Active manual control of object views facilitates visual recognition

Karin L. Harman, G. Keith Humphrey and Melvyn A. Goodale

Active exploration of large-scale environments leads to better learning of spatial layout than does passive observation [1–3]. But active exploration might also help us to remember the appearance of individual objects in a scene. In fact, when we encounter new objects, we often manipulate them so that they can be seen from a variety of perspectives. We present here the first evidence that active control of the visual input in this way facilitates later recognition of objects. Observers who actively rotated novel, three-dimensional objects on a computer screen later showed more efficient visual recognition than observers who passively viewed the exact same sequence of images of these virtual objects. During active exploration, the observers focused mainly on the 'side' or 'front' views of the objects (see also [4–6]). The results demonstrate that how an object is represented for later recognition is influenced by whether or not one controls the presentation of visual input during learning.

Address: Psychology Department, University of Western Ontario, London, Ontario N6A 5C2, Canada.

Correspondence: G. Keith Humphrey
E-mail: keith@julian.uwo.co

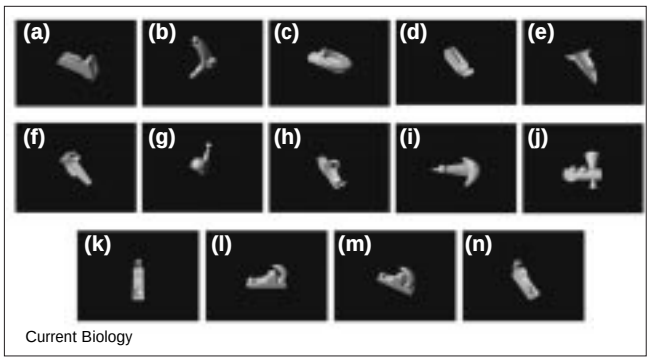
Received: 9 August 1999
Revised: 20 September 1999
Accepted: 8 October 1999

Published: 8 November 1999

Current Biology 1999, 9:1315–1318

0960-9822/99/$ – see front matter

Figure 1



(a–j) Examples of the novel, computer-rendered, three-dimensional objects used in the present study. (k–n) Examples of the views used during the old/new test session. (k) Front or foreshortened view, where the principal axis of elongation is perpendicular to the viewer's line of sight. (l) Side view, where the axis of elongation is parallel to the viewer's line of sight. (m) Three-quarter back view, a 45° or intermediate view between a side and a back view. (n) Three-quarter front view (sometimes called canonical view [7]), a 45° intermediate view between the front and the side views.

Results

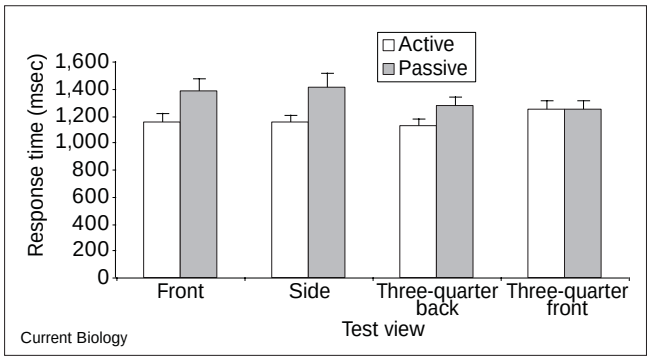
Recognition performance

We measured the response latency and accuracy of subjects as they performed an 'old/new' discrimination between two classes of object: ones they had seen during a study period and ones they had never seen before. The objects were novel, computer-generated, three-dimensional objects. As Figure 1 illustrates, the objects were constructed of 'geon'-like parts [7] and were elongated along a single axis. During the earlier study period, each subject viewed half the objects using active exploration and half using passive observation. A yoked-control design was used such that the passive viewing sequence for a particular object viewed by a subject during the study period was simply a 'replay' of an active exploration of that same object by another subject.

A within-subjects (between-objects) analysis of variance demonstrated that actively explored objects were recognized faster than were passively viewed objects ($F(1,21) = 16.1$, $p < 0.001$). As can be seen in Figure 2, the active–passive difference in the speed of recognition of the studied objects (that is, correctly responding 'old' to studied objects) was evident in three of the four different views of the objects that were presented during the old/new task. Specifically, active exploration facilitated recognition of the front ($p < 0.004$), side ($p < 0.003$) and the three-quarter back ($p < 0.03$) view of the objects. Speed of recognition of the other three-quarter view, the so-called canonical view, did not depend on whether or not the object had been studied actively. The same pattern of results was seen when a between-subjects analysis (yoked subjects, within objects) was carried out ($F(1,18) = 3.8$, $p < 0.02$).

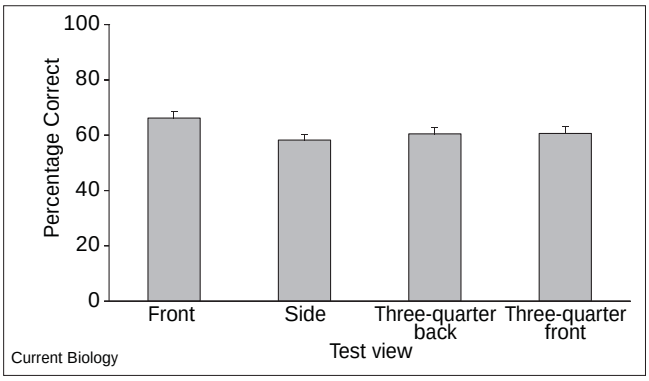
A within-subjects analyses of variance showed no effect of active exploration on the accuracy of recognition. Accuracy was, however, affected by the view of the object that was presented during the old/new task ($F(3,21) = 10.77$, $p < 0.0001$). As is evident in Figure 3, the front or foreshortened view of objects was recognized best. The same pattern of results was found with a between-subjects analysis ($F(3,20) = 9.9$, $p < 0.0001$). It is interesting that accuracy in general was quite low in both conditions, probably because of the similarity among the target and distracter

Figure 2



Response latencies to target objects during the test session. Actively studied objects were recognized faster than passively studied objects, except for the three-quarter front view. Note that generalization to a less-studied view (three-quarter back view, see also Figure 4) was greater for the active group than for the passive group. This generalization difference between the two study groups was, however, less pronounced for the three-quarter front view (see also Figure 4). Error bars indicate one standard error above the mean.

Figure 3



The percentage of correctly recognized target objects as a function of test angle. The front view is recognized more accurately than the other test views. Error bars indicate one standard error above the mean.

items. In fact, we deliberately designed the old/new task to be difficult so that we could increase the response latency enough to reveal a difference between study conditions. But why accuracy was not sensitive to the active–passive manipulation is unclear. Of course, continuous measures, such as reaction time, are often more sensitive than simple accuracy scores.

Exploration data analyses

We also examined how subjects distributed their looking time in the active exploration condition. In particular, we examined the amount of time that subjects spent on different views of the objects. We calculated peak dwell times for each subject and found that, when these values were averaged across subjects, a distinct pattern of exploration emerged. Rather than exploring the objects in an idiosyncratic manner, the subjects spent most of their time studying only four views of the objects, all of which were rotations about the vertical axis (see Figure 4). These four views corresponded to the front, back and two side views of the objects. Subjects tended to spend very little time studying particular intermediate views between these angles.

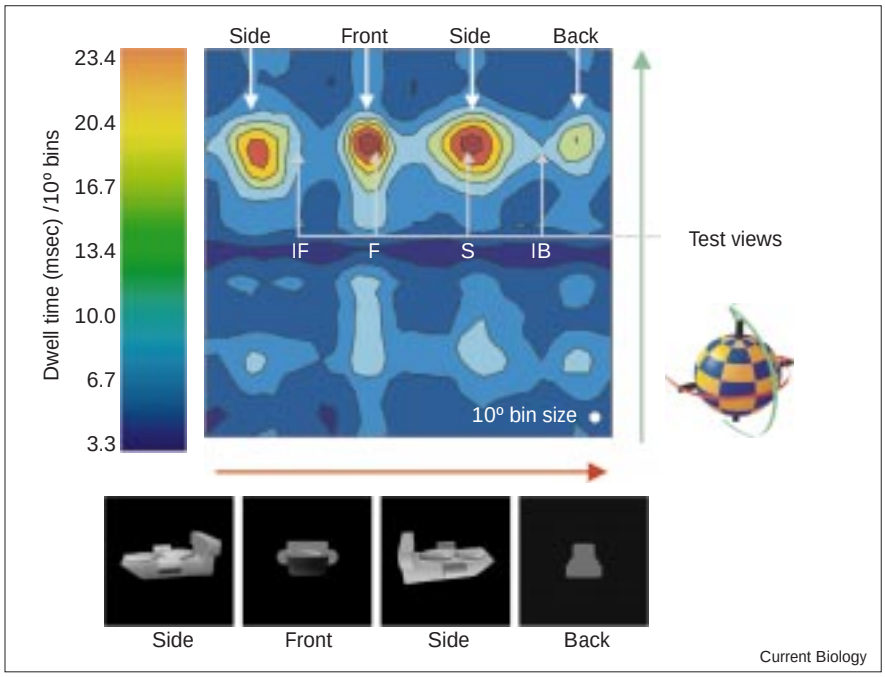
Discussion and conclusions

The results provide the first demonstration that active control of visual input during perceptual learning leads to more efficient object recognition. We found that subjects who actively rotated novel, three-dimensional objects on a computer screen recognized objects more rapidly than did subjects who passively viewed the exact same sequence of images of these virtual objects. In addition, we found that, when exploring such novel objects, subjects concentrated on particular views.

Although other studies have demonstrated that active exploration can improve scene recognition through the detection of changes in a stimulus array [3], our study provides convincing evidence that fundamental mechanisms mediating object recognition can be influenced by active exploration. In other words, active control over the way in which the different views of an object are revealed leads to faster recognition. Just why this occurs is not clear. It could be that direct manual control over the sequence of views provides efference copy and/or proprioceptive information (see also [3]) that helps to integrate the different views by allowing subjects to anticipate the upcoming view and relate it to the previous view. Alternatively, or at the same time, active exploration could allow subjects to test 'predictions' about the expected deformations in the image that would occur when the object is rotated in a particular way. The advantage observed with active exploration in our experiment might have depended critically on the fact that the movement of the object on the computer screen was, in some ways, an isomorphic reflection of the movement of the trackball. This relationship between visual input and manual control resembles, in some respects, the way in which we might visually inspect an actual object that we are holding in our hands.

Of course, integrating views and/or testing hypotheses about the structure of an object would involve attention. But attentional resources would not necessarily be distributed the same way in the two study conditions. In other words, subjects in the active exploration condition might have deployed their attention strategically — increasing their attention when a particular view of the object was on the screen. Indeed, they might have anticipated the need to increase their attention at this time. At other times, their attention might not have been as well focused. This strategic manipulation of attention would be expected to occur less often in the passive viewing condition where attention

Figure 4



Contour map depicting dwell times during the exploration (study) session. The map is a representation of the flattened viewing sphere (right). This particular map is a mean of all actively explored objects and all subjects. Red areas, higher dwell times; blue areas, lower dwell times. The top half of the map depicts dwell times about the vertical axis when objects are upright, the bottom half depicts dwell times when objects are inverted (most objects had a flat 'bottom' allowing us to determine upright and inverted orientations). The 'start' orientation is in the center of the map and is a view of the object from the top. Thus, the object required a rotation before it was in an 'upright' orientation. Therefore, the pattern in this figure could not be an artifact of starting position. The spatial resolution of the dwell time calculation was 10°. The spatial location on this map of the test views that were used are depicted with gray arrows: IF, intermediate (three-quarter) front view; F, front (foreshortened) view; S, side view; IB, intermediate (three-quarter) back view.

might be deployed more evenly over the entire sequence of views. In short, it is unlikely that a simple argument that subjects attended more in one condition than the other can account for the difference in performance.

Subjects spent more time inspecting certain views of the objects compared with others. In the active exploration condition, subjects tended to rotate the objects mainly around the axis that was perpendicular to the main axis of elongation of the objects. As a consequence, the object was rotated so that it moved between a fully elongated view to a completely foreshortened view. Subjects also treated the flat surface of the object as the 'bottom' and generally kept the objects oriented so that this surface was always face down on the monitor. Thus, both the geometry of the object and the convention of a top–bottom rule appeared to be driving the inspection strategies. The subjects constrained their viewing even more by concentrating on only a few particular views around the 'primary' (or chosen) axis of rotation. In particular, the front and side views received the most looking time. These two views represent ones in which the primary axis of elongation of the object is either perpendicular or parallel to the line of sight. Perrett and colleagues also found that, when subjects explored objects, they concentrated their inspection time on front and side views whether the objects were potatoes [4], heads [6] or machine-tooled 'widgets' [5].

Perrett *et al.* [5] have proposed that observers concentrate on 'plan' views (views in which the principal axis of the object is parallel or perpendicular to the line of sight), like front and side views, because these views are 'unstable' and can be thought of as singularities in the viewing space of an object. In other words, these are the views in which there is the greatest amount of change in the visibility of the object features as the object is rotated by a small amount. Inspection strategies that concentrate on such views would be important in coding these particular views. We can see now why subjects would not dwell on any particular intermediate views. The intermediate views are all perceptually similar: all the major features of the objects are visible over a wide range of image projections. Thus, subjects do not need to concentrate on one particular intermediate angle because of the high similarity among many of the successive images. This might explain why subjects deviate only a little from side to side when viewing a plan view; larger excursions would not produce much more information than they already have.

There is a long history of research that has investigated the role of various types of visual information in the representation and recognition of objects (for example [8]). For almost all of the accounts of object recognition that have grown out of this research, the observer has been assumed to be a passive recipient of information. Although some have studied the role of eye movements in directing the fovea to different parts of the visible object, little attention has been paid to the fact that, in the real world, the observer can actually manipulate the view of the object that is being scrutinized. The present experiment is the first to examine

the role of active manual control in object representation and recognition. Our results show that perceptual knowledge of objects is facilitated when one controls the sequence of images that convey the structure of the object. It now remains to determine just how this active exploration promotes more efficient perceptual learning.

Materials and methods

Subjects

Subjects were 22 right-handed undergraduate students (9 males, 13 females) ranging in age from 18 to 23 years (mean = 19 years) and all had normal or corrected to normal visual acuity.

Materials

Study stimuli were 20 computer-rendered images of novel, three-dimensional gray scale objects (see Figure 1a–j for examples). They were presented on a 15 inch computer monitor on a black background and were 'illuminated' with an ambient light source. Presentation of the images and recording of subjects' responses were controlled by a Macintosh G3 computer. The object images could be rotated by the subject about any axis using a 5 cm diameter trackball. All objects had a central axis of elongation and 'geon-like' parts [7] attached to a central body [9,10]. The object images were viewed from a distance of 60 cm. For the views in which the long axis of the object was perpendicular to the line of sight, the mean image size was 9 cm for the X dimension and 6 cm for the Y dimension. For images in which the axis of elongation of the object was parallel to the line of sight, the mean size was 5 cm for the X dimension and 6 cm for the Y dimension. During the active condition, subjects were free to rotate each object for 20 sec about any axis. During the passive viewing condition, subjects viewed a 20 sec recording of the previous subject's active exploration of that object. The data from the first subject was not used, but his active study was recorded and used as the passive component of the second subject's study session, thus the yoked design began with the second subject. The order in which the objects were presented for active and passive study was presented in a pseudo-random fashion and counterbalanced to eliminate any possible effects of order of study. Test stimuli were static images of four different orientations of each of the 40 objects (20 studied, 20 new), resulting in 160 test images. The four test angles (side, front, three-quarter front and three-quarter back) that were used are shown in Figure 1k–n. Note that these test images were two plan views (front and side) and two intermediate views (three-quarter front and back). As can be seen in Figure 4, two of these views were focused on during study, while the other two were not. These particular views were chosen because we were interested in investigating any recognition differences in the views that have classically been found to be difficult (front) and those that tend to be less difficult (three-quarter and side) to recognize [11].

Procedure

Study. During active exploration, subjects were told to explore each object, as they would be asked to recognize it during a test session. Subjects then moved the track ball with their right hand to rotate the object in a possible 360° on any axis on the computer screen. Although the subjects were allowed 20 sec in total to manipulate the objects, the rate at which they moved the objects was not controlled. That is, they could move the objects as fast or slowly as they wanted. During passive viewing, the subjects were told to study each object because they would be tested on their recognition during a test session. A subject's study of each object was initiated by the experimenter, which resulted in an average inter-trial interval of 7 sec. After studying each of the 20 objects (10 active blocked, then 10 passive, counterbalancing for order), the test session began.

Test. Each test trial was composed of a 1000 msec fixation cross, followed by a 100 msec blank screen and then the presentation of the test image. On appearance of the test image, subjects were required to press keys on a keyboard to indicate whether or not they had studied the particular object shown (an old/new decision). Response latency and accuracy were recorded by a Macintosh G3 computer. After the subject's response, an interval of 500 msec was followed by the next fixation cross signaling the next trial. This procedure continued until the subjects responded to the 20 'old' objects and the 20 'new' objects at the four test orientations.

Acknowledgements

The research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and from the Medical Research Council of Canada. Thanks to Tom James for his help with Matlab data analysis. Thanks also to Robert Sekuler and an anonymous referee for their helpful comments on an earlier version of this paper.

References

1. Held R, Hein A: Movement-produced stimulation in the development of visually guided behaviour. *J Comp Physiol Psychol* 1963, 56:872-876.
2. Held R, Freedman SJ: Plasticity in human sensorimotor control. *Science* 1963, 142:455-462.
3. Simons DJ, Wang RF: Perceiving real-world viewpoint changes. *Psychol Sci* 1998, 9:315-320.
4. Perrett DI, Harries MH: Characteristic views and the visual inspection of simple faceted and smooth objects: 'tetrahedra and potatoes'. *Percep* 1989, 17:703-720.
5. Perret DI, Harries MH, Looker S: Use of preferential inspection to define the viewing sphere and characteristic views of an arbitrary machined tool part. *Percep* 1992, 21:497-515.
6. Harries MH, Perrett DI, Lavender A: Preferential inspection of views of 3-D model heads. *Percep* 1991, 20:669-680.
7. Biederman I: Recognition-by-components: a theory of human image understanding. *Psychol Rev* 1989, 94:115-147.
8. Humphrey GK, Goodale MA, Jakobson LS, Servos P: The role of surface information in object recognition: studies of a visual form agnosic and normal subjects. *Percep* 1994, 23:1457-1481.
9. Humphrey GK, Khan SC: Recognizing novel views of 3D objects. *Can J Psychol* 1991, 46:170-190.
10. Harman KL, Humphrey GK: Encoding 'regular' and 'random' sequences of views of novel 3D objects. *Percep* 1999, 28:601-616.
11. Humphrey GK, Jolicoeur P: An examination of the effects of axis foreshortening, monocular depth cues and visual field on object identification. *Quart J Exp Psychol* 1993, *A* 46:137-159.