These outline regions seem to carry more information than positively curved (convex) regions, as evidenced by a number of experimental studies. For instance, changes in shape are more noticeable if they involve concave outline regions (Barenholtz, Cohen, Feldman & Sing, 2003; Cohen, Barenholtz, Singh, & Feldman, 2005), concave targets are more easily detected in visual search (Humphreys & Müller, 2000), people consciously segment natural objects along negative minima of curvature (De Winter & Wagemans, 2004; 2006), and 6 months old infants are better able to discriminate concavities, particularly when they are part of closed shapes (Bhatt, Hayden, Reed, Bertin, & Joseph, 2006). Some models further refine the role of contour features and integrate them with other global shape properties, such as width, axes and size (e.g. Burbeck, Pizer, Morse, Ariely, Zauberman & Rolland, 1996; Siddiqi, Tresness & Kimia, 1996; Siqqidi, Kimia, Rannenbaum & Zucker, 2001).

Other, more model independent research, confirms the importance of outline shapes in object recognition and encoding. The study by Hayward (1998) indicates that silhouettes of objects are recognised nearly as effectively as shaded images, and that changes in outline shape predict identification performance. In an early visual memory experiment by Rock, Halper and Clayton (1972) it was found that when discriminating already seen figures from novel figures, participants more reliably remembered outline shape than the internal details of the image.

Given the importance of object boundaries in object recognition, one would suppose that integrating information from a relatively large spatial area (the area of the whole shape) would be beneficial to the recognition process. In other words, attempting to view the object as a whole, rather than focusing on its particular details, would be a sensible strategy. Thus, one would expect eye-movements during recognition of single objects to reflect this by displaying a certain reliance on the visual strategy underlying the global eye movement effect. Furthermore, and from the same reasoning, reliance on a global strategy would likely depend on the difficulty of processing the to-be-recognised shape. More difficult shapes would require more sampling of outline and contour information, thus a heavier reliance on a global effect strategy.

Now, it is a well-known phenomenon that rotating a visual pattern, shape or object increases the demands on recognition. Apart from a number of experiments aimed at informing the so-called viewpoint debate (e.g. Biederman & Gerhardstein, 1993; Hayward, 1998; Tarr & Pinker, 1989; Wilson & Farah, 2003), the classic study by Palmer, Rosch and Chase (1981) explores the effects and correlates of perspective change on object perception. These researchers use the concept of a 'canonical perspective' to express the fact that people are better able to identify objects that are presented in the perspective most people would consider to yield the best image of the object. In the present experiment, we challenged the shape perception of our participants by manipulating the perspective (i.e. the depth rotation around the vertical axis) of the to-be-identified objects.

Category-effects in object identification

Within another current of vision research there has, in recent years, been mounting interest in the topic of category-specific effects in object identification. In certain conditions, objects belonging to specific categories or domains (e.g. living vs. nonliving objects) are identified (at basic level) more accurately and rapidly than objects belonging to other categories (see e.g. Capitani, Laiacona, Barbarotto & Trivelli, 1994; Gerlach, 2001, Laws, 2000; Laws & Neve, 1999; Låg, 2005). Specifically, in experiments where the cross-category influence of factors such as concept familiarity and visual complexity are controlled, the identification of objects belonging to the domain of living things is more accurate and rapid than identification of nonliving things.

Although the causes of these effects are far from evident, there are indications that the informativeness of objects' overall global shape may statistically differ according to category and in turn exert an influence that yields a living things advantage in identification. For instance, in an experiment by Lloyd-Jones and Luckhurst (2002) participants performed an object decision task (i.e. deciding whether an object is real or not). In one condition, the object stimuli were visible only as silhouettes, whereas in the other condition the objects were presented as normal line-drawings complete with internal details. There was a general advantage for living things, but this difference between living and nonliving things was larger in the silhouette condition. Lloyd-Jones and Luckhurst (2002) suggested that there is less useful information in nonliving things' outline contours. Låg, Hveem, Ruud and Laeng (2006) obtained a similar result. Using picture-name verification and both blurred and clear images, it was found that the living-things advantage was considerably increased with blurred pictures compared to clear. Both these studies clearly indicate that the overall global shape, or outline, of an object's 2D representation is more informative with regard to identity for living than for nonliving things. An experiment by Vannucci, Viggiano and Argenti (2001) also provides support for the special role of the global shape of living things. Their participants identified images of living and nonliving objects. These images were spatially filtered at nine different levels of resolution. This made it possible to determine the identification threshold for each of three categories. Results showed that animals were on average identified at a lower level of resolution than tools and vegetables, indicating that the information provided by the global shape of animals reveals more of their identity, whereas the global shape of tools is less helpful to their identification.

If it really is the case that living things provide more "stable" (Laws & Neve, 1999, p.1268) visual representations in the sense of a low structural variability within a basic level object class (e.g. think of the contour similarity shared by all horses), then matching their shapes to stored models should require less intense visual processing than the recognition of nonliving things. In other words, one would expect a heavier reliance on the global effect strategy for nonliving things, since nonliving things present more challenging shapes.

Summary

The present experiment was aimed at exploring the use of eye movement methodology in a single object recognition paradigm. We presented our subjects with images of common objects belonging to the domains of living and nonliving things presented in three different depth rotations. We looked specifically for changes in the participants' tendency to position their eyes on or near the COG of the presented object as a function of canonicalness of perspective (cf. Palmer et al., 1981) and object domain (living or nonliving). If it is the case, as the research of He and Kowler (1989; 1991) suggests, that the global effect in eye movements reflects an efficient default information gathering and search strategy, then we should see an increased tendency to focus the COG of the presented figures when recognition of their shapes become more taxing. Given the considerations set forth in the previous sections, this would imply increased COG-viewing with rotations away from the canonical viewpoint of the object. Also, if the shapes of nonliving things present more of a challenge to our perceptual system, then we should see a heavier reliance on COG-viewing with nonliving compared to living things.

Method

Participants

42 students and employees at the University of Tromsø, 24 males and 18 females (age range 18-54 years), volunteered to participate in an experiment on object identification. All participants had normal, or corrected to normal (with contact lenses) vision, and all were native speakers of Norwegian.

Apparatus and stimuli

Eye movements were recorded by means of the Remote Eye Tracking Device, R.E.D., built by SMI-SensoMotoric Instruments in Teltow (Germany). Analyses of recordings were then computed by use of the iView software, also developed by SMI. The R.E.D. II can operate at a distance of 0.5-1.5 m and the recording eye-tracking sample rate is 50 Hz., with resolution better than 0.1 degree. The eye-tracking device operates by determining the positions of two elements of the eye: The pupil and the corneal reflection. The sensor is an infrared light sensitive video camera typically centred on the left eye of the subject. The coordinates of all the boundary points are fed to a computer that, in turn, determines the centroids of the two elements. The vectorial difference between the two centroids determines the "raw" computed eye position.

Stimuli consisted of a total of sixty realistic 3D models of common objects from the categories of animals, fruits/vegetables, vehicles, musical instruments, and other inanimate objects (mostly house appliances and furniture). Five different sample views of each of the 60 objects were downloaded from the Internet site of Digimation Inc. (Digimation, 2002) or from the "Object Databank, Tarrlab" CD-ROM of object stimuli (Tarr, 1996). In a pilot study, 25 participants named each object from one of its five viewpoints (thus, each object was viewed by 5 pilot participants). Given previous results (Palmer, Rosch & Chase, 1981) of better identification of perspectives of objects that are a) judged to be seen in good views and b) rated as displaying more important visual information ("canonical views"), we selected (based on pilot participants' ease of naming; as indexed by their accuracy rates and RTs), three views of each object, one canonical, one intermediate and one non-canonical. On screen, the vertical and horizontal axes of the objects subtended from 5.7 to 16.3 degrees of visual angle, averaging about 12.3 degrees. Hence, the outlines of the objects could never be included as a whole within the foveal or parafoveal regions of the retinae. Sample stimuli are displayed in Figure 1.

Insert Figure 1 about here

Procedure

Pictures were distributed into three blocks, with each object being presented once in each block. Canonical, intermediate and non-canonical views were equally distributed between the three blocks. Objects were ordered in the same fixed random sequence in each block. Participants were randomly assigned to one of the three blocks.

Participants were seated in front of a 15-inch flat screen monitor at a distance of 60 cm with their heads in a chin-and-forehead rest to limit head movements. The eye-tracker was calibrated before each session according to a standard routine where participants fixated nine calibration points corresponding to a regularly spaced 3x3 matrix. Participants were then told that they would see a number of common objects presented one after the other on the screen, and that their task was to name the object as soon as it was recognised. In each trial, before an object image was presented, there was a black fixation cross that appeared on a blank white background and randomly in one of the four corners of the screen. Participants were requested to fixate on the cross. While the participant maintained fixation, the experimenter initiated the presentation of an object. Eye movements were recorded from the start of each object image presentation and until the participant gave a response.

Results

Determining the centre of gravity

First we estimated the COG of each object image. The following algorithm was used: For every point in the image, a vertical and horizontal line was drawn through it. Then, the number of pixels actually part of the figure (and not the background) on each side of these two lines was summed. The point at which the numbers of pixels belonging to the figure were most closely matched on both sides of both the horizontal and vertical lines was taken as the COG. In addition, a circle with a diameter corresponding to two degrees of visual angle (i.e. approximately matching the area of the fovea when fixating precisely on a given location) was defined around each estimated COG point. For analysis purposes, eye position was scored as being in the region of the COG whenever both the horizontal and vertical eye position fell within this circle. Figure 2 shows two stimulus items with COGs determined by this method.

Insert Figure 2 about here

Percentage of time spent by the eye within the object's centre

To examine the degree to which participants looked within the 2º area centred on the COG, we determined the percentage of time on each trial that the eye position was within the defined circular region. In other words, this measure provides an estimate of the total time spent looking at the centre of each object relative to the total view-time for each image. Specifically, the average percentage across participants for each object in each rotation was computed. These averages were then subjected to repeated measures ANOVA by-items with Rotation (canonical, intermediate, non-canonical) as a single within-subjects factor. This analysis revealed a significant effect of Rotation, $F(2, 118) = 13.3$, $MSE = 56.4$, $p < .001$. Simple effects analyses showed this effect to be due to a higher percentage of time spent looking in the COG of non-canonical ($M = 23.7\%$) compared to canonical ($M = 16.9\%$) and intermediate ($M = 18.5\%$) rotations, $F(1, 59) = 20.4$, MSE 67.5, $p < .001$ and $F(1, 59) = 11.5$, $MSE = 69.6$, $p < .001$ respectively. There was no difference between the canonical and intermediate rotations, $F(1, 59) = 2.4$, n.s. These results are illustrated in Figure 3.

Insert Figure 3 about here

However, it is possible that when using the above computations based on the percentage of time the eye is within 2º of the COG, the effect of rotation is confounded with changes in the total area of each object's figures. Indeed, as the objects are rotated from canonical, through intermediate and to non-canonical, the area of their figures in the images (defined as the percentage of total image pixels belonging to the actual figure) tends to diminish (canonical $M = 25.8$, intermediate $M = 24.5$, non-canonical $M = 21.7$; $F(2, 118) = 14.7$, $MSE = 18.1$, $p < .001$). Thus, the fact that participants spend more time looking in the COG with non-canonical rotations could just reflect the fact that the 2° of visual angle around the COG makes up a proportionally larger area of the object's figure. To control for this factor, the area of each object's figure in each rotation was regressed, along with the rotation factor, on the percentage of time the eye spent in the COG. In a hierarchical regression analysis, area of figure was entered in Step 1, while rotation was entered in Step 2. If area of figure can account for the observed effect of rotation on percentage time in COG, one would expect Step 2 not to contribute to the predictive power of the model. Table 1 summarises the results of this analysis.

Insert Table 1 about here

It is clear from this analysis that although there was a relationship between diminishing area of the figure and the time the eye position is in the COG, this was insufficient to account for the effect of rotation. Adding rotation as a variable significantly increases the model's predictive power (*F change* $(1, 177) = 11.7$, $p = .001$).

Another possible confound in the %-of-time-in-COG-measure is the following: If participants fixate the COG initially, in order to obtain an overview of the visual information, and only subsequently let their eyes drift or saccade to other parts of the figure, then the short trials will have a higher percentage of time with eye-positions in the COG-region. Even though this seems to be a very unlikely explanation for the rotation effects observed here (since trials tend to become longer with increasingly non-canonical views), we nevertheless computed correlations of trial duration and % time in the COG for each of the three rotation conditions and for all conditions together. All *r* were close to zero, and all $p > .25$.

Spread of ocular positions

Although the participants spend more time looking within the central area around the COG when objects are presented in a non-canonical rotation, this finding may be compatible with different patterns of eye-movement. For instance, even though more eye positions in the non-canonical condition were in the COG, it is possible that participants made brief forays away from the COG to fixate specific elements of the image. This should be reflected in an increased spread of eye positions. To investigate this possibility, we calculated the deviation of both the horizontal (x) and the vertical (y) coordinate positions of the eye from the x and y coordinate positions of the centre of the COG. Using these deviations, we then, for each participant, computed the standard deviation of the x and y coordinate eye position for each image. These standard deviations were then averaged across participants and across coordinate axes, and subjected to repeated-measures ANOVA by items with Rotation (canonical, intermediate, non-canonical) as a single within-subjects factor. This revealed a small, but significant, effect of Rotation, $F\ (2, 118) = 3.4$, $MSE = 327$, $p < .05$. Simple effects analyses showed this effect to be due to significantly higher spread in the canonical ($M = 175$) compared to the non-canonical ($M = 167$) condition, $F\ (1, 59) = 5.5$, $MSE = 402$, $p < .05$. No other comparisons were significant. These findings indicate that the spread of eye positions did not increase with intermediate or non-canonical rotations. On the contrary, the spread of eye positions actually diminished somewhat when viewing the non-canonical perspectives. The results of this analysis are illustrated in Figure 4.

Insert Figure 4 about here

The pattern of results from the analyses reported so far is graphically illustrated in Figure 5, where the origin of each panel corresponds to the COG of the object for the relevant rotation condition and each point corresponds to the eye-position during one 20 ms sampling period. The Figure is based on data from a total of 14 subjects for each condition, while they viewed two representative objects – the honeybee and the car. An inspection of Figure 5 should show that the eye was mainly positioned in the immediate area around the centre of the COG in the non-canonical condition, and that there was a slightly decreasing spread of eye-position as the object is rotated away from the canonical view.

Insert Figure 5 about here

Average eye position relative to the COG

An alternative to the strategy of making brief saccades away from the COG, which was rejected by the previous analysis, would be a strategy in which the eye remains at all times in the neighbourhood of the COG but drifts of eye position in the direction of salient parts of the outline are either tolerated, or used strategically to increase the level of resolution of diagnostic portions of the image. To investigate this possibility, we also computed the averages of the deviations of both the horizontal (x) and the vertical (y) coordinate positions of the eye from the x and y coordinate positions of the centre of the COG for each image and across participants and coordinate axes. There were no differences between the canonical rotations (M = 12.5, SD = 73.6), the intermediate rotations (*M* = 5.0, *SD* = 84.2), and the non-canonical rotations (*M* = 21.3, *SD* = 81.5), $F(2, 118) = 0.57$, n.s.

Category differences in looking at the COG

To examine the degree to which participants' viewing time in the COG differed between categories, we used the same measure (% of time in COG) as for the first analysis. For each participant, an average percentage of time in the COG region across all living objects (animals: *n* = 12; produce: *n* = 12) and across all nonliving objects (vehicles: *n* = 12; musical instruments: *n* = 12; appliances: *n* = 12) was computed. Averages were subjected to repeated measures ANOVA by subjects, with Rotation (canonical, intermediate and non-canonical) and Domain (living and nonliving) as the within-subjects factors. The analysis revealed a significant effect of Rotation, $F\ (2, 82) = 13.3$, $MSE = 32.1$, $p < .001$, with, as already seen in the first analysis, increasing time spent within the COG region from canonical to non-canonical views. There was also a main effect of Domain, $F\ (1, 41) = 6.1$, $MSE = 16.9$, $p < .05$, with less time spent in the COG region for living things (*M* = 18.3%) than for nonliving things (*M* = 19.6%). However, these main effects need to be interpreted in the light of an interaction between Rotation and Domain, $F\ (2, 82) = 5.8$, $MSE = 21.3$, $p < .01$. Paired samples t-tests on all cell mean comparisons (with *p*-level Bonferroni-adjusted to .005 for multiple comparisons) revealed a significant difference in % time spent looking in the COG region for nonliving things when comparing non-canonical to intermediate and canonical views, $t(41) = 5.1$, $p < .001$ and $t(41) = 7.5$, $p < .001$ respectively. Again, the % time increased from canonical and intermediate to non-canonical views. For the non-canonical views, there was also a significant difference in COG permanence for living compared to nonliving things, $t(41) = 4.3$, $p < .001$, with more time spent looking in the COG region for nonliving (*M* = 23.4) than for living things (*M* = 19.5). No other comparisons were significant. These results are illustrated in Figure 6.

Insert Figure 6 about here

Fixation analyses

In the foregoing sections we have presented results from relatively unrefined eye-position data. We reasoned that given the ease with which normal humans recognise common objects, fixations might be relatively sparse in our displays. Indeed, research on eye movements in face perception suggests that additional fixations within a single face stimulus display do little to facilitate recognition (Walker-Smith, Gale, & Findlay, 1977). Thus, using data from every 20 millisecond sampling period seemed sensible, in order to retain as much of the raw eye position data as possible.

Commonly, though, eye-movement researchers tend to analyse their data in terms of fixations (i.e. the eye position remaining relatively stable over time before saccading to a different fixation). We therefore decided to provide fixation analyses in addition to the analyses already presented. We defined a fixation as eye position remaining within a 60 pixel area for a minimum of 150 milliseconds. For the following analyses we counted the proportion of total fixations on each trial that fell within the COG.

First, the average proportions of COG fixations across participants for each object in each rotation were computed. These averages were then subjected to repeated measures ANOVA by-items with Rotation (canonical, intermediate, non-canonical) as a single within-subjects factor. This analysis revealed a significant effect of Rotation, $F(2, 118) = 10.4$, $MSE = 0.01$, $p < .001$. Simple effects analyses showed this effect to be due to a higher proportion of fixations in the COG of non-canonical ($M = 0.29$) compared to intermediate ($M = 0.23$) rotations, $F(1, 59) = 6.6$, MSE 0.02, $p < .05$ and a higher proportion of COG fixations in the intermediate compared to the canonical ($M = 0.19$) rotations, $F(1, 59) = 5.5$, $MSE = 0.007$, $p < .05$. Thus, in fact, the pattern of increasing COG-viewing with increasingly non-canonical rotations seems at least as strong with the fixation analyses as with the raw eye position analyses.

Second, in order to test the effect of object domain, we computed, for each participant, an average proportion of total trial fixations in the COG region across all living objects (animals: $n = 12$; produce: $n = 12$) and across all nonliving objects (vehicles: $n = 12$; musical instruments: $n = 12$; appliances: $n = 12$). These averages were subjected to repeated measures ANOVA by subjects, with Rotation (canonical, intermediate and non-canonical) and Domain (living and nonliving) as the within-subjects factors. The analysis revealed a significant effect of Rotation, $F\ (2, 82) = 7.7$, $MSE = 0.02$, $p < .01$, with a larger proportion of fixations in the COG region from canonical to non-canonical views. The main effect of Domain was not significant, $F\ (1, 41) = 2.4$, $MSE = 0.013$, $p = .13$. However, there was an interaction between Rotation and Domain, $F\ (2, 82) = 5.1$, $MSE = 0.013$, $p < .01$. Paired samples t-tests on all cell mean comparisons (with p-level Bonferroni-adjusted to .005 for multiple comparisons) revealed a significant difference in proportion of fixations in the COG region for nonliving things when comparing non-canonical to intermediate and canonical views, $t(41) = 4.2$, $p < .001$ and $t(41) = 6.3$, $p < .001$ respectively. For the non-canonical views, there was also a significant difference in COG fixations for living compared to nonliving things, $t(41) = 3.4$, $p = .002$, with a higher proportion of COG fixations for nonliving ($M = 0.30$) than for living things ($M = 0.22$). No other comparisons were significant. These results are illustrated in Figure 7.

Insert Figure 7 about here

Discussion

The present experiment aimed to explore the use of eye movement methods in the study of object recognition. Specifically, we wanted to investigate whether the tendency to view a stimulus' centre of gravity (COG) could be modulated by object stimulus parameters; depth rotation and object domain. We found that (i) participants tend to increase the time spent looking at an object's COG when the object's rotation goes from canonical (cf. Plamer et al., 1981) to non-canonical. This tendency is not combined with brief saccades to the object outline, or with minor drifts in average eye-position away from the COG; (ii) this effect of rotation is only evident for nonliving objects, however.

These results are interesting on several levels. First, the fact that quite subtle manipulations produce observable effects on global measures of eye-movements is important in its own right. It implies that such measures do index object recognition processes, even when using single-object displays. Previous eye movement studies have tended to use more complex visual displays and to focus on scene perception (e.g. Henderson & Hollingworth, 1998) and visual search tasks (e.g. Findlay, 2004). To the extent that object identification has been an issue at all, manipulations of the displays have tended to be quite drastic, such as removing objects from a scene by means of artificial scotomas (e.g. Henderson, McClure, Pierce & Schrock, 1997), or manipulating the location of objects within scenes (e.g. Loftus & Macworth, 1978). Rotations of one and the same object and manipulations of objects' semantic category have, to our knowledge, not been studied by means of eye movements, despite the frequent use of these kinds of manipulations in research on object identification (see e.g. Hayward, 2003; Låg et al., 2006). Thus, the present findings provide initial support for the notion that measures of eye movements can reveal important clues to the information processes underlying object recognition in simple stimulus presentation paradigms of single objects.

Second, although interpreting the specific implications of the present findings in terms of computational processes is not a straightforward task, we do believe that some qualified speculations, based on previous findings from the object recognition literature along with the present results, can provide a platform for new studies, and thus, in the future, for firmer conclusions regarding specific information processes. For instance, the amount of evidence converging on the conclusion that object outline or bounding contours plays a prominent role in object recognition is substantial (e.g. Barenholz, Cohen, Feldman & Singh, 2003; De Winter & Wagemans, 2004; Driver & Baylis, 1995; Feldman & Singh, 2005; Hoffman & Richards, 1984 Hoffman & Singh, 1997; Paupathy & Connor, 2001). Furthermore, there is evidence that changes in outline shape predicts identification performance when simple 3D objects are rotated in depth (Hayward, 1998). That the effect of 3D-rotation on eye movements observed in the present experiment at least partially reflects the processing of changes in outline shape therefore seems like a plausible contention. This is further supported by findings indicating that when participants are instructed to view an object as a whole, they tend to fixate the object's COG (He & Kowler, 1991). On this account, then, effective localisation and processing of outline features (the global shape of the object) becomes increasingly important and challenging when an object is rotated away from its canonical view. If one accepts that fixating the COG of any stimulus array is an effective default strategy for information extraction from its global shape features (cf. He & Kowler, 1989; 1991), an increase in COG viewing-time would be the natural consequence of depth rotation. In fact, the finding of a slight decrease in spread of ocular position with increasingly difficult rotations suggests that for canonical and less demanding views, identification from contour shape is efficient, allowing the eye to actually fixate other parts of the display without compromising task performance. In contrast, for more demanding shapes (non-canonical), looking away from the COG would hamper identification, and is thus avoided.

A similar line of reasoning can be applied to make sense of the finding that depth rotation affects nonliving, but not, at least to the same extent, living objects. If, when attempting to extract information about an object's overall shape, the most effective strategy is viewing the object as a whole, and therefore positioning the gaze in the region of the COG, then more challenging shapes should induce more use of this strategy. The current pattern of results thus fits well with an account of category specific living things identification advantages that emphasises the relative stability and informativeness of the global shapes of living things (e.g. Gerlach, 2001; Gerlach, Law & Poulson, 2004; Laws & Neve, 1999). However, the fact the category difference in eye movements only emerges strongly for the non-canonical rotations may imply a slightly different sense for the expression "stable visual representation" than the one originally intended. If our reasoning concerning the increased use of a global effect strategy is valid, then the stability of living things' visual representations may extend beyond low structural variability among exemplars of a basic level object to include the stability of living things contour shapes across depth rotations. Conversely, rotating a nonliving thing induces more drastic contour alterations, thus taxing the shape-recognition processes. Accurate and objective measures of shape similarity may in the future confirm these speculations.

Another aspect of the present results worth commenting is the fact that they indicate that object identification is quite effective with only low-resolution visual information. The area typically foveated (2 degrees of visual angle) when fixating the COG is considerably smaller than the average area of the objects presented to the observer (about 12 degrees of visual angle). Thus, the contours of the shapes will not have been available for the high acuity fovea when the participants focused on the COG. This probably means that the low-resolution portions of the visual field provide the informational basis for the object identifications. This conclusion is strengthened by previous results from an experiment by Henderson, McClure, Pierce and Schrock (1997) in which the presence of a foveal scotoma did not disrupt identification of objects in a multi-object display.

In sum, then, our results support the idea that it is primarily an object's silhouette or global shape that provides the information that is crucial for recognition (Hayward, 1998). By extension, it indicates that explaining category-specific effects in normal object recognition by means of variations in structural or global shape stability and similarity (e.g. Gerlach et al., 2004; Laws & Neve, 1999; Låg et al., 2006) is at the very least a plausible approach.

References

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

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

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

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

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 USA, 89*, 60-64.

Burbeck, C.A., Pizer, S.M., Morse, B.S., Ariely, D., Zauberman, G.S., & Rolland, J.P. (1996). Linking object boundaries at scale: A common mechanism for size and shape judgements. *Vision Research, 36*, 361-372.

Capitani, E., Laiacona, M., Barbarotto, R., & Trivelli, C. (1994). Living and non-living categories. Is there a "normal" asymmetry? *Neuropsychologia, 32,* 1453-1463.

Coëffé, C., & O'Regan, J.K. (1987). Reducing the influence of non-target stimuli on saccade accuracy: Predictability and latency effects. *Vision Research, 27*, 227-240.

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*, 313-321.

Coren, S., & Hoenig, P. (1972). The effect of non-target stimuli on length of voluntary saccades. *Perceptual and Motor Skills, 34*, 499-508.

De Winter, J., & Wagemans, J. (2004). Contour-based object identification and segmentation: Stimuli, norms and data, and software tools. *Behavior Research Methods, Instruments, & Computers, 36*, 604-624.

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

Digimation. (2002). Sample pictures of 3D models. Retreived April 2002 from http://www.digimation.com.

Driver, J., & Baylis, G.C. (1995). One-sided edge assignemnt in cision: 2. Part decomposition, shape description, and attention to objects. *Current Directions in Psychological Science, 4*, 201-206.

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

Findlay, J.M. (1982). Global visual processing for saccadic eye movements. *Vision Research, 22*, 1033-1045.

Findlay, J.M. (2004). Eye scanning and visual search. In J.M. Henderson and F. Ferreira (eds.). *The Interface of Language, Vision and Action: Eye Movements and the Visual World*. New York: Psychology Press.

Findlay, J.M., & Brown, V. (2006a). Eye scanning of multi-element displays: I. Scanpath planning. *Vision Research, 46*, 179-195.

Findlay, J.M., & Brown, V. (2006b). Eye scanning of multi-element displays: II. Saccade planning. *Vision Research, 46*, 216-227.

Findlay, J.M., & Gilchrist, I.D. (2003) *Active Vision: The Psychology of Looking and seeing*. Oxford: Oxford University Press.

Gerlach, C. (2001). Structural similarity causes different category-effects depending on task characteristics. *Neuropsychologia*, *39,* 895-900.

Gerlach, C., Law, I., & Paulson, O.B. (2004). Structural similarity and category-specificity: A refined account. *Neuropsychologia, 42*, 1543-1553.

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. (2003). After the viewpoint debate: Where next in object recognition. *Trends in Cognitive Sciences, 7*, 425-427.

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.

He, P., & Kowler, E. (1991). Saccadic localization of eccentric forms. *Journal of the Optical Society of America, A, 8*, 440-449.

Henderson, J.M., & Ferreira, F. (eds.) (2004). *The Interface of Language, Vision, and Action. Eye Movements and the Visual World*. New York: Psychology Press.

Henderson, J.M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (ed.), *Eye Guidance in Reading and Scene Perception* (pp. 269-293). Amsterdam: Elsevier.

Henderson, J.M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243-271.

Henderson, J.M., & Hollingworth, A. (2003). Eye movements, visual memory, and scene representation. In M.A. Peterson and G. Rhodes (eds.), *Perception of Faces, Objects and Scenes: Analytic and Holistic Processes* (pp. 356-383). New York: Oxford University Press.

Henderson, J.M., & McClure, K.K., Pierce, S., & Schrock, G. (1997). Object identification without foveal vision: Evidence from an artificial scotoma paradigm. *Perception & Psychophysics, 59*, 323-346.

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.

Humphreys, G.W., & Müller, H. (2000). A search asymmetry reversed by figure-ground assignment. *Psychological Science, 11*, 196-201.

Huttenlocher, D.P., & Ullman, S. (1999). Recognizing solid objects by alignment with an image. *International Journal of Computer Vision, 5*, 195-212.

Kowler, E., & Blaser, E. (1995). The accuracy and precision of saccades to small and large targets. *Vision Research, 35*, 1741-1754.

Laws, K.R. (2000) Category-specific naming errors in normal subjects: The influence of evolution and experience. *Brain and Language, 75,* 123-133.

Laws, K.R., & Neve, C. (1999). A 'normal' category-specific advantage for naming living things. *Neuropsychologia*, *37,* 1263-1269.

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.

Loftus, G.R., & Mackworth, N.H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 565-572.

Låg, T. (2005). Category-specific effects in object identification: What is 'normal'? *Cortex*, *41*, 833-841.

Låg, T., Hveem, K., Ruud, K.P.E., & Laeng, B. (2006). The visual basis of category effects in object identification: Evidence from the visual hemifield paradigm. *Brain & Cognition, 60*, 1-10.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

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.

McGowan, J.W., Kowler, E., Sharma, A., & Chubb, C. (1998). Saccadic localization of random dot targets. *Vision Research, 38*, 895-909.

Najemnik, J., & Geisler, W.S. (2005). Optimal eye movement strategies in visual search. *Nature, 434*, 387-391.

Nelson, W.W., & Loftus, G.R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 391.399.

Ottes, F.P., Van Gisbergen, J.A.M., & Eggermont, J.J. (1985). Latency dependence of colour-based target vs nontarget discrimination by the saccadic system. *Vision Research, 26*, 849-862.

Palmer, S., Rosch, E., & Chase, P. (1981). Canonical perspectives and the perception of objects. In J. Long & A. Baddeley (eds.) *Attention and Performance IX*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Palmeri, T.J., & Gauthier, I. (2004). Visual object understanding. *Nature Neuroscience, 5*, 291-303.

Parker, R.E. (1978). Picture processing during recognition. *Journal of ExperimentalPsychology: Human Perception and Performance, 4*, 284-293.

Paupathy, A., & Connor, C.E. (2001). Shape representation in area V4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology, 86*, 2505-2519.

Rayner, K., & Juhasz, B.J. (2004). Eye movements in reading: Old questions and new directions. *European Journal of Cognitive Psychology, 16*, 340-352.

Rock, I., Halper, F., & Clayton, T. (1972). The perception and recognition of complex figures. *Cognitive Psychology, 3*, 655-673.

Siddiqi, K., Kimia, B.B., Tannenbaum, A., & Zucker, S.W. (2001). On the psychophysics of the shape triangle. *Vision Research, 41*, 1153-1178.

Siddiqi, K., Tresness, K.J., & Kimia, B.B. (1996). Parts of visual form: Psychophysical aspects. *Perception, 25*, 399-424.

Tarr, M.J. (1996). Tarrlab CD-ROM.

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

Tarr, M.J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21*, 233-282.

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

Vanucci, M., Viggiano, M.P., & Argenti, F. (2001). Identification of spatially filtered stimuli as function of the semantic category. *Cognitive Brain Research, 12*, 475-478.

Vishwanath, D., & Kowler, E. (2003). Localization of shapes: eye movements and perception compared. *Vision Research, 43*, 1637-1653.

Vishwanath, D., & Kowler, E. (2004). Saccadic localization in the presence of cues to three-dimensional shape. *Journal of Vision, 4*, 445-458.

Walker-Smith, G.J., Gale, A.G., & Findlay, J.M. (1977). Eye movement strategies involved in face perception. *Perception, 6*, 313-326.

Wilson, K.D., & Farah, M.J. (2003). When does the visual system use viewpoint-invariant representations during recognition? *Cognitive Brain Research, 16*, 399-415.

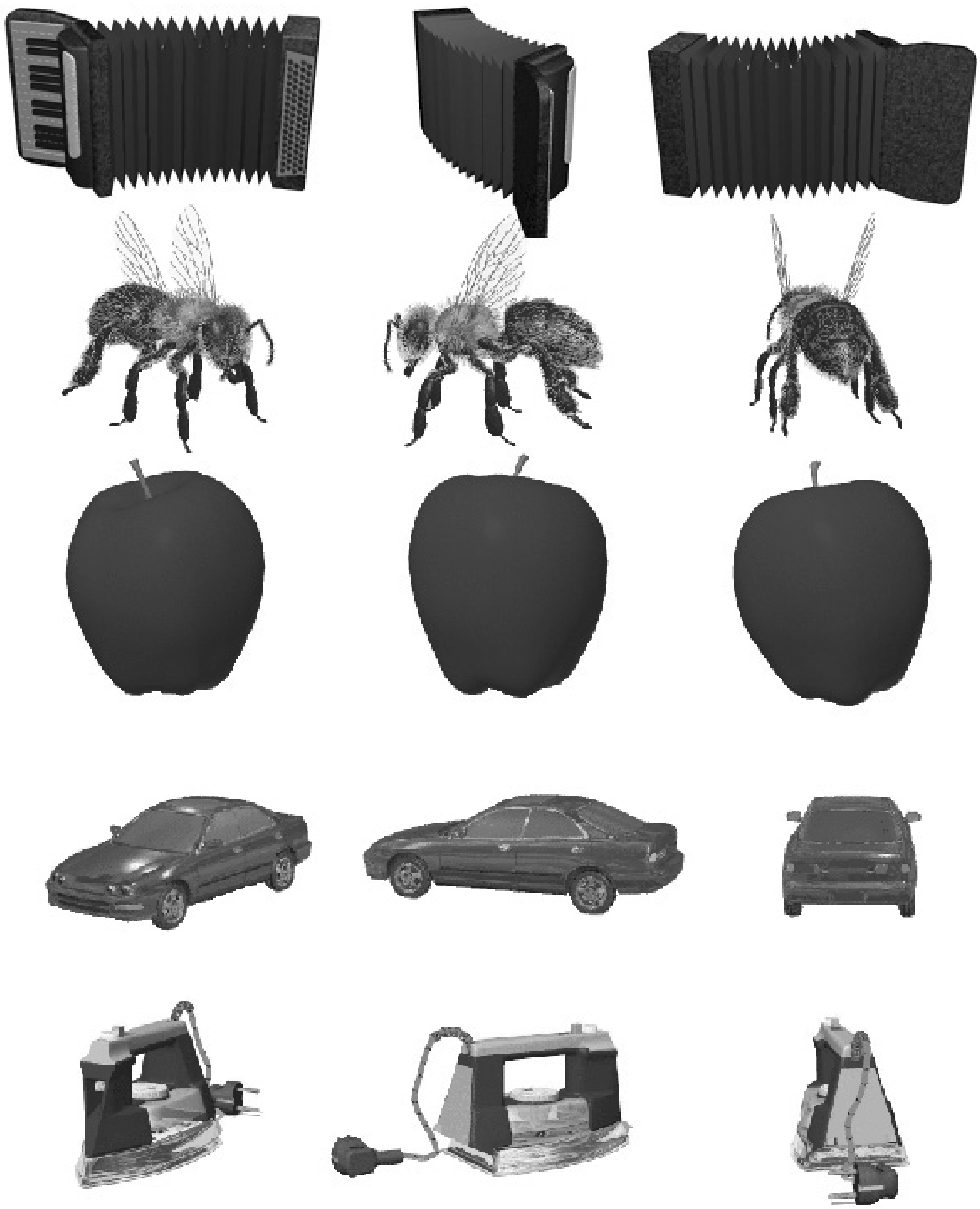


Figure 1. Sample stimuli from each of the 5 categories used in the experiment. From left to right: Canonical, intermediate, and non-canonical views.

Figure 2. Sample stimuli with COG: Intermediate accordion and canonical banana. Red circle marks the approximate COG.

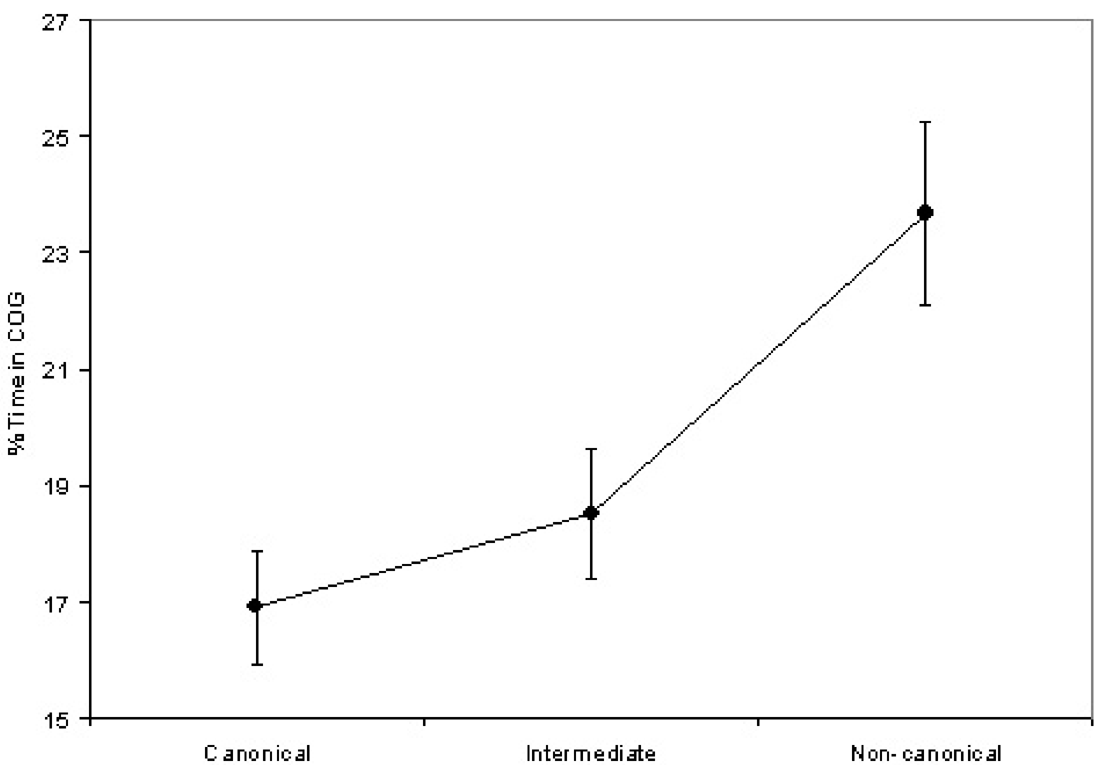


Figure 3. Average % time (across participants) of eye position in the COG. Bars represent standard errors.

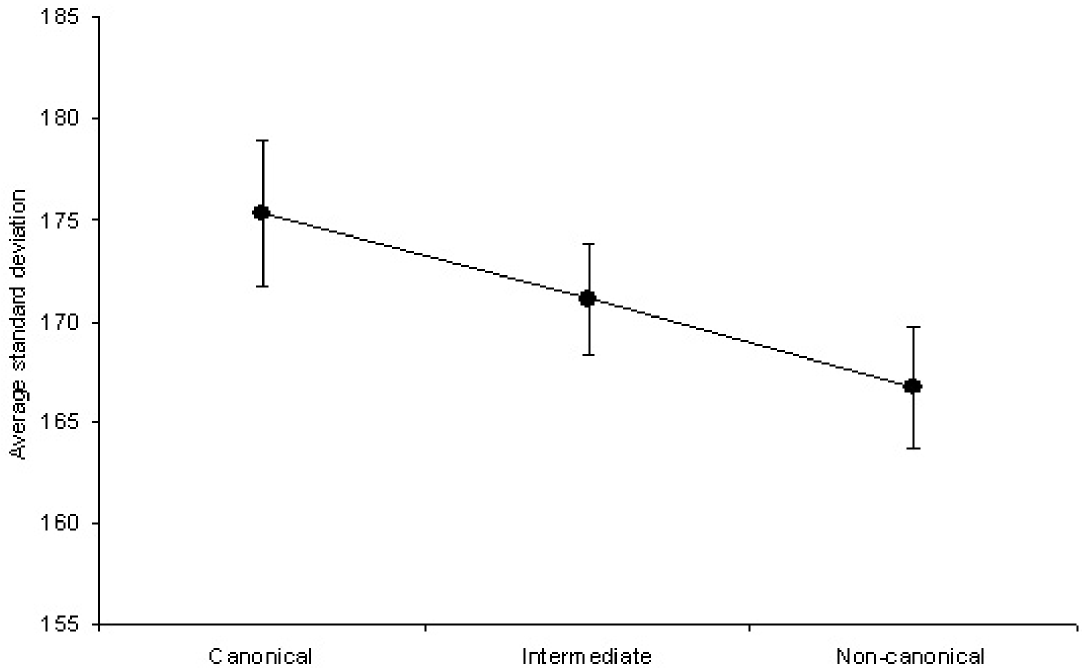
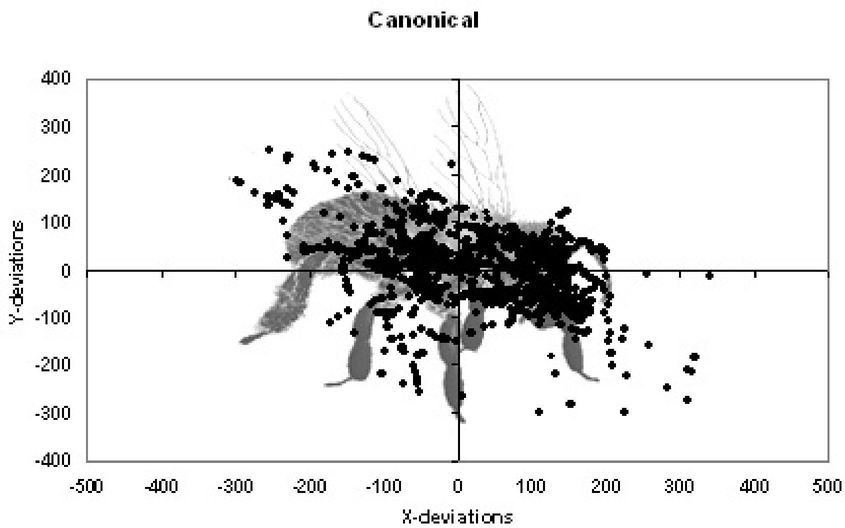
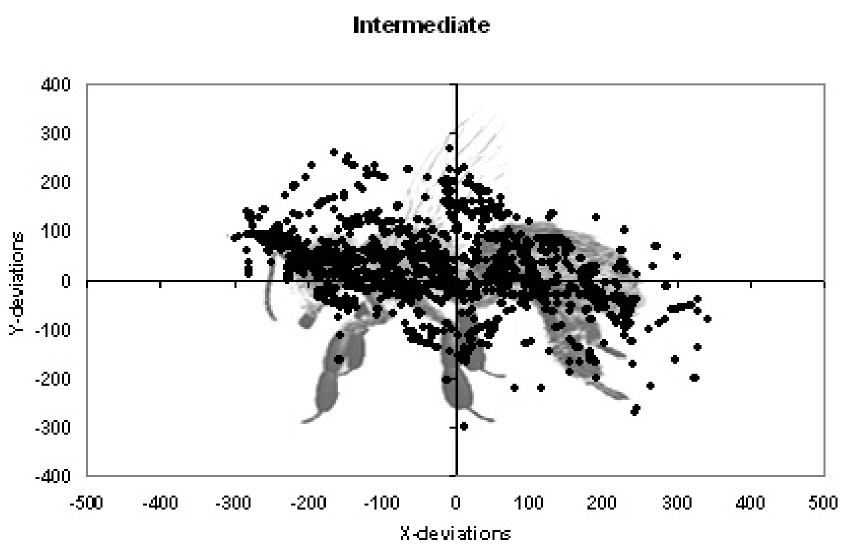
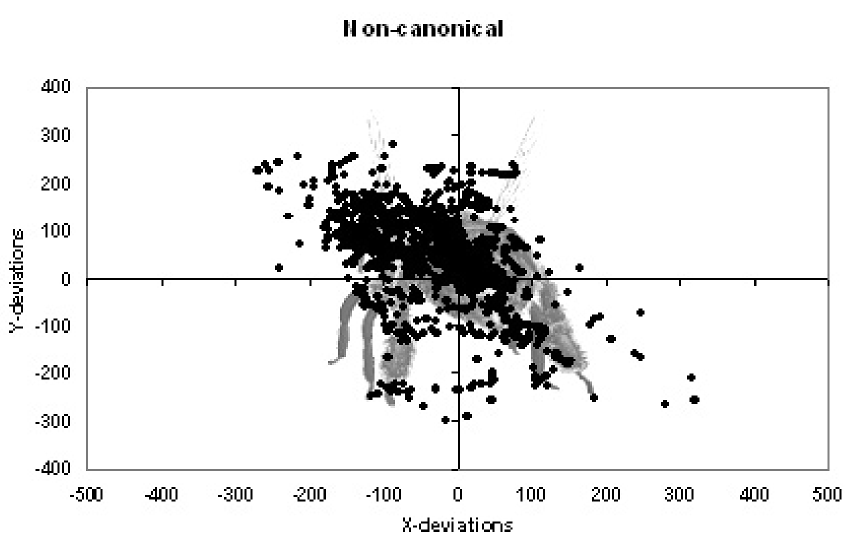
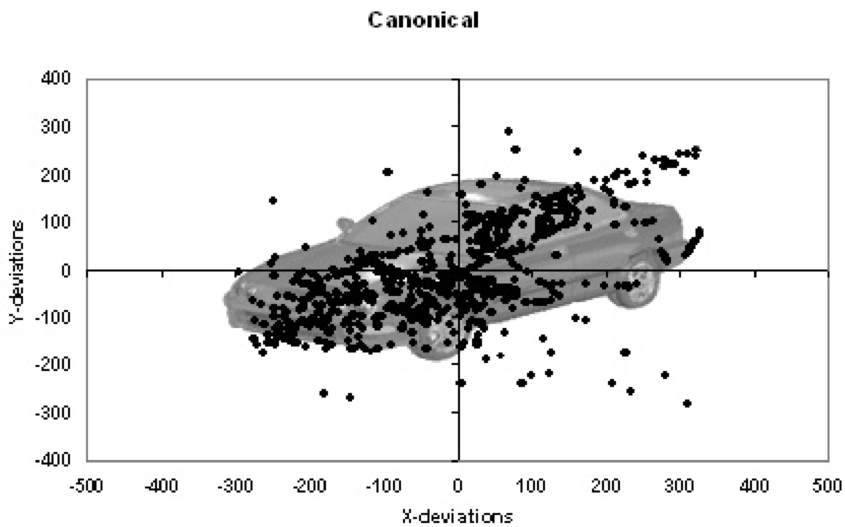


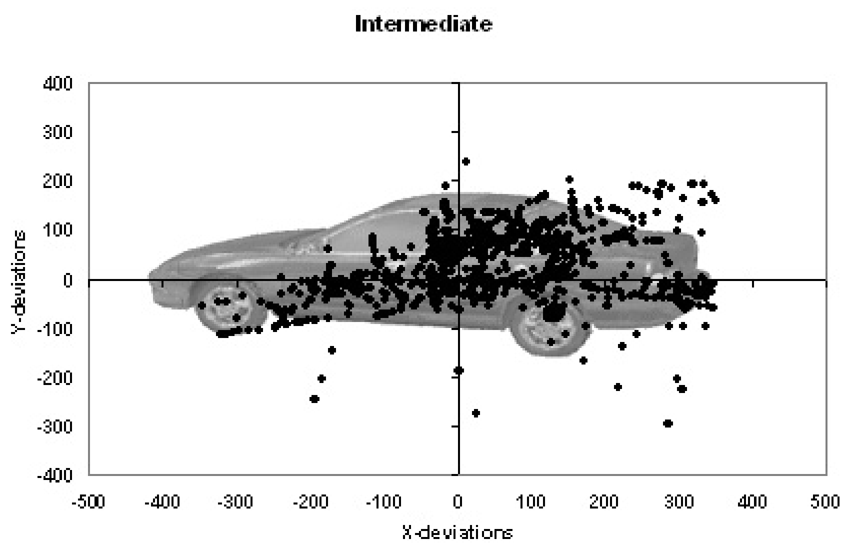
Figure 4. Average (across participants and coordinate axes) standard deviations in image pixels from the COG as a function of Rotation.


Canonical
Y-deviations
X-deviations
400
300
200
100
0
-100
-200
-300
-400
-500
-400
-300
-200
-100
0
100
200
300
400
500


Intermediate
Y-deviations
X-deviations
400
300
200
100
0
-100
-200
-300
-400
-500
-400
-300
-200
-100
0
100
200
300
400
500


Non-canonical
Y-deviations
X-deviations
400
300
200
100
0
-100
-200
-300
-400
-500
-400
-300
-200
-100
0
100
200
300
400
500


Canonical
Y-deviations
X-deviations
400
300
200
100
0
-100
-200
-300
-400
-500
-400
-300
-200
-100
0
100
200
300
400
500


Intermediate
Y-deviations
X-deviations

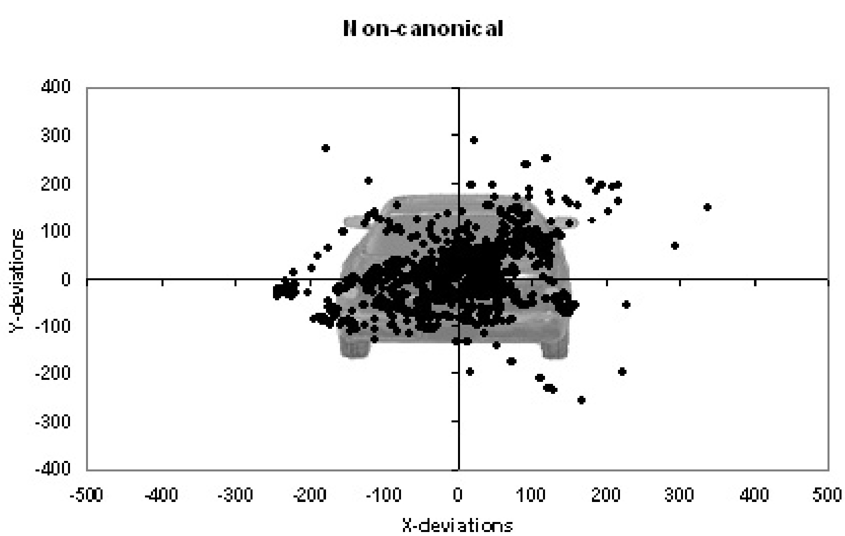

Non-canonical
Y-deviations
X-deviations

Figure 5. Eye positions of participants in each of the three rotation conditions when viewing two representative objects: A.) The honeybee. B.) The car. The origin of each panel corresponds to the object's COG in each rotation. Each point corresponds to the eye position (relative to the COG) for a 20 ms sampling period. The x-axis corresponds to horizontal position in image pixels and the y-axis corresponds to vertical position in image pixels. The object images are ghosted into each panel with the COG placed over the origin.

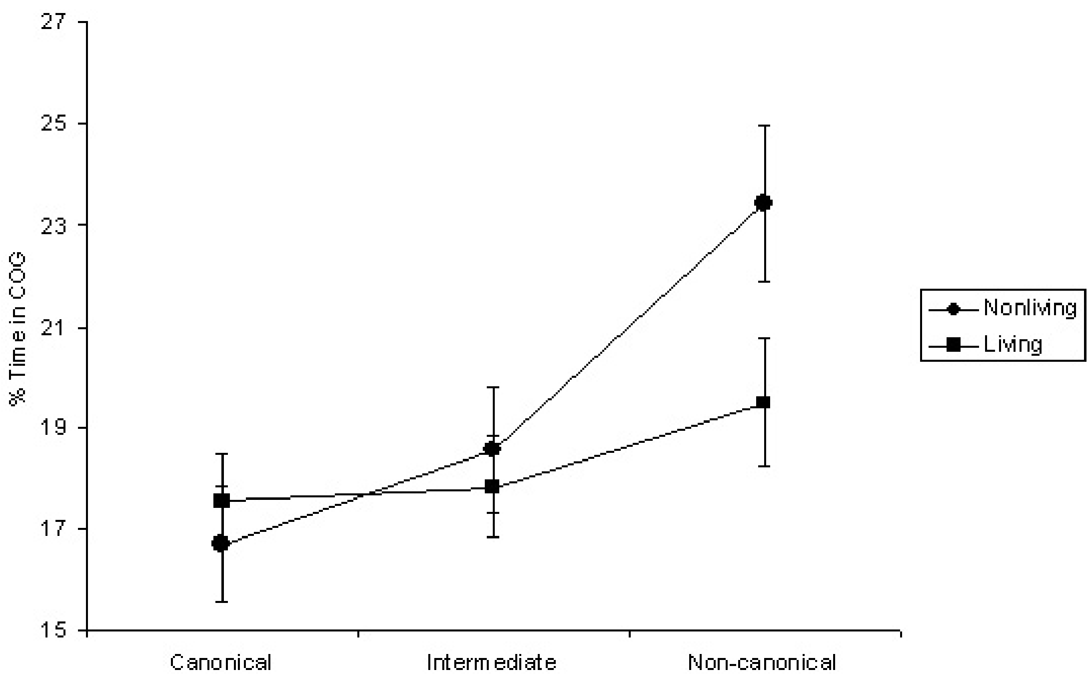


Figure 6. Average % time (across items) of eye position in the COG for living and nonliving things. Bars represent standard errors.

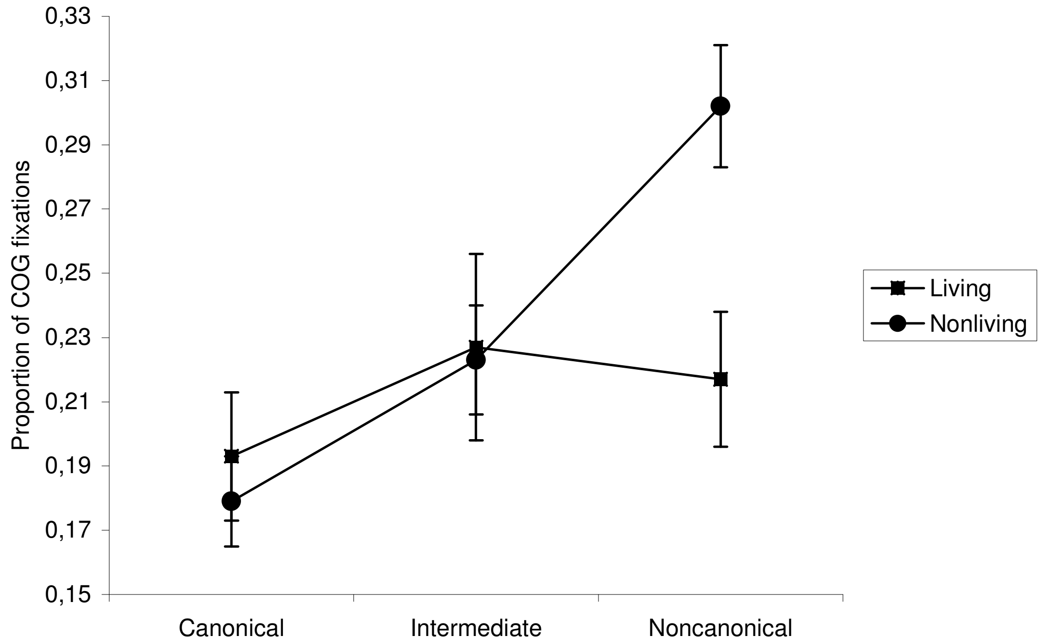


Figure 7. Average proportion of total trial fixations in the COG for living and nonliving things. Bars represent standard errors.

Table 1. Summary of hierarchical regression analysis using figure area and rotation as independent variables.

Factor	*B*	*SE B*	β
Step 1			
Figure area	-0.16	0.6	-.19*
Step 2			
Figure area	-0.13	0.6	-.16*
Rotation	-3.1	0.9	.25**

Note: R^2 = .036 for Step 1; ΔR^2 = 0.60 for Step 2 (p = .001), *p < .05, **p < .01