Max- Planck- Institut
für biologische Kybernetik
Spemannstraße 38 • 72076 Tübingen • Germany

Technical Report No. 68

# Viewpoint Information Provided by a Familiar Environment Facilitates Object Identification

Chris G. Christou[1], Bosco S. Tjan[2]
& Heinrich H. Bülthoff[3]

March, 1999

[1] Unilever Research, Wirral, UK.
[2] NEC Research Institute, Princeton, USA
[3] Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany

Address correspondence to:
Chris Christou, Unilever Research, Port Sunlight Laboratory, Quarry Road East, Bebington, Wirral L63 3JW, UK. Email: chris.christou@unilever.com. Tel. 0151 641 3104. Fax 0151 641 1841

# Viewpoint information provided by a familiar environment facilitates object identification

*Chris G. Christou, Bosco S. Tjan & Heinrich H. Bülthoff*

**Abstract.** We studied whether contextual information regarding an observer's location within a familiar scene could influence the identification of objects. The context was provided by a 3D virtual living room, which allowed natural familiarization of the scene and objects together with a high level of interactivity. Results of initial self-orientation judgments obtained in the room showed observers could make accurate judgments of their instantaneous orientation with respect to a reference point. We wanted to know if this information could in turn be used as an aid to identify objects from unfamiliar viewpoints. Our main experiment showed that after familiarization of objects within the virtual room, the presence of the room during identification produced significantly fewer errors than when the objects were shown in isolation. This reduction in error was attributed to the provision of a consistent reference frame by the room. This was tested by a control experiment, in which we randomly varied the orientation of the objects with respect to the room. In this case, the observer's relative orientation with respect to the objects could not be derived from the room. Results showed that recognition accuracy dropped significantly in this case. The results in general suggest that object identification can be aided by knowledge of where we are in space and in which direction we are looking.

## Introduction

One of the most intriguing abilities of human vision is the apparent ease with which it enables the identification of complex three-dimensional (3-D) objects. Visual identification in a simplistic sense can be construed as matching the retinal image of an object against stored mental representations. Exact matches are unlikely to occur, owing to variations in an observer's viewpoint and changes in the visual environment in the background. Robustness of the identification process is therefore indicated by its ability to overcome uncertainty introduced by deviations of the retinal image as compared to a mental representation. In this paper we ask how well people can cope with deviations brought about by changes in the observer's viewing perspective. We extend previous studies on this topic by considering whether subsidiary information regarding the viewing context could help the identification process, considering that objects in a natural environment never appear against a featureless background. This subsidiary information is implicitly specified by the visual background to the object. For observers who are familiar with a scene, a single view of the scene is sufficient for them to determine where they are in the scene and where they are looking at. We wanted to know whether such information about viewpoint, extrinsic to a target object, could enhance object identification. In the past, object identification has been studied in isolation and contextual studies, although possible (c.f. Wang & Simons, 1998; Shelton & McNamara, 1997), were difficult to perform because of the problems in moving observers around in real scenes. We overcome this by using a virtual-reality simulation in which both visual realism and interactivity are used to mimic a realistic learning scenario.

In order to assess how well people can encode the geometry of newly learned objects, Rock & DiVita (1987) showed observers twisted wire-like shapes from one direction and subsequently tested generalization to novel views by incrementally rotating these objects in depth (see Figure 1). They found that recognition performance negatively correlated with the amount of rotation. That is, errors increased as a function of miss-orientation from the familiar view. These so-called 'wire-frame' or 'paper-clip' objects reduced the significance of self occlusion and allowed us to focus on the question of whether geometrical detail encoded from one view of the object could be used to recognize the same object from another view. This lack of

generalization to novel views, or view specificity, was taken to mean that spatial encoding is egocentric in nature; that is, spatial detail is encoded within a viewer-centred reference frame as opposed to an object-centred or world-centred reference frame.

More recent experiments using images of computer generated 3-D simulations of 'paper-clip' objects have revealed similar view-dependent recognition performance in both humans (Bülthoff & Edelman, 1992) and monkeys (Logothetis, Pauls, Bülthoff & Poggio, 1994). Furthermore, Bülthoff & Edelman (1992) also showed that one's ability to recognize an object is a function of the angular distance from familiar views. Bülthoff & Edelman's results suggest that the principal means of overcoming changes in orientation is by *interpolation* of detail between known views. This 'image-based' account of the recognition process is supported by similar view-dependent results using other kinds of objects (e.g., Tarr, 1995) and also by the isolation of 'view-tuned' neurons in the inferior-temporal cortex of the monkey (Logothetis et al., 1994). What is clear from these studies is that recognition performance for these objects depends on familiarity of view. One possibility is that view-dependency results from the observers not knowing their viewpoint with respect to the test object. In deciding whether a given stimulus is a view of a familiar object, observers may benefit from knowing from where in the scene they are looking. The reason why poor performance has been observed for recognition from novel views may be because previous experiments have studied object recognition without a visual background.

Under natural circumstances, information regarding self-orientation in space is derived from several sources, such as vestibular cues, proprioceptive feedback, and visual inputs. In the last case visual features of the environment can be used to specify an observer's relative location and orientation within that environment. Neurological properties of the encoding of spatial location and viewing direction is steadily accumulating since the discovery of 'place-cells' in rats (OKeefe & Dostrovsky, 1971) which were later shown to be active only when the rat is in specific locations of a familiar environment (OKeefe & Conway, 1978). Furthermore, Taube, Muller & Ranck (1990) have also identified direction-sensitive neurons in rats, which are active only when the rats point their head in a particular direction in an environment. When these rats are kept in the dark, this specialization disappears and must be re-established by visual experience.

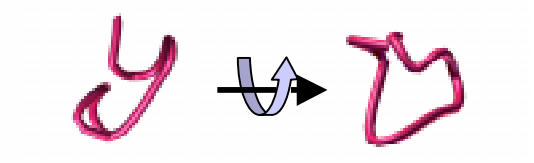

***Figure 1.*** *Two views of the same object. A counter-clockwise rotation of 90º about the indicated axis will rotate the left view into the right view.*

Complementary research in humans is more difficult to obtain and requires non-invasive methods. For example, Maguire et al. (1998) used Positron Emission Tomography (PET) to identify the involvement of brain structures during human observers' navigation through a virtual environment. Although space encoding and viewing direction are clearly important in navigation, there has been no research concerning the interaction between disparate sources of information such as 'where am I' and 'what is this object'. A positive use for such information is demonstrated in Figure 1. Verifying whether these two images of wire-like objects derive from the same 3D object is difficult. However, the task is made much easier if the reader is told that the only possible transformation is a 90° rotation about the horizontal axis. (The two images do derive from the same object.) The additional information that we have supplied specifies the intermediate transformation. Most studies in object recognition have assumed that such transformations originate from the objects themselves. Furthermore, they have assumed that transformations in view brought about by rotations of an object and rotations of the observer are equivalent. While this may be the case in a void, ordinarily object learning and identification take place within a context – a 3D environment. The environmental context could therefore tell observers where they are stationed and to where they are looking. In the absence of volitional information derived from the observer's own movements, a visual environment could provide information about the transformation used in the example of Figure 1.

From a different perspective, a number

of researchers have investigated the facilitation of object identification by consistent context. For instance it has been reported that object detection (Biederman, 1972, 1981) and object identification (Palmer, 1975) is facilitated by the presence of consistent context. Thus, the context of a kitchen scene facilitates the recognition of a loaf of bread, for example. The implication of these results is that top-down knowledge can influence the matching between retinal stimulation and mental representation. However, Hollingworth & Henderson (1998) suggest that this apparent facilitation is only the result of response biases and that when methods which eliminate response biases are used no facilitation is observed.

Other benefits of having visual context may derive from the ability to account for the distorted appearance of objects. For instance, Humphrey & Jolicoeur (1993) found that presenting foreshortened views of familiar objects on backgrounds with strong depth cues improves recognition performance. Even so, many of the recent experiments on object recognition and identification have tended to isolate objects from any particular scenic context. The assumption probably being that the segmentation of objects in complex scenes is a hindrance, which increases the complexity of the recognition or identification process. However, no studies have looked at the natural, interactive, means by which objects and scenes are connected and the specific reference frame imposed by a scene where the 'connectedness' of object and scene is established over time. The existence of a spatial reference frame is important both for encoding and recognition of objects (e.g. Marr, 1982; Feldman, 1985). If people customarily experience objects in a given spatial relationship with their surround environment then perhaps they can use their spatial location within the environment to aid identification.

In the experiments reported here, we tested if contextual and positional knowledge (viewpoint) obtained from the environment can help in reducing or eliminating the view-dependency observed in object identification. An observer's viewpoint with respect to a given object is determined by the observer's spatial position in the scene and by the observer's viewing-direction. Both of these can be derived by interacting with an environment. Perceived changes in viewpoint can in principle be obtained from two sources: the observer's bodily movements (kinesthesis) and the visual appearance (perspective) of the environment. The latter is possible because a 3D environment imposes a reference frame that can be used to specify an observer's relative viewpoint much the same way as the position of the camera can be derived from the contents of a photograph. This however implies established knowledge of the spatial layout of the environment, which is essential in relating a view of the scene to a spatial location of the observer. In these experiments we established this familiarization with the test environment by using a Virtual Environment Simulation in which a richly decorated room could be explored. Our first task was to determine if changes in view with respect to some chosen reference direction could indeed be assessed. This proved to be the case. The second and third experiments utilized a naturalistic long-term object-encoding task and determined if information about viewpoint could benefit object identification from novel views.

## General Methodology

### *The Virtual Environment*

In the past, spatial cognition experiments involving animals have taken place in specially constructed mazes or environments. The use of real environments for studying spatial cognition in humans however is complicated both by the need to control subsidiary cues or is usually impractical owing to the required dimensions of the environment. The use of Virtual Environment Simulations (VES) is becoming an increasingly popular alternative (e.g., Maguire et al., 1998; Péruch, Vercher & Gauthier, 1995; Christou & Bülthoff, 1999). For these experiments we devised a learning and test paradigm based entirely within a simulated, richly decorated, model of a natural environment, a living room (Figure 2). It seems reasonable to assume that the contextual richness of an environment enhances its memorization. Also, realistic textural detail and illumination improve 3-D cues. The room was created using 3D Studio Max (Kinetix, USA), a 3-D modeling program that allowed us to incorporate realistic furniture and fittings. Furthermore, the illumination was realistically modeled using Lightscape (Discreet Logic) which simulates the inter-reflective nature of light reflection. A further benefit of this software was that luminance values could be

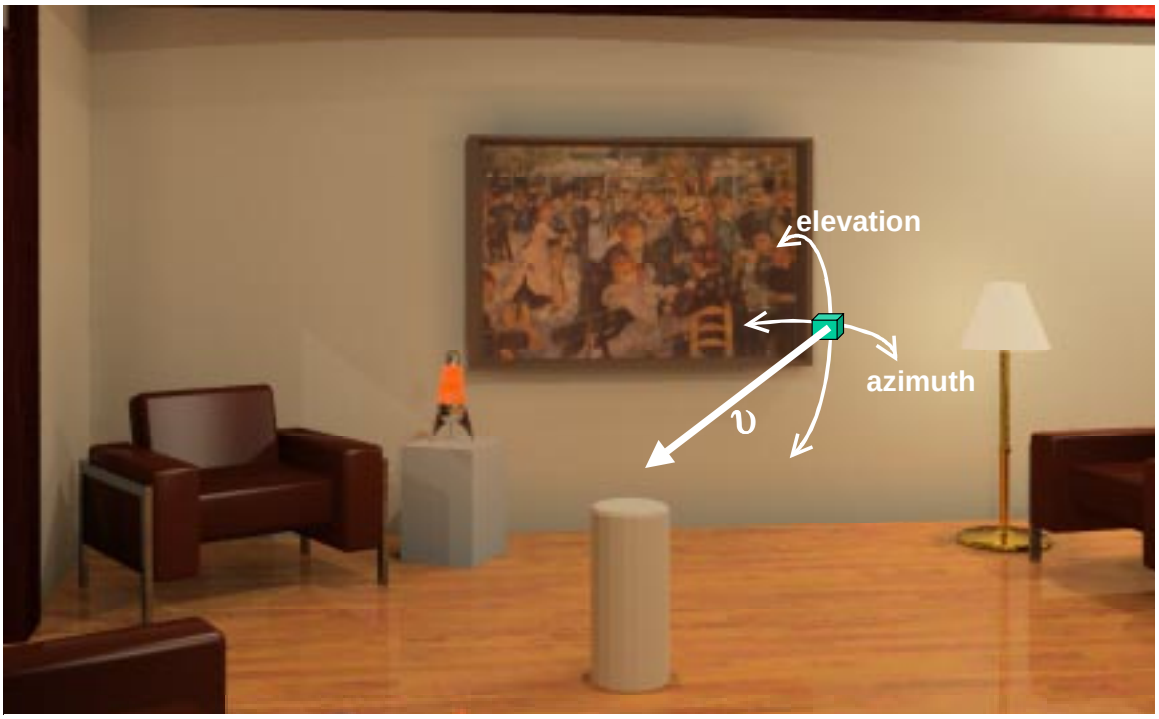


***Figure 2.*** *Rendered image of the Virtual Environment used in the experiments showing the pedestal around which simulated movements were performed. The instantaneous viewpoint of the observer is specified by* $\upsilon$ *which was always tethered to a point just above the pedestal. Changing the azimuth (horizontal) angle and the angle of elevation altered the viewpoint.*

calculated just once for each polygonal facet comprising the scene. This greatly enhanced the rendering speed during the experiment. In total, the final 3D model consisted of around 32,000 polygonal facets.

### *Novel Objects: Paperclips*

The target objects were computer generated 3-D geometric forms that consisted of 10 cylindrical segments starting from a vertical 'stem'. In order to produce geometrical forms that were similar, the 'arms' of each paperclip were produced by joining, with cylindrical segments, 9 points on the surface of an imaginary sphere. The surface of the sphere was subdivided into 8 equal sized sectors and each point could occur at a random azimuth and elevation within the bounds of one of these sectors. The ninth point always occurred at the base of the sphere. The order in which each of the points were connected was fixed, thus producing similar overall shapes for all targets. A set of 32 objects was generated separately and with random variations, although noticeably degenerate examples were replaced. Such degeneracy, for example, included intersections or near-intersections of arms.

### *Interactive Manipulation of View*

Self-control of movement was an important attribute in these experiments. Active movement through the environment allowed observers to form a better impression of its spatial layout. Active manipulation of view around objects allowed observers to perform natural behavioral patterns of examination and familiarization. Interactive control of simulated observer movement was facilitated using real-time computer graphics (implemented with SGI's IRIS Performer libraries). Observer movements were input using a Logitech SpaceMouse (Spacetec, USA). Users could control movement within the simulated room by applying pressure on the SpaceMouse in the direction they wished to move. Thus, pushing the cap forward moved observers forward, pushing to the left moved their simulated position to the left, etc. The users were also able to change their heading direction (i.e. tilt their view towards the ground) by applying differential force on one side of the

SpaceMouse cap.

### *Viewing Conditions and Stimuli*

Dynamic views of the scene consisting of 1280x1024 pixels were presented in 24-bit color across the entire drawing area of a RGB monitor housed within a view-reduction box. The box was tapered toward the viewer. Two rectangular holes separated by a septum allowed visual access and restricted viewing to the central portion of the screen. Viewing distance for all experiments was constant at 80cm producing a visual angle of approximately 34.5° at the eye. The video update frequency (i.e., the number of refresh updates of the scene) was 60 Hz.

## Experiment 1 – Orientation Judgments

This experiment sought to determine how well observers were able to judge shifts in viewpoint based on the content of the visual environment. In essence, this involved making judgements concerning the observer's movements (in azimuth, see Figure 2) within the simulated room implied by the contents of two images of the room. Constraints were imposed on how these images were produced. These constraints were consistent with an observer tethered to a point just above the pedestal in the center of the room (see Figure 2) and where the distance between observer and pedestal was kept constant. We wanted to gauge the magnitude of errors made in judging shifts away from a reference viewpoint.

Prior to the experiment, observers were familiarized with the room by performing simulated movements through the environment. This familiarization process was considered highly important given that the observers had to make judgements of changes in view based on static visual information.

### *Observers*

Seven people between 18 and 25 years of age were recruited for the experiments. They received payment for each hour of participation. In all cases, observers received verbal instruction and they performed an initial two blocks of settings to familiarize them with the task. Data from these preliminary trials were discarded from the analysis. Procedure

### *Procedure*

Following an initial familiarization exercise in the virtual room (see "room familiarization" procedure in Experiment 2 for detail), observers made judgments of orientation shift between two display intervals. The first interval consisted of an image from a reference viewing direction within the room under the constraints mentioned previously. This reference direction was randomly chosen for each block of 24 trials. The first interval was presented for 1 second. The second interval in each trial consisted of a 500 ms presentation of another view of the room after an intermediate change in both azimuth and elevation with respect to the pedestal. These new views were therefore generated by moving the observer's viewpoint across the surface of an imaginary viewing sphere centered just above the pedestal. The azimuth could change by an amount up to ±180 degrees from the reference direction. The elevation differences varied between ±15 degrees. Observers had to judge only the intervening shift in azimuth, while ignoring the change in elevation. Observers responded by adjusting an on-screen dial which was presented immediately after the second interval. The observers were told that the dial represented a plan view of the room with the familiar direction clearly defined by the vertical 'up' direction. Observers set the dial to indicate the viewing direction of the second interval relative to the first by turning the cap of the SpaceMouse. Each subject performed 3 sessions of the experiment. Each session consisted of 3 blocks of 24 trials. A new reference view point and direction was chosen for each block. The results for each subject were averages of 216 judgments in total.

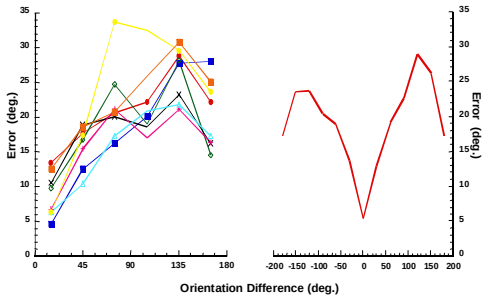


***Figure 3.*** *Graphs showing errors for viewpoint displacement judgments. The graph on the left shows the absolute error plotted individually for all 8 observer against actual displacement. The graph on the right shows the signed error averaged across all observers.*

### *Results*

The results were assessed in terms of the angular difference between the true shift in azimuth and observer's settings. The data for all observers as a function of actual displacement from the reference directions are shown in Figure 3. The graph on the left shows that the absolute error profiles appear quite similar across observer, with an initial increased error with increasing orientation difference followed by a decrease approaching 180 degrees. This indicates that the opposite view of the reference direction was more easily detected than oblique deviations from the reference direction. This is further illustrated by the signed means collapsed across all observers shown on the right panel of Figure 3. Some observers also showed enhanced performance for 90-degree views, which perhaps indicates a special significance for all canonical directions in egocentric space (i.e. front, back, left and right). Overall, the mean absolute error was smallest (about 8 degrees) within the first 30 degree interval. This is probably due to the overlapping of visual features across such stimulus pairs. Overall, errors were always less than 45 degrees. This shows that although the ability of judging viewpoint changes is not particularly fine-tuned and varies with actual displacement, it is still possible to infer one's displacement in view with respect to a fixed reference direction. In essence, these results show that it is possible to make gross judgements of relative view (e.g. front, back, left and right) based only on the content of a single image. The next experiment assessed whether such knowledge could be used as an aid to object identification.

## Experiment 2 – Effects of Context

The purpose of this experiment was to provide some indication of whether people can use viewpoint information of the kind described previously in order to correctly identify newly learned geometrical objects from novel views. To do this, we used 'paper-clip' or wire-like forms (see Figure 5) whose identification has been identified as view dependent in previous experiments. We devised a naturalistic learning paradigm in which observer had to first learn the layout of the virtual room and then learn to identify four of these objects which could be individually displayed on top of the pedestal in the middle of the room. Our observer could rotate around the objects and therefore benefited from shape cues such as motion-parallax. Training was the same for all observers and for all conditions. Testing took one of two forms: Either the room remained visible or was replaced with a light-gray background during identification of the target objects.

### *Procedure*

The experiment involved essentially four stages: (1) room familiarization by simulated locomotion through the scene, (2) interactive object learning, (3) criterion test, and (4) main test. The latter three stages (shown in Figure 4) were repeated three times in succession for each block. The initial locomotion stage of each block familiarized observers with the 3D spatial layout of the environment from all perspectives. In order to encourage observer to explore the room they were instructed to locate and acknowledge randomly positioned two-digit codes. These codes only became visible when viewed closer than one (simulated) meter.

After approximately three minutes of self-locomotion, the training stage commenced automatically. Each block involved the familiarization with four new 'paperclips' rendered in real-time resting on the top of the pedestal in the middle of the room (see Figure 5). With four fingers of their preferred hand placed on four buttons of a computer keyboard, the observers could 'flip' between each of the four objects which was immediately displayed on the pedestal. Thus, they learned to associate each object with each finger. With their other hand they could manipulate the SpaceMouse which allowed them to change their simulated viewpoint of these objects within bounds. The nature of this movement was analogous to being

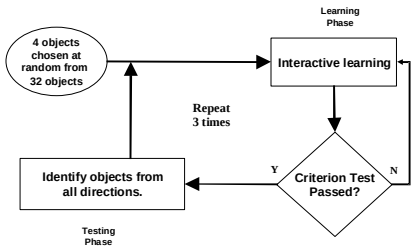


***Figure 4.*** *Depicts the various components of the experiment in flow-chart form.*

tethered to an invisible point just above the pedestal (as in the previous experiment) and with the freedom to alter azimuth and elevation by up to 15 degrees to either side of a reference viewing direction. This reference direction was randomly chosen for each block.

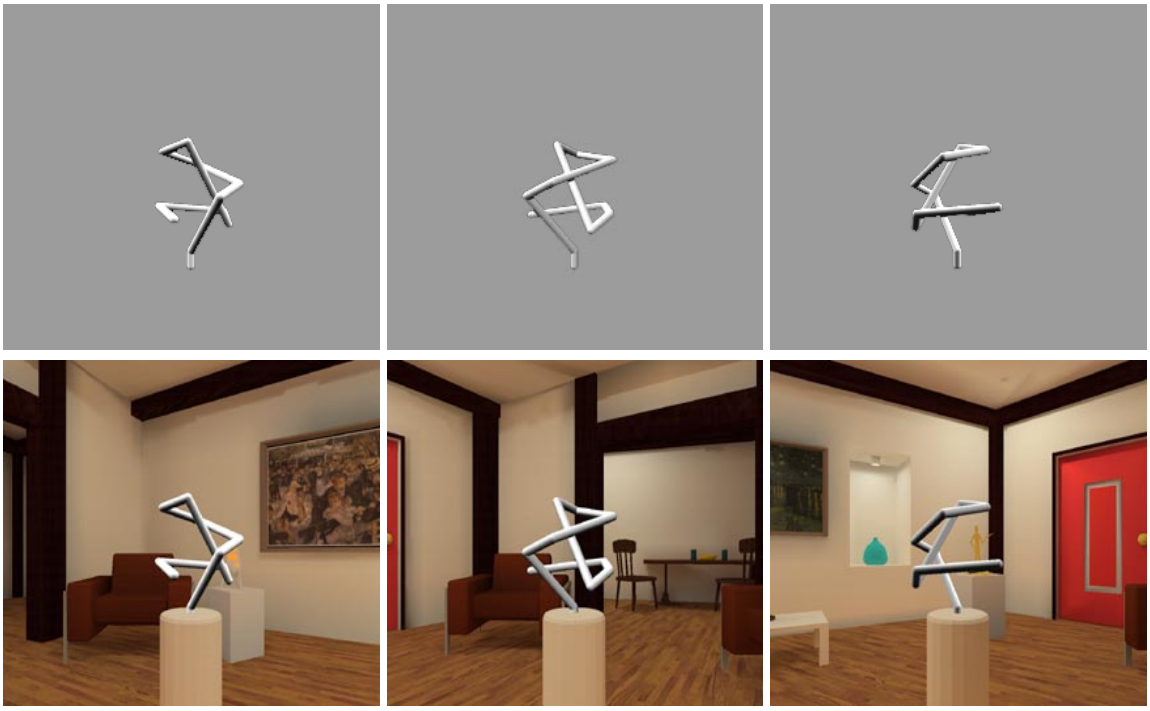

***Figure 5.*** *Example images for each of the two conditions in Experiment 1. From left to right the images show the same paperclip after observer rotates by 0º, 90º and 180º around it. These correspond to the front, right and back views of the object. The images for the room condition portrayed in the second row show appropriate changes in background consistent with the observer's movements.*

The learning stage lasted two minutes after which time the subject performed a criterion test. In this test they received randomly chosen static views of each object within the bounds of their movement during learning (e.g., ±15° for azimuth and elevation) and had to identify each. In each trial, a view of the 'context' (i.e., vacant pedestal and room), determined by the randomly chosen viewpoint, would be presented for an indefinite period until the subject initiated the display of the test object by pressing a computer key. The test object was displayed for 500ms after which time it disappeared leaving only the background context again. Observers were allowed to make a response as soon as the test object was displayed.

Each object was presented 4 times. This made a total of 16 trials for each criterion test. Observers passed the criterion test if 14 out of 16 trials were correct. Otherwise, they were again placed at the training stage with the same object set and reference viewing direction. The number of attempts required by observers to pass the criterion tests was recorded.

Once the criterion test was passed we could be sure that observers were able to differentiate between the four objects from a familiar range of viewpoints. We then tested this ability from all viewpoints around the objects in an identification test. This test was in all other respects similar to the criterion test. However, for room-absent trials observers saw only a blank screen before, during, and after the presentation of the object; that is, the room was absent from view. Each object was shown to observers 12 times; once from each of 12 positions around the object, where each position differed from the familiar azimuth direction by a multiple of 30° around the pedestal. From each position a further randomly chosen offset of between ±15° in azimuth and elevation perturbed the viewing direction. This reduced the possibility that observer might rely on accidental features of any given view.

Criterion test and main test were repeated three times to assess learning effects. Learning was measured primarily by the number of attempts required in passing the criterion test for each set of four objects. In total, each block consisted of 3 (repeats) x 4 (objects) x 12 (orientations) = 144 trials. There were four blocks of four new objects for each room-present/room-absent treatment, resulting in a combined total of 4x144=576 trials per room condition.

### Design

A two-factor repeated-measures design was used. The two factors were (1) room presence during test (two levels: present/absent. The room was always present during learning) and (2) orientation shift in viewpoint (i.e. difference in azimuth between new viewpoint and studied/ reference viewpoint). The latter consisted of six levels corresponding to the mean of each of six 30-degree bins in which view changes in azimuth were collected (i.e., with mean azimuths at ±15°, ±45°, ±75°, ±105°, ±135°, 165°). Percentage-correct identifications were averaged within each bin for each observer. The experiment for each observer consisted of 8 blocks, evenly divided between the room-present/room-absent conditions. Blocks involving room-present and room-absent trials were randomly interleaved.

### Observers

The 11 observers were aged between 17 and 31 years and were paid for each hour of participation. All were given prior instruction in all conditions of the experiment and in the use of the SpaceMouse. Two of the observers participated in the view displacement experiment described previously. All observers were naïve as to the purposes of the current experiment and performed this experiment for the first time. They were instructed to use any means to discriminate between the objects shown to them and any method of identification that maximized correct responses. They were also instructed to respond as quickly and as accurately as possible.

### Results

#### Criterion Tests

The criterion test always involved the presentation of the test object within the room context. The number of successive attempts at the criterion test before progression to the main identification test varied as a function of the block number for each session. Table 1 shows that the first criterion test in each block was always the hardest to pass, understandably. Because training always involved the presence of the room context no difference was expected according to whether room was present or absent during main test. Table 1 shows that this is indeed the case. An analysis of variance (ANOVA) with room presence during main test and block number (1, 2 or 3) as factors showed that the effect of block on number of attempts was significant ($F(2,20) = 18.75$, $p<0.0001$). The presence or absence of the room during main test had no effect on learning ($F(1,10)=0.05$, ns).

| *Room* | *Block 1* | *Block 2* | *Block 3* |
|---|---|---|---|
| *Present* | *2.6* | *1.3* | *1.3* |
| *Absent* | *2.4* | *1.6* | *1.1* |

***Table 1.*** *Shows the average number of attempts before passing the criterion. Results are tabulated according to the room-present/room-absent test blocks for the first, second and third blocks, averaged across all sets of four objects. We expected no difference between the room-present/room-absent conditions because the training phase always the same (i.e. room was always present during criterion test)*

#### Main Identification Tests

Responses not made within 4 seconds of each presentation during identification tests were discarded from the analysis (approximately 5%). The data for each subject was averaged within each of the six bins of mean-azimuth changes . The data were also collapsed across changes in elevation. The response times (RT) for correct responses were grouped in a similar manner. Figure 6 shows the proportion of errors and RT averages as a function of orientation difference from the reference view. For both conditions, errors increased as a function of orientation shift, reached a maximum around 90° and began to drop approaching the rear view of the objects.

This relationship is also depicted by the RTs. The pattern of errors for the two 'room' conditions are also clearly different, with the room-present condition producing significantly fewer errors for all orientations than that of the room-absent condition. An analysis of variance (ANOVA) with orientation and room-presence as within-observer factors showed an overall main effect of orientation ($F(5,50)=44.1$, $p<0.0001$) and of room presence ($F(1,10)=31.6$, $p<0.0005$). The interaction between orientation and room presence was not significant ($F(5,50)=1.1$, ns). A similar analysis on RTs revealed a significant main effect of orientation ($F(5,50)=29.1$, $p < 0.0001$) but no significant main effect of room presence ($F(1,10)=1.2$, ns) although the latter produced visibly distinct RT curves (see Figure 6) which might have reflected a small speed/ accuracy trade-off.

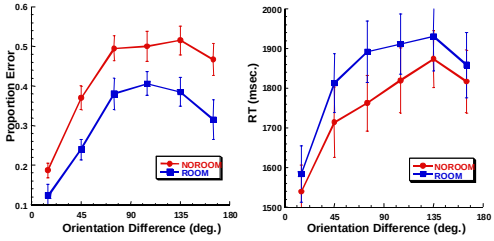


***Figure 6.*** *Results of Experiment 2. Proportion of errors (left) and mean RTs for correct responses (right) are plotted against the average angle of misalignment from the familiar, or reference, viewing direction.*

### *Discussion*

In summary, after natural interactive learning and repeated testing, the identification of 'paperclips' is still view dependent as reported by previous studies. In particular, the pattern of errors found here are very similar to those observed in psychophysical studies with both humans (e.g., Bülthoff & Edelman, 1992) and animals (Logothetis et al., 1994) using similar objects. Many of our observers reported that they found the differences between objects extremely difficult to classify at first. The most prominent means of differentiation was with respect to 'features' such as conjunctions of arms producing patterns which looked like familiar objects or which were comparatively distinctive. These allowed observers to distinguish one object from another within only a narrow range of views. However, these features appear to be maximally eliminated from view with 90° rotations of the observer and appear again (only in mirror image form) as rotations approached 180°. This could explain the inverted 'U' shaped performance functions obtained, which are qualitatively similar to the single-cell response profiles obtained by Logothetis & Pauls (1995). That is, cells were found to be sensitive to the orientation of the objects viewed and this sensitivity was at a minimum for 90° rotation views.

The most interesting finding however is that the room-present condition resulted in a significantly improved ability to identify novel views of objects. If objects are encoded and identified solely in terms of object features, then the presence of the context in which learning occurred should not influence performance. We therefore ask just how the presence of the room affects performance in recognition. One possibility is that the context conveys to an observer his/her viewpoint and thereby reduces the orientation uncertainty about the stimulus presented from a novel view. Signal detection theory suggests that such a reduction in signal uncertainty will lead to improved sensitivity (Green & Swets, 1974). Intuitively, the process of matching a target image to the stored representations will become earlier if the orientation of the target object is known. . However, it is also possible that the difference between conditions resulted because observers were simply disturbed by the removal of the background and that we were not looking at a benefit of the context but simply a deficit due to the removal of room. In order to test for a positive benefit of having a fixed reference frame (provided by the context) with respect to the object, we performed another experiment in which we directly manipulated the spatial relationship between object and room.

## Experiment 3 – Fixed/Rotating Room

In this experiment we wanted to preserve as much of the detail of Experiment 2 while manipulating the relationship between object and room. We did this by altering the software used previously to produce two conditions. The first was a repeat of the room-present condition of Experiment 2. The second condition was also based on the first but now we added a random

perturbation in both the azimuth (>30° & < 360°) and elevation (±15°) of the room (and pedestal) with respect to the objects. Unlike Experiment 2, the room remained present for both conditions. Another difference was that the learning and criterion tests reflected each of these two conditions. This allowed us to determine the effect of disrupting the spatial reference frame on learning as well as on the ability to generalize to novel views.

ble orders.

In addition, this experiment utilized an audible prompt after 2 seconds from stimulus onset for each identification trial. In the previous experiment we noticed a tendency for room-present responses to take longer than room absent responses (although this was not statistically significant) which could be interpreted as a speed/accuracy trade-off between the two conditions. We therefore tried to reduce this by

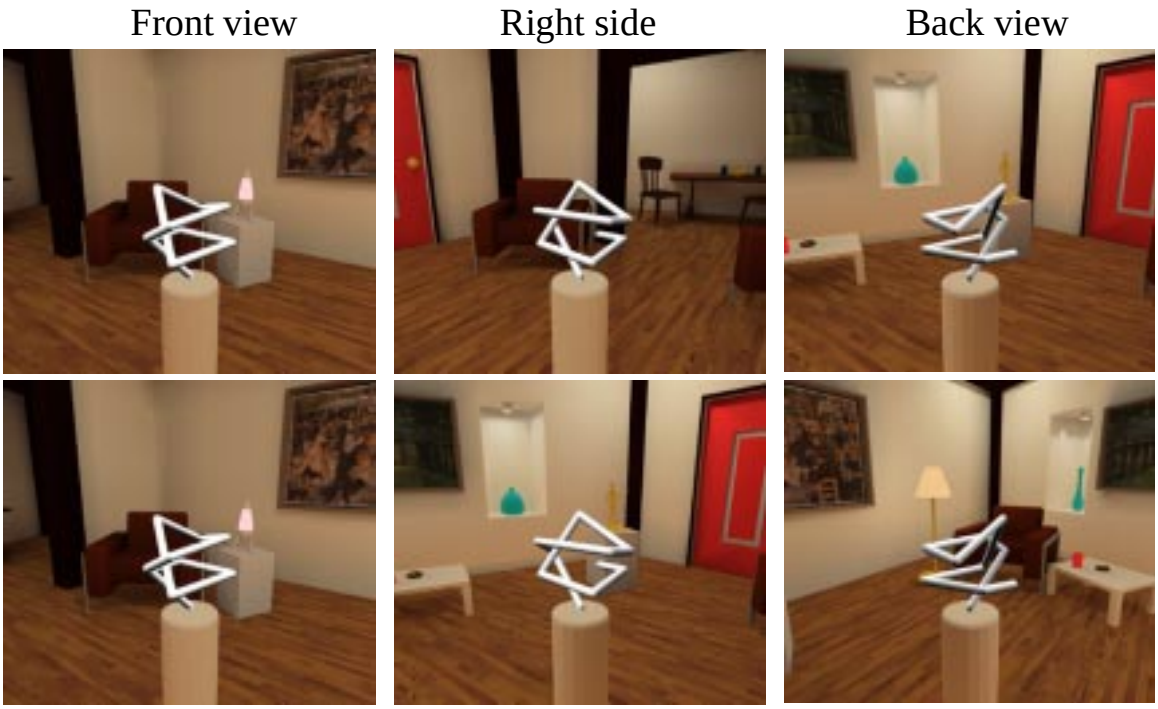


***Figure 7.*** *Images similar to stimuli used in Experiment 2. Top row shows front, side and back views (0, 90 and 180° rotations) of the same paperclip with appropriate changes in the visual background (fixed condition). The second row shows the same object but with arbitrary changes in the rooms orientation (rotating condition).*

### *Procedure*

The procedure was very similar to that of the previous experiment and consisted of four main stages (room familiarization, interactive learning, criterion test and main test). However, blocks from both 'room' conditions were performed separately and on two different occasions. This was to reduce the possibility that observers adopted a common strategy for both conditions and thereby ignoring the room (since it is uninformative in one of the two conditions). To reduce order effects, observers were alternately assigned initially to one of the two possi-

forcing observers to answer within a similar time frame for both conditions.

With the exception of the above-mentioned changes, the overall format of the experiment remained the same as the previous experiment.

The principle aim of the experiment was to determine if altering the learned spatial relationship between the room and objects (and therefore the observer) affected performance, compared with when this relationship was not altered. By randomly rotating the room between each trial relative to the object, we could remove environmental or contextual information about

the relative position of the observer with respect to the objects (see examples in Figure 7). That is, observers could not tell from the room whether they were viewing each object from the front, the back or the side. If this information is not used during identification, then no difference in performance between the conditions should be observed.

### Observers

The 12 participants were aged between 17 and 28 and were all naïve as to the purposes of the experiment and had not participated in these experiments before. Observers received payment for participation and were required to make at least two visits on separate occasions to perform both conditions of the experiment.

### Design

The experiment utilized a within-subject design with two factors: (1) orientation shifts in viewpoint (6 levels as defined in Experiment 2) and (2) room rotation (2 levels: room-fixed or room-rotating). The dependent variables were again the proportion of errors and RTs for correct responses.

### Results

#### Criterion Test

The average numbers of attempts an observer required to pass the criterion test for each of the two room conditions are shown in Table 2. Because the criterion tests corresponding to each condition reflected the nature of the main test, namely had room fixed with respect to the object or rotating randomly, we expected to see some differences between observers in terms of learning. From visual inspection of Table 2, the fixed room condition appears to have facilitated faster learning (at least initially). An analysis of variance with block (1,2,3) and room (fixed, rotating) as within-observers factors showed that the overall effect of room rotation on number of attempts was not significant, although the effect of block was significant ($F(2,22)=33.8$, $p<0.0001$) as was the interaction between these two ($F(2,22)=5.32$, $p<0.05$). A post hoc analysis revealed that the difference between fixed and rotating rooms was only significant for the first block which, as stated previously, was always the hardest to pass.

| *Room* | *Block 1* | *Block 2* | *Block 3* |
|---|---|---|---|
| *Present* | *2.8* | *1.3* | *1.4* |
| *Absent* | *3.7* | *1.4* | *1.2* |

***Table 2.*** *Shows the mean number of attempts required to pass the criterion test in Experiment 3.*

#### Identification Test

An analysis of variance on error rates and RTs with two within-observers factors (room rotation and observers' shift in orientation) was used to analyze the data. The error rates and RTs as a function of orientation shift are shown in Figure 8. The effect of orientation on error rates was significant ($F(5,55)=45.0$, $p<0.001$) as was the difference between the fixed/rotating room conditions ($F(1,11)=9.5$, $p<0.01$). The interaction between room rotation and observer orientation was not significant ($F(5,55)=1.85$, ns). With respect to RTs, the effect of orientation was significant ($F(5,55)=16.4$, $p<0.0001$), although there was no significant effect of room ($F(1,11)=0.6$, ns) and no interaction between these two ($F(5,55)=1.4$, ns). The RT profiles for the room rotation conditions in Figure 8 are not clearly separated, which reflects our use of the auditory response prompt in this experiment.

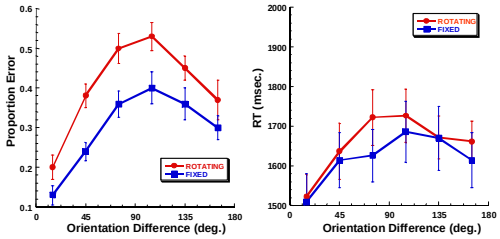


***Figure 8.*** *Results for Experiment 3 in which the room was either fixed or rotating with respect to the objects. The RT data are plotted with same ordinate as Experiment 2 for comparison. Overall, RTs were faster because of the time limit imposed for responses.*

### Discussion

The results provide a clear indication that the benefits afforded to observers by the room's presence is related to having a fixed relationship between object and room. Response errors for the room-fixed condition were significantly lower than for the room-rotating condition. If a fixed spatial relationship between the room and

the objects was not encoded by observers, then this difference should not have been observed. There was, however, no difference in terms of RTs showing that the benefit did not result from, for example, a longer viewing of fixed-room stimuli. Furthermore, we observed a significant reduction in the number of attempts required to pass the criterion test for the first time. Again, if the room's presence was ignored then having a fixed spatial reference frame between room and objects should have had no effect on accuracy.

## General Discussion and Conclusion

The experiments reported here produced two clear results. First, they show that even after repeated testing in a naturalistic learning task, generalization to novel views for geometric objects like 'paperclips' is still incomplete. Performance is a function of mis-orientation from a familiar viewing direction. Second, the spatial environment in which objects are learned can be used to facilitate identification. Addressing the issue of view dependency first, it should be recalled that the rear views of objects were often identified with apparently greater ease than oblique 90º views. This was reflected both by reduced errors and reduced response times for rear views of objects. Similar performance has previously been shown in monkeys (Logothetis & Pauls, 1995) and in humans (e.g., Vetter, Poggio & Bülthoff, 1994). Our impression is that even though such paper-clip objects do not consist of explicit parts, the means by which they are learned involves the identification of featural conjunctions, which differentiate one object from another. These features however, are view-dependent in the sense that they are specific to a given viewpoint and are maximally extinguished from view after oblique changes in viewpoint. When looking at an object from the rear these features may again become visible (albeit in mirror form) and this, in turn, may facilitate recognition. This interpretation is also consistent with the idea of 'virtual views' proposed by Vetter, Poggio & Bülthoff (1994), who described how a single view of a bilateral symmetric object can be used to produce three additional views without resorting to any specific knowledge of 3-D object structure at all.

Returning to the main objectives of the experiments, we have shown with two object identification experiments a significant positive effect of contextual background. Experiment 2 established that the room's presence improved performance, and Experiment 3 showed that the critical factor is the fixed relationship between objects and room. These two results together suggest that the room can be used to impose an objective reference frame, which allows an observer to tell from which direction they are looking at an object (e.g. from the side or the back.) The availability of such information was revealed by Experiment 1. In turn, this information regarding observer viewpoint can be used to reduce identification errors.

There are however important questions, which we have not addressed and which will be the subject of future research. First, it would have been desirable to compare the results from the no-room condition of Experiment 2 with the room-rotating condition of Experiment 3. It should be recalled that the only difference between these two was that the room was present in the latter case although it provided no fixed spatial reference. This would have allowed an additional verification of the benefits of a fixed reference frame. Alas, such an analysis was not feasible because there were differences between the two experiments, both in training and testing.

Another question relates to what aspects of the scene in particular afford the reference frame information. We took great effort to simulate a realistic setting, but would merely local information (from the specific shape of the pedestal for instance) suffice? We believe that we have taken the right course in starting with a rich scene first. A result of 'no facilitation' provided by arbitrary local information would have met with the argument that the setting was not ecologically valid. In order to avoid making arbitrary assumptions, one must start from the top and work down. Future research therefore would entail isolating the kinds of information, which allow such facilitation to take place.

Finally, if knowing one's orientation with respect to an object improves the ability to identify the object, how does this facilitation occur and how can this be reconciled with current thinking on recognition and identification? The benefit may originate, for instance, from an 'expectancy' of how an object and its features should look from a given direction (or perhaps how it shouldn't look). This is an explanation based on priming in the sense that one's view-

point focuses activation of specific components of memory. An alternative hypothesis is that knowing one's relative viewpoint 'explains' the appearance of the current visual stimulus. In this case facilitation may result because an inverse transformation (which can be determined from the given viewpoint information) is used to match the stimulus with contents of memory. An explanation for this observed facilitation would perhaps allow us to gain a better understanding of the mechanics of identification and recognition and will be the subject of future research.

In conclusion we have shown a positive benefit in the identification of objects derived from the presence of a familiar background. This can be attributed to viewpoint information provided by the environment. The provision of such information within a Virtual Environment Simulation however, did not eliminate the dependence of identification on the extent of mis-orientation between learning and testing views. These results must be taken into account in theories of object encoding and identification and also when considering the *interaction* of spatial knowledge derived from disparate sources, as is the case when identification is studied within a natural context.

## References


Biederman, I. (1972) Perceiving real-world scenes, Science, 177, 77-80.

Biederman, I. (1981) On the semantics of a glance at a scene, in M. Kubovy & J. R. Pomerantz (Eds.) Perceptual Organization, Hillsdale NJ: Earlbaum, 213-253.

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, 89, 60-64.

Christou, C. G. & Bülthoff, H. H. (In press) View dependence in scene recognition after active learning, Memory & Cognition.

Feldman, J. A. (1985) Four frames suffice: A provisional model of vision and space, The Behavioral and Brain Sciences, 8, 2, 203-289.

Green, D. M. & Swets, J. A. (1974) Signal Detection Theory and Psychophysics. Huntington, New York: Robert E. Krieger Publishing Company.

Hollingworth, A. & Henderson J. M. (1998) Does consistent scene context facilitate object perception? Journal of Experimental Psychology: General, 3, 1-28.

Humphrey, G. K. & Jolicoeur, P. (1993) An examination of the effects of axis foreshortening, monocular depth cues, and visual field on object identification, The Quarterly Journal of Experimental Psychology, 46A, 1, 137-159.

Logothetis, N. K., Pauls, J., Bülthoff, H. H. & Poggio, T. (1994) View-dependent object recognition by monkeys, Current Biology, 4, 401-414.

Logothetis, N. K. & Pauls, J. (1995) Psychophysical and physiological evidence for viewer-centered object representations in the primate. Cerebral. Cortex, 3, 270-288.

Maguire, E.A, Burgess, N, Donnett, J.G., Frackowiak, R.S, Frith, C.D., & O'Keefe, J. (1998) Knowing where and getting there: A human navigation network, Science, 280, 921-924.

Marr, D. (1982) Vision: A computational investigation into the human representation and processing of visual information, San Francisco: W.H. Freeman and Co.

O'Keefe, J & Dostrovsky, J (1971) The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely moving rat, Brain Research, 34, 171-175.

O'Keefe, J & Conway D. H. (1978) Hippocampal place units in the freely moving rat: why they fire where they fire, Experimental Brain Research, 31, 573-590.

Palmer, S. E. (1975) The effects of contextual scenes on the identification of objects, Memory & Cognition, 3, 5, 519-526.

Péruch, P, Vercher, J.L., & Gauthier, G.M. (1995) Acquisition of Spatial knowledge through visual exploration of simulated environments, Ecological Psychology, 7,1,1-20.

Rock, I. & DiVita, J. (1987) A case of viewer-centered object perception, *Cognitive Psychology*, 19, 280-293.

Shelton, A. L. & McNamara, T. P. (1997) Multiple views of spatial memory, *Psychonomics Bulletin & Review*, 4, 1, 102-106.

Tarr, M. J. (1995) Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review*, 2(1), 55-82.

Taube, J. S., Muller, R.U., Ranck, J.B. Jr (1990) Head-direction cells recorded from the postsubiculum in freely moving, Journal of Neuroscience, 10, 436-447.

Vetter, T., Poggio, T. & Bülthoff, H.H. (1994) The importance of symmetry and virtual views in three-dimensional object recognition, Current Biology, 4, 1, 18-23.