Extrinsic cues aid shape recognition from novel viewpoints

Chris G. Christou Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Bosco S. Tjan University of Southern California, Los Angeles, CA, USA

Heinrich H. Bülthoff Max Planck Institute for Biological Cybernetics, Tübingen, Germany

It has been shown previously that the visual recognition of shape is susceptible to the mismatch between the retinal input and its representation in long-term memory, especially when this mismatch arises from rotations in depth. One possibility is that the visual recognition system deals with such mismatch by a transformation of the input or the representation thereby bringing both into alignment for comparison. In either case, knowing what transformation has taken place should facilitate recognition. In natural circumstances, objects do not disappear and appear in different orientations inexplicably and an observer usually knows what to expect according to the context. This context includes the environment, and the history of the observers' movements, which specify the transient relationship between the object, the environment and the observer. We used interactive computer graphics to study the effects of providing observers with either implicit or explicit indications of their view transformations in the recognition of a class of shape found previously to be highly view-dependent. Results show that these cues aid recognition to varying degrees but mostly for oblique views and primarily in terms of accuracy not response times. These results provide evidence for egocentric encoding of shape and suggest that knowing ones' transformation in view helps to reduce the problem space involved in matching a shape percept with a mental representation.

Keywords: shape recognition, view-dependency, extrinsic cues, context, virtual environments, computer graphics

Introduction

Visual recognition of 3D objects involves finding a positive match between a two-dimensional retinal projection and a stored mental representation. This match may be performed as a comparison of the defining properties of the percept and items in memory. These properties may include non-accidental (non-transient) features that persist over time including the objects' shape, color, surface texture and material composition and perhaps even its position in space. It is reasonable to assume that humans use as many, or as few, properties as necessary in object identification. However, regardless of which property is used the recognition process will always involve uncertainty due to the variations that can occur in environmental variables. These include changes in illumination, position, decomposition of materials, changes in viewpoint, atmospheric changes (e.g. fog), etc. The goodness of a match is a function of the number of properties that may be relied upon and the weights assigned to each property. We may assume that these weights are derived during the object learning process and are dependent on how uniquely a given property specifies the identity of the object and how readily the property changes under environmental or viewing transformation.

In this paper we are concerned with the influences of viewpoint changes on object identification using just its shape. Physical object shape is a non-transitory property that usually persists over time. However, whereas the physical shape of an object may remain invariant, the perception of shape may change according to the factors listed above. For example, dramatic changes occur in the projection of an objects' shape as an observer moves around it. Our ability to recognize an object in this instance is dependent on the mental representation we have of it. Clearly, if we were unfamiliar with a particular facet or view of an object then we would have difficulty in recognizing it from this view. The nature of how a shape is represented in memory is a subject of on-going debate (Biederman, 1987; Biederman & Gerhardstein, 1993; Bülthoff, Edelman & Tarr, 1995). A rough distinction between two prominent theories is that on the one hand shape is represented mentally in terms of its 3D components (Marr, 1982; Marr & Nishihara, 1978; Biederman, 1987; Biederman & Gerhardstein, 1993) and on the other as a collection of views (Bülthoff, Edelman & Tarr, 1995). Regardless of the validity of any theory, there is overwhelming empirical evidence showing that recognition performance often decreases when an observer's viewing position is different between when an object is learned, and when it is to be recognized. This

DOI 10:1167/3.3.1 Received October 23, 2001; published April 8, 2003 ISSN 1534-7362

has been shown in several psychophysical experiments Rock, DiVita & Barbeito, 1981; Bülthoff & Edelman, 1992) and more recently by neurophysiology (Logothetis & Pauls, 1995). Thus, in psychophysical experiments subjects might initially view the shapes oriented in one position and later have to recognize them after they had been rotated in depth in the azimuthal direction by an amount referred to as the displacement angle. The error rate and/or the response times in such studies were found to be a function of the displacement. This view-dependency has been used to argue that shapes are represented within an ego-centric frame of reference and objects are represented in memory as we view them with little or no interpretation of their shape in three dimensions.

Although view dependent recognition has been reported for familiar objects Bartram, 1974; Cave & Kosslyn, 1993; Srinivas, 1993) many studies have used novel geometrical figures (e.g. Rock, DiVita & Barbeito, 1981; Bülthoff & Edelman, 1992). This was in order to isolate shape, reduce the effects of prior experience and eliminate unwanted surface cues (such as color). In general, where recognition time is shown to be dependent on viewpoint change it has been suggested that generalization to novel views is the result of normalization procedures that transform the retinal stimulus in order to align it with the ego-centric memory representation (e.g., Jolicoeur, 1985; Tarr & Pinker 1989; Ullman, 1989). Such normalization procedures were also suggested by Shepard (e.g., Shepard & Metzler, 1971; Shepard & Cooper, 1982) in which the increased response latencies for rotated views of objects was claimed to be a result of an analogue mental rotation of the visual stimulus to bring it into alignment with the contents of memory. Since the procedure is purported to be an analogue process, the time taken to perform the transformation is a function of the displacement angle (see Jolicoeur and Humphrey, 1998).

In this scenario, one important question is how the visual system determines in which direction and by how much the stimulus must be rotated to bring it into alignment (for review see Palmer, 1989). The importance of this was suggested in experiments by Shepard & Cooper (1982). In 2D mental rotation exercises they presented subjects with 2D shapes and then gave them a prior indication of the direction in which the shape might be rotated. Subjects were asked to first imagine what the object would look like from the cued orientation prior to the presentation of the test stimulus. The time taken to perform this imagined rotation was suggestive of an analogue process, as in previous experiments. However, the time taken to respond to the self-initiated stimulus was not view-dependent. This indicates that the time-consuming variable in responses is the time taken to work out the appropriate transformation. If the difference between the learned view and the current view is known this could allow for preparatory processing that facilitates view-independent recognition. A parallel analysis that reaches essentially the same qualitative conclusion on the utility of such extrinsic view information (information that is not part of the target object) can begin without mental rotation or any other sequential matching operations. Eckstein, Thomas, Palmer & Shimozaki (2000), for example, showed that signal uncertainty alone can explain size set effect on visual search, without having to postulate serial search with limited capacity. For object recognition, not knowing the viewpoint difference between the learned and current views means that an optimal detector, even with unlimited capacity, must simultaneously monitor a large number of possible views for each object. To keep false-alarm rates in check, the threshold of each of these detectors must be increased relative to what would be required if there were only one possible view per object (hence no viewpoint uncertainty). A higher detection threshold, however, will take more time to reach or result in lower accuracy. Any extrinsic view information will help to reduce signal uncertainty and shorten the time needed for accurate recognition.

Where could extrinsic view information be derived from? Visual cues from the surroundings and vestibular and-or proprioceptive cues are important sources. The importance of non-visual cues in the perception and recognition of spatial layout has been documented by Simons & Wang (1998) who found that subjects who perform their own movements around a collection of objects are less prone to make mistakes in identification of the spatial layout of these objects than when the objects are rotated by the same amount and the subject stands still. Simons & Wang attributed this to an ability to spatially update ones' mental representation according to the knowledge of their own movement. This also amounts to a form of spatial cueing. More recent studies have revealed similar influences of extra-retinal cues in the recognition of novel objects (Simons, Wang & Roddenberry, 2002).

It may be conjectured from the foregoing that much of the difficulty found in recognizing novel objects in previous studies relates to their presentation in isolation, without contextual or background information, and without knowledge of the extent or direction of the transformations in view. To perform a systematic, yet controlled, study of extrinsic view information in shape recognition, we used interactive computer graphics to convey both a realistic context in which shapes are learned and provide rigid control over stimulus generation and presentation. We attempted to make the learning process as natural as possible. Subjects were allowed to manipulate their orientation with respect to the object by manipulation of a 3D mouse. Using real-time computer graphics, the observer was provided with visual feedback of their movement as they rotated about the shapes. They were therefore free to implement natural learning strategies to discriminate between self-similar shapes. By using realistic lighting, texture and shading

(see Appendix A) we further optimized visual cues. Below we describe four experiments. The first experiment assessed whether the environmental background influences recognition performance. The second addressed more specifically the benefits of the environment as a fixed frame of reference for specifying the changes in viewpoint. The third experiment tested the utility of an abstract but explicit indication of the observer's original viewpoint presented *simultaneously* with the test objects. The final experiment assessed the effect of an explicit viewpoint indicator available only *prior* to the presentation of the object.

Experiment 1: Implicit Cues – Paperclip in a Room

The first experiment tested if the presence of a visual background during learning influences later recognition. Subjects were allowed to make small rotational movements around the shapes that were presented within a rich visual context (see Figure 1). Their rotation around the shapes was implied by the transformations occurring in the background and these movements in turn helped them to fully appreciate the 3D nature of each shape. If the geometry of the shapes is encoded or represented without recourse to the surrounding context then we would expect to find no differences either in speed of learning (reaching a criterion level of performance) or in the ability to recognize completely novel views.

Method

The task consisted of learning to differentiate between four 'paper-clip' shapes followed by identification tests in which the subjects had to identify each shape viewed from several unfamiliar orientations. The experiment involved four stages: (1) room familiarization by simulated locomotion, (2) interactive learning of the individual shapes, (3) criterion test to assess performance from familiar viewpoints, and (4) novel view identification. The latter three stages (shown in **Figure 2**)

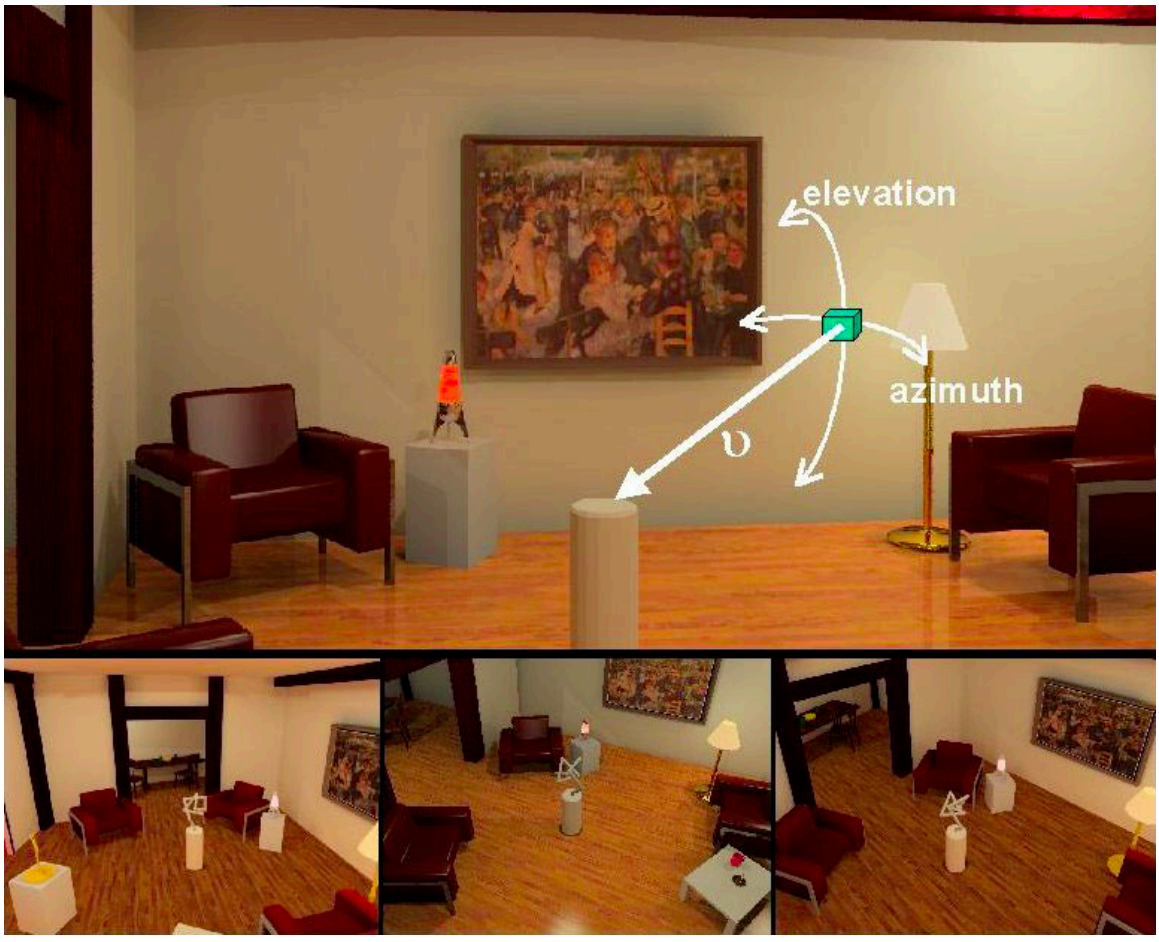

Figure 1. (a) Rendered image of the virtual environment used in the experiments showing the pedestal around which simulated movements were performed. The viewpoint of the observer is specified by υ , a direction vector whose origin varied with viewpoint and which was always directed to the same point just above the pedestal on which the shapes appeared. (b) Three rendered images showing an example wire-like shape resting on the pedestal.

were repeated three times in succession for each block. The initial free locomotion stage familiarized subjects with the spatial layout of the environment from many perspectives. In order to encourage subjects to explore the room they were instructed to locate and acknowledge randomly positioned two-digit codes. These codes only appeared when viewed within a short distance and this game-like procedure was useful in teaching the subjects the layout of the room.

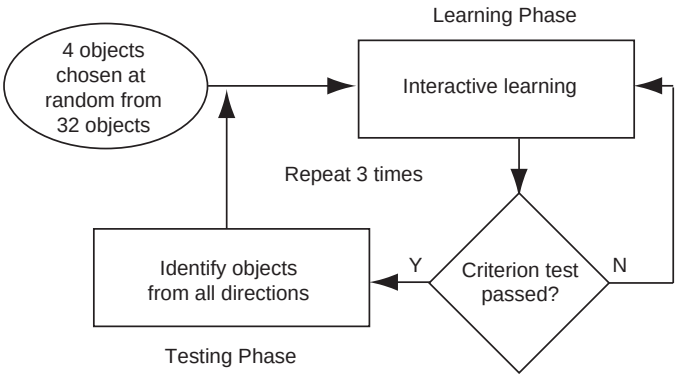

Figure 2. Flowchart of the experimental procedure for each block. Four new shapes were chosen for each block. Interactive training was repeated until criterion performance was reached. The entire procedure was repeated 3 times for each set of four shapes (for each block).

After three minutes of room familiarization, the shape learning stage commenced automatically. Each block involved the familiarization of four new shapes, which were loaded individually and appeared to be resting on the top of the pedestal in the middle of the room (see Movie 1). With four fingers of their preferred hand placed on four different buttons of a computer keyboard, the subjects could alternate between each of the four shapes. Thus, they learned to associate a shape with each finger as this was hoped to remove any additional latency that would be introduced had subjects been required to learn arbitrary shape names. With their other hand they could manipulate a six-degrees of freedom SpaceMouse which allowed them to change both the azimuth and elevation of their viewpoint of these objects by up to 15° on either side of a randomly chosen reference viewpoint. The nature of this movement was analogous to being tethered to an invisible point just above the pedestal. This learning viewpoint remained constant for each block of the experiment.

The learning stage lasted two minutes after which a criterion test was performed. In this test, 4 randomly chosen static views of each shape had to be identified by pressing the pre-designated keys on the computer keyboard. The views were always within the bounds of subjects' movement during learning (e.g., ± 15° for azimuth and elevation). Each trial consisted of the following steps:

1. The vacant pedestal and complete view of the room was displayed for an indefinite period.
2. Subjects initiated the presentation of the test shape.
3. The test shape was displayed for 500msec. and disappeared, leaving only the view of the room.

Each object was presented 4 times during the criterion test. Subjects passed the criterion test if 14 out of 16 responses were correct. Otherwise, they repeated the learning stage with the same 4 objects and viewing direction. The number of attempts required to pass the criterion tests was recorded.

Movie 1. Demonstrates interactive learning of the test shapes. Subjects rotated their view of the shapes by manipulating the SpaceMouse. Each of the four objects could be viewed individually by pressing one of four keys on the computer keyboard. This learning stage lasted 2 minutes.

Once the criterion test was passed, the ability to generalize to novel viewing directions was tested. This constituted the main test of the experiment and was in all respects similar to the criterion test except that it utilized not only familiar views but also novel views of each shape. Views were generated from 12 viewing positions evenly distributed around the objects. The influence of the environment was tested with two conditions. One in which the wire object was presented on a gray background and one in which the shapes were presented with the visual background visible and consistent with the transformation in viewpoint (Figure 3). Each shape was presented 12 times: once from each of 12 viewpoints. Each viewpoint differed from the familiar azimuth direction by a multiple of 30° around the pedestal and was perturbed by a random offset of between ± 15° in azimuth and elevation (see Figure 4). This reduced the

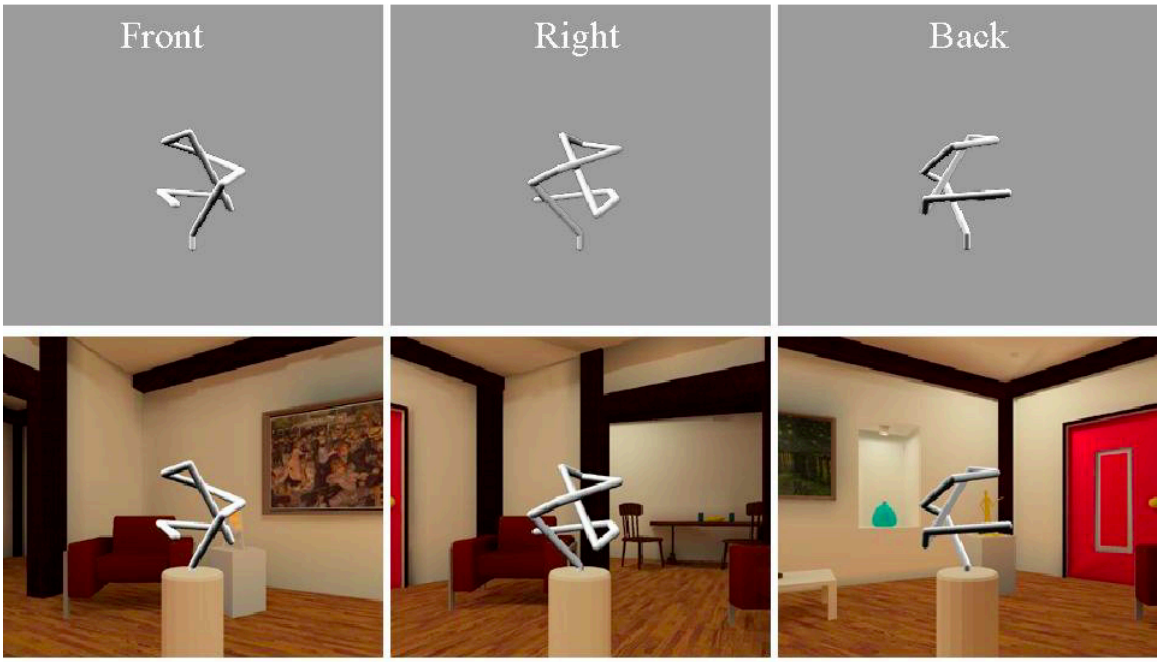

Figure 3. Example images of stimuli used for each of the two conditions in Experiment 1. From left to right the images show the same shape after the observer rotates by 0, 90 and 180° anti-clockwise. These correspond to the front, right and back views of the object. It is easier to see this by considering the images in the second row in which the depicted environment helps establish a frame of reference that specifies what the implied change in view is. However, the ability to do this depends on the fact that the environment is a familiar one. For this reason subjects spent the first portion of each block familiarizing themselves with the layout of the room.

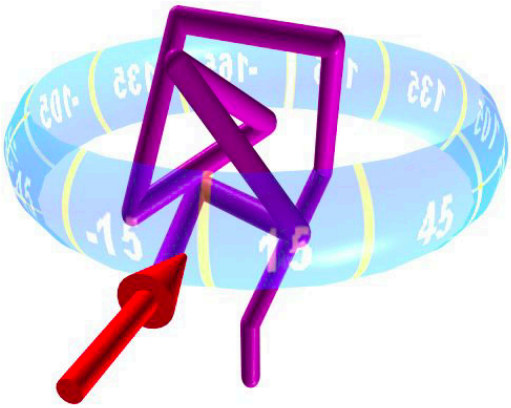

Figure 4. Shows an example shape with a depiction of the 12 view segments used. The actual viewing direction was chosen to lie within the bounds of each of these segments. This was determined by calculating a fixed offset multiple of 15° from the familiar direction (red arrow) and then adding a random perturbation of ± 15° in both azimuth and elevation.

possibility of relying on accidental features of any particular view of the room and/or the test shapes.

Criterion test and main test were repeated three times to collect sufficient repetitions for data analysis and to assess learning effects. Learning across blocks 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 4 (objects) x 3 (repeats) x12 (orientations) = 144 trials. There were four blocks of four new objects for each condition. Thus, each condition consisted of 4x144=576 trials in which the proportion correct responses and response times were recorded for later analysis.

A two-factor repeated-measures design was used. The two factors were (1) room presence during test with two levels: present/absent and (2) angular displacement of viewpoint (i.e. the difference in azimuth between familiar and test 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 responses 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. Each block used four new objects chosen at random from a previously generated database of 32 shapes.

The 11 observers were between 17 and 31 years of age and paid for each hour of participation. All were given prior instruction in all conditions of the experiment and

in the use of the SpaceMouse. All observers were naïve as to the purposes of the 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 as possible.

Results

Criterion Tests

The criterion test always involved the presentation of the test shape within the room context. The number of successive attempts at the criterion test varied as a function of the block number. Table 1 shows that the first criterion test in each block was always the hardest to pass. Because training always involved the presence of the room 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 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).

Table 1. Average Number of Attempts Before Passing the Criterion Test.

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

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 was always the same (i.e. room was always present during criterion test).

Identification (Main) Tests

Responses not made within 4 seconds of each presentation during main tests were discarded from the analysis (this happened in approximately 5% of all trials). The data for each subject was averaged within each of the six bins of mean-azimuthal displacement angles and over all elevation changes. The response times (RT) for correct responses were averaged in an identical manner. Figure 5 shows the proportion of errors and RT averages as a function of displacement angle. 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 shows up also in the reaction times. The pattern of errors for the two conditions is also clearly different, with the room-present condition producing fewer errors for nearly all orientations than the room-absent condition. A 2x6 repeated measures analysis of variance with room (present, absent) and angular displacement (6 levels; one for each displacement bin) as within-subject factors showed a significant effect of displacement angle ($F_{5,50}$=44.1, p<0.0001) and of room presence ($F_{1,10}$=31.6, p<0.0005). The interaction between rotation and room presence was not significant ($F_{5,50}$=1.1, ns). Individual contrasts (student t-tests with paired samples) between corresponding error rates for room-present and room-absent conditions showed that all differences were significant at $p < 0.05$ except the familiar view (within ± 15°), which was approaching significance (t_{10}=2.1, p=0.06).

A similar ANOVA was performed on the response times, which showed a significant main effect of displacement angle ($F_{5,50}$=29.1, $p < 0.0001$) but no significant effect of the rooms' presence ($F_{1,10}$=1.2, ns).

Discussion

The identification of these shapes was found to be dependent on the displacement angle for both conditions. In particular, the pattern of errors found here are similar to those observed in studies involving both humans (e.g., Edelman & Bülthoff, 1992) and primates (Logothetis, Pauls, Bülthoff & Poggio, 1994) using similar shapes. Many of our observers reported that the objects were very similar and difficult to distinguish at first. The most prominent means of differentiation was with respect to 'features' such as conjunctions of arms producing patterns that could be used to differentiate one object from another. However, these features must have been specific to a limited set of views as they did not facilitate view invariant recognition for displacement angles greater than 30° . Peak error rates occurred at approximately 90° rotation of the observers' viewpoint, which suggests that these 'features' are most difficult to detect at 90° rotations and become easier at or approaching 180° rotations. The presence of the visual environment did not eliminate view dependence although it did reduce errors significantly. There are a number of possibilities for this:

1. The rooms' disappearance during main tests perturbed subjects causing an increase in error rates.
2. Features of each shape were encoded with respect to features in the room. Recognition was facilitated by identify room/shape correlations.
3. The rooms' presence during testing provides contextual depth cues, which helps to recover the depth dimension of the object (Humphrey & Jolicoeur, 1993).
4. The room was used as a frame of reference during training and testing and was used to judge displacement angle.

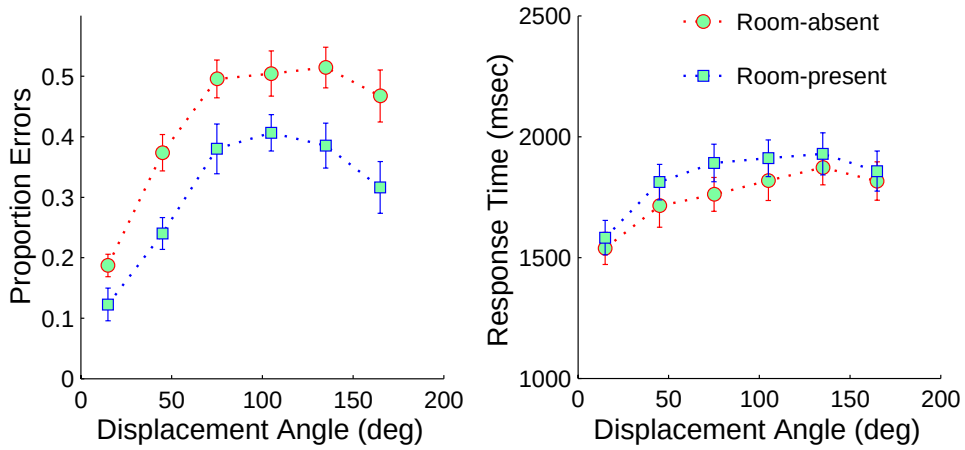

Figure 5. Results of Experiment 1. Proportion of errors (left) and mean RTs for correct responses (right) are plotted against the average angle of rotation from the familiar, or reference, viewing direction.

Possibility 2 seems unlikely given the pattern of errors. An encoding based on visible room features during training would be expected to help to varying degrees for displacement angles between 0 and 90° . Maximum facilitation would be expected at 0° and this would reduce to minimal facilitation for 90° displacements. This is because at 90° the room features visible from the familiar viewing direction would disappear from view (given the limited viewing distance and field of view). For displacement angles greater than 90° the facilitation would be zero. In contrast, we found no facilitation due to the presence of the room at 0° , but a significant amount of facilitation at 90° or greater.

Possibilities 1 and 4 cannot be ruled out by the current experiment. Experiment 2 was designed to address these issues. In Experiments 3 and 4 we reduced subsidiary depth cues completely, addressing possibility number 3 and testing the effects of explicit view change cues.

Experiment 2: Implicit Cues – When the Room Rotates

If the room facilitated recognition in Experiment 1 by serving as a frame of reference, then the orientation of the room relative to each shape must remain constant between training and testing. Otherwise, observers cannot use it to judge their displacement. In the second experiment, we manipulated this relationship between shape and room by making random perturbations in the orientation of the room relative to the shapes. Two conditions were used: the 'fixed-room' condition was the same as the 'room-present' condition in Experiment 1. In the 'rotating-room' condition, we introduced random perturbations of the orientation of the room with respect to the shapes. Since the room was present in both cases this served as a test for possibility number 1. If subjects were merely perturbed by the disappearance of the room then no difference would be observed between these two conditions.

Method

The procedure was identical to Experiment 1. The rotating-room condition consisted of rotated views of the shapes seen during training but with an additional random perturbation in both the azimuth (>30° & < 360°) and elevation (± 15°) of the room with respect to the shapes (see Figure 6). In addition, unlike Experiment 1, the current experiment also utilized training regimes that reflected the test condition (i.e. during blocks of the rotating-room condition, the room orientation was also perturbed relative to the shape during criterion trials).

The 12 participants were aged between 17 and 28 years and had not participated in the experiment before.

Results

Criterion Test

The average number of attempts an observer required to pass the criterion test for each of the two conditions is shown in Table 2. Because the criterion tests corresponding to each condition reflected the nature of the main identification test (namely that the visual environment was either fixed or rotated with respect to the test shapes) we expected to see differences in learning time if observers were influenced by the visual environment. By inspection of Table 2, the fixed room condition appears to have facilitated faster learning (at least initially). However, a repeated measures ANOVA 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 ($F_{1,11}$=1.29, p=.28), although the effect of block was significant ($F_{2,22}$=50.2, p<0.0001). The interaction between these two did not reach significance ($F_{2,22}$=2.59, p=0.09). A post hoc analysis (Newman-Keuls test) revealed that the difference between fixed and rotating rooms was

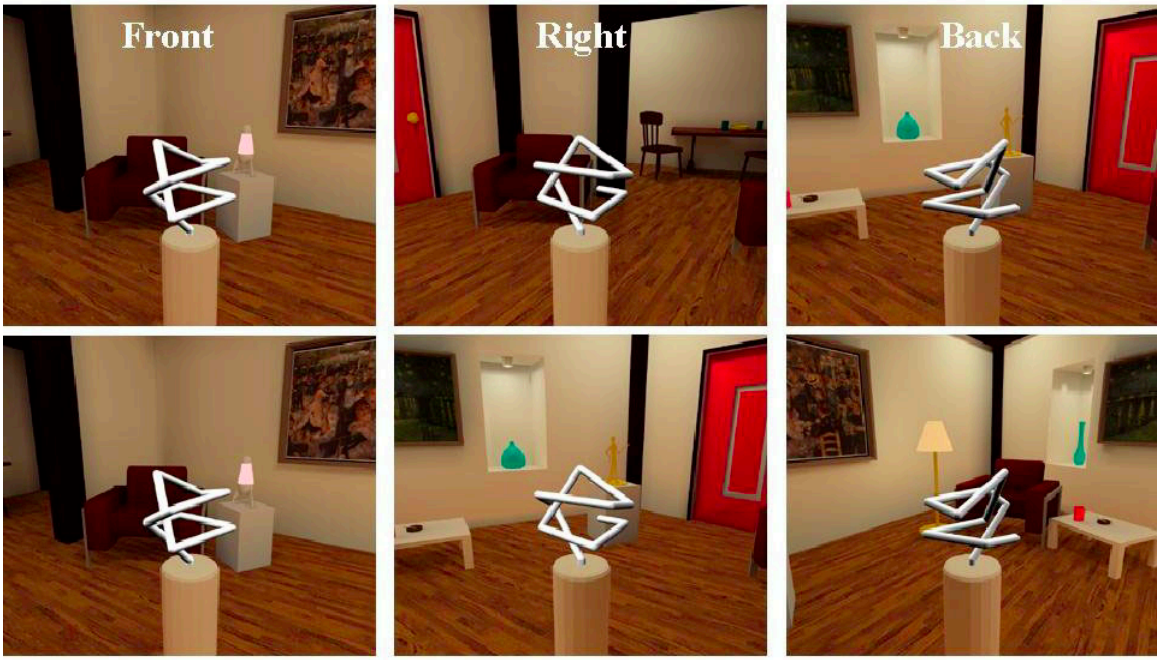

Figure 6. Images similar to stimuli used in Experiment 2. Top row shows front, side and back views (0° , 90° and 180° rotations) of the same object with consistent changes in the visual background (fixed-room condition). The second row shows the same object but with random rotations in the rooms orientation with respect to the shapes (rotating-room condition).

significant only for the first block, which was always the hardest to pass.

Table 2. Mean Number of Attempts Required to Pass Criterion Test in Experiment 2.

Room	Block 1	Block 2	Block 3
Fixed	2.4	1.3	1.3
Rotating	3.0	1.4	1.1

Identification Test

An analysis of variance of error rates and reaction times with two within-observers factors (room rotation and displacement angle) was used to analyze the data. The error rates and reaction times as a function of orientation shift are shown in Figure 7. The effect of displacement angle on error rates was significant ($F_{5,55}$=45.0, $p<0.001$) showing once again that identification from novel views was view-dependent. The main effect of room (fixed/rotating) was also significant ($F_{1,11}$=9.5, $p<0.01$). The interaction between room condition and viewpoint was not significant ($F_{5,55}$=1.85, ns). Individual paired student t-tests comparing the means of each room condition for each displacement angle showed that only the 45° , 75° & 105° displacements produced significant differences between room fixed and room rotating conditions. For 0-30° displacements, the difference in means was only approaching significance (t_{11}=1.98, $p=0.07$). With respect to response times, the effect of displacement angle 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).

Discussion

The results from both the average numbers of blocks required to reach criterion and the proportion of errors during the main test reveal an advantage of having a fixed spatial relationship between the shapes and their surrounding environment. Response errors for the fixed-room condition were significantly lower than for the rotating-room condition although the amount of facilitation was found to be optimal for displacements of around 90° . At its peak the facilitation amounts to a reduction of around 15% in error rate. This suggests that the benefit of the visual surround is that it provides a stable frame of reference with which to gauge the extent of the view transformation.

Results from Experiment 1 and 2 suggest that the room was used as a frame of reference to determine the direction and extent of the displacement angle. The extent and direction of the change in view was implicit in the changes observed in the background detail. What if the displacement angle is given more explicitly? Will there still be a facilitation effect? Our next two experiments investigated further.

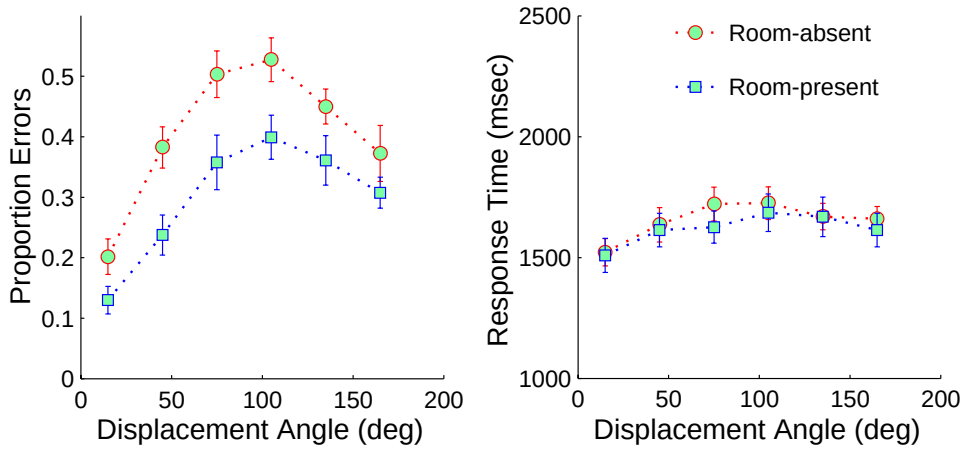

Figure 7. Results for Experiment 2 in which the room was either fixed or rotating with respect to the objects. The RT data are plotted with same ordinate as Experiment 1 for comparison. Overall, RTs were faster because of the time limit imposed for responses.

Experiment 3: Explicit Indication of Original Viewpoint

The results of the previous two experiments suggest that a stable environmental reference frame can improve the ability to recognize novel 3D shapes from unfamiliar directions. One reason for this is that the background can be used as an implicit indicator of the kind of transformation that has occurred. This in turn may be used in a process of mental rotation (as suggested by Shepard & Cooper (1982)). Although it is not clear how this occurs one would predict that in this case the same level of facilitation can be achieved by using a simple indicator that *explicitly* conveys the information that the room conveyed *implicitly*.

Method

An experimental procedure similar to Experiments 1 and 2 is used. This experiment again consisted of two conditions. In the 'indicator-absent' condition subjects learned to differentiate between the shapes portrayed on the computer screen on a gray background (as in Experiment 1) and identification ability was also tested on a gray background. In the 'indicator-present' condition the geometric shapes were portrayed as resting on a solid block with an arrow indicating the training viewing direction (see Figure 8). This indicator was also visible during the main identification tests, and observers were instructed as to its purpose. All other aspects of this experiment were the same as in Experiment 1 although the presentation of the geometric shapes was not initiated by the observers themselves but was presented after a fixed period of 3 seconds. Thus, in the indicator-absent condition observers viewed a gray screen for 3 seconds prior to the presentation of the test shape and in the indicator-present condition they viewed the rotated indicator oriented appropriately relative to the original training view prior to the appearance of the test shape. In the indicator-present condition observers therefore had a chance to appreciate their original viewing direction with respect to the new viewing direction.

Subjects

All 9 subject were initially naïve as to the purpose of the experiment and none had participated in the previous experiments. They were given initial instruction on how to perform the experiment. Because these experiments did not utilize the visual environment used previously, no room familiarization was required.

Results

Criterion Test

The average number of attempts an observer required to pass the criterion test for each of the two conditions is shown in Table 3. By inspection of Table 3, the indicator appears to have facilitated faster learning, at least for the first block in each experiment. A repeated measures ANOVA with block (1,2,3) and indicator (present, absent) as within-observers factors showed that the effect of block number was significant ($F_{2,16}$=16.5, p<0.0005). The effect of indicator showed a trend towards significance ($F_{1,8}$=4.8, p=0.059). The interaction between these two was not significant ($F_{2,16}$=2.3, p=.13). A post hoc analysis (Newman-Keuls) showed that the difference between fixed and rotating rooms was again only significant for the first block (p<0.01).

Table 3. Mean Number of Attempts Required to Pass Criterion Test in Experiment 3.

Room	Block 1	Block 2	Block 3
Fixed	1.7	1.3	1.2
Rotating	2.3	1.3	1.3

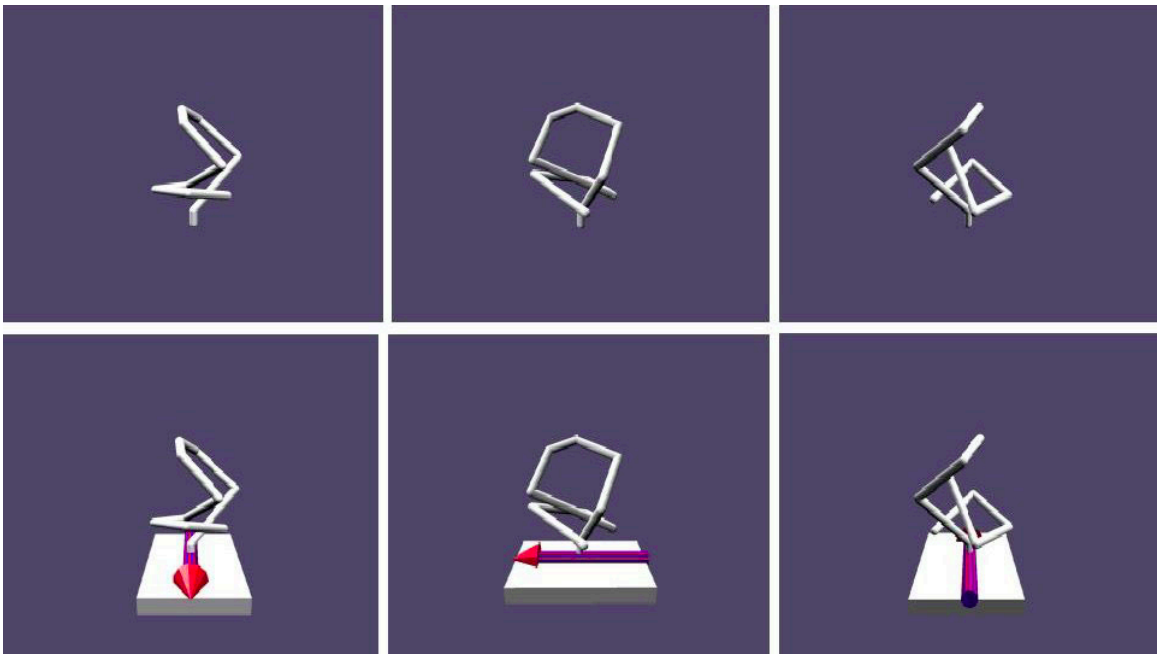

Figure 8. Shows example stimuli from the two conditions in Experiment 3. The indicator (bottom row) pointed to the original viewing direction (left column) and could be used to give an advance indication that the shapes would be observed from the side (middle column) or from the back (right column).

Identification Test

Results for the main identification test in this experiment resemble those of the other two experiments (see Figure 9). An analysis of variance (ANOVA) was performed on the proportion of errors as in previous experiments. This revealed a significant effect of the indicator ($F_{1,8}$=22.8, p=0.005), a significant effect of displacement angle ($F_{5,40}$=33.9, p=0.0001), but no significant interaction between these two ($F_{5,40}$=1.67, p=0.16). A similar analysis on the response times associated with correct responses revealed no significant effect of the indicator ($F_{1,8}$=.17, ns), but a highly significant effect of the displacement angle ($F_{5,40}$=21.36, p=0.0001). There was no significant interaction between these two ($F_{5,40}$=0.75, ns).

Discussion

We conjectured from the results of Experiments 1 & 2 that the presence of the room allowed subjects to gauge the extent of the transformation in viewpoint around the shapes. The purpose of the indicator in this experiment was to serve as an abstract indication of what the room may have provided; namely, an indication of shift in viewpoint. Results show that the indicator yields similar

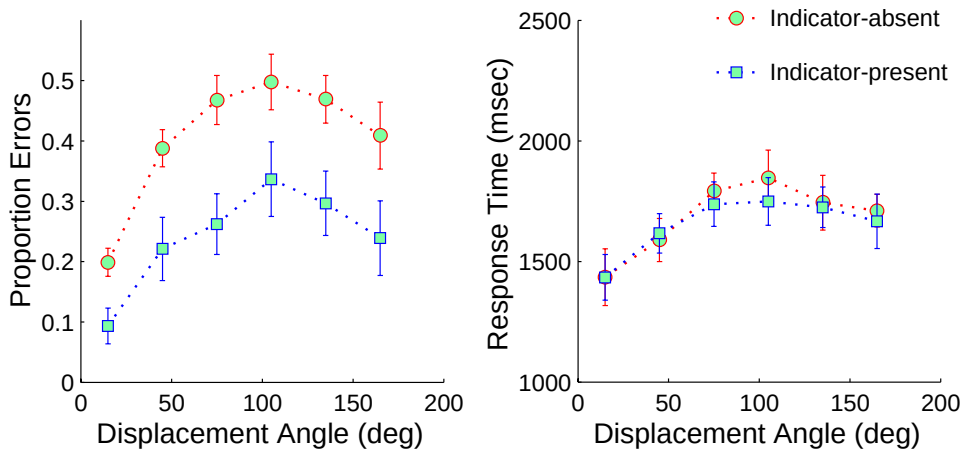

Figure 9. Results for Experiment 3 testing the effect of a geometrical indication of observer's original viewpoint.

facilitation, as did the room. One possibility however, and this was alluded to in the discussion of the first experiment, is that the main benefit derives from encoding each shapes' features with respect to the indicator. Since the indicator was present during learn and test phases (for the indicator present conditions) this is a possibility. In our last experiment, in which the utility of indicating the new (as opposed to original) viewing position is assessed, we removed all additional detail between learn and test phases.

Experiment 4: Explicit Indication of New Viewpoint

The purpose of this experiment was to determine if facilitation in shape identification is apparent when subjects are given prior indication of their new viewpoint. This experiment utilized the most abstract conditions. That is, no additional detail persisted from training to identification test apart from the shapes themselves. The subjects' new viewpoint with respect to the objects was conveyed to them by means of an arrow pointing to different positions on a globe. We reasoned that by indexing the viewing perspective in this manner (no room context, and indicator shown prior to shape presentation) that any shape encoding and identification based on conjunction of features would be eliminated. Furthermore, all subsidiary depth cues not arising from the geometry of the shapes are also eliminated. In this case, any facilitation that would be observed is a result of the extrinsic viewpoint information alone.

Method

The procedure was the same as in the previous experiments. The viewpoint indicator consisted of a globe with a ring (see Figure 10) plus a pointer that indicated from which position they would view the test shape. A within-subjects design was used with two conditions. In the indicator-present condition a red shaded pointer would point somewhere on the globe to indicate the new viewpoint. In the 'indicator-absent' trials, the globe and pointer were not presented and subjects viewed just the empty screen for the same amount of time.

The experiment began with a learning phase, followed by a criterion-level recognition test, and then the main test. The criterion recognition tests reflected the current experimental condition.

In total, 8 subjects performed this experiment. Ages ranged between 19 and 39 years. Subjects performed both the indicator-present and indicator-absent trials in blocks, the order of which was randomized.

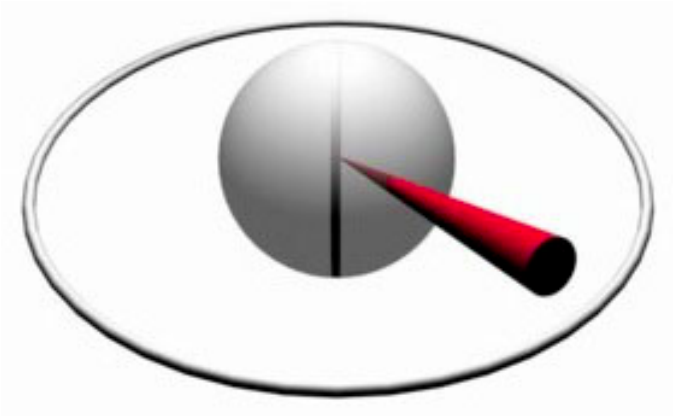

Figure 10. The globe stands in place of the objects and the ring specifies the original inclination of the observers viewpoint. The red pointer indicates the new viewpoint.

Results

Criterion Test

As before the number of attempts required to achieve criterion level performance was noted and the mean number of attempts are shown in Table 4. The most difficult block was, as expected, the first. Subsequent blocks became easier, as in previous experiments. A repeated measures ANOVA with block (1,2,3) and indicator (present, absent) as within-observers factors showed that the effect of block number was significant ($F_{2,14}$=11.6, p<0.001). The effect of indicator showed a trend towards significance ($F_{1,7}$=4.5, p=0.07), but the interaction between these two was not significant ($F_{2,14}$=0.1,ns).

Table 4. Mean Number of Attempts Required to Pass Criterion Test in Experiment 4.

Room	Block 1	Block 2	Block 3
Fixed	1.9	1.1	1.0
Rotating	2.1	1.3	1.1

Novel View Identification Test

Once again the proportion of errors are a non-linear function of displacement angle with least errors for the familiar viewing direction and maximum error at the oblique 90° view (Figure 11). Familiar views produced approximately 20% errors for both conditions. An analysis of variance (ANOVA) was performed on the proportion of errors with two repeated measures: indicator (present, absent) and displacement angle (6 levels, one for each displacement bin). This revealed a significant effect of the indicator ($F_{1,7}$=8.5, p<0.05), a significant effect of displacement angle ($F_{5,35}$=24.6,

p=0.001), and no interaction between the two ($F_{5,35}$=1.2, p=0.3).

The response times are also related to the displacement angle but appear to level out at 90° displacements. A similar analysis of variance (ANOVA) performed on the response times as on the error rates. This revealed no significant effect of the indicator ($F_{1,7}$=0.4, p=.5), a significant effect of displacement angle ($F_{5,35}$=8.7, p=0.001), and no interaction ($F_{5,35}$=0.1, ns).

Discussion

These results show a reduced yet significant facilitation for the identification of novel views when prior indication of the new viewpoint was given. This reduced facilitation compared to, for example, the room-present condition of Experiment 1 may reflect the level of abstraction used in the current experiment. For example, the only detail present during training and testing was the geometry of each shape (appropriately transformed by the viewing transformation). The facilitation therefore consisted entirely in the information about the new viewpoint relative to the original viewing directions. Accidental feature conjunctions between the shape and its environment that might facilitate performance were absent. Another reason for a reduced facilitation was that the view indication was given prior to the shape and not simultaneously. The additional memory load coupled with the possibility that subjects may have been inattentive of the globe prior to the appearance of the test shape may have contributed to its reduced effectiveness. Regardless of these considerations, the view indicator still appears to serve a useful role in increasing accuracy. The response times were once again unaffected by the presence of the view indicator.

Experiment 4 is similar to that of Shepard & Cooper (1982) who also investigated view-dependency by giving prior indication of the extent of the transformation of objects. In their experiment however, they found that if a prior indication was given, the response times ceased to depend on the angle of rotation of the test shapes. In our experiment the benefit is one of reduced error rates, with response times remaining unaffected. These differences probably reflect the difference in the task; ours was a 3D-depth rotation task in which subjects did not initiate the presentation of the test shapes.

Conclusions

We studied the effects of an environmental context on learning and recognition of complex geometrical shapes. Interactive control during learning allowed subjects to utilize natural learning strategies of what have proven to be extremely difficult shapes to learn. Since subjects were allowed to move their position, depth information regarding the shapes was available from motion parallax. Environmental influences were studied by using a naturalistic computer generated 3D context in which the shapes were embedded.

In general, recognition of the shapes after interactive learning was found to be view dependent. More specifically, identification errors were lowest for the small displacements, increased to a maximum at 90° and decreased again for 180° . From our own impression of performing these tests and the comments made by subjects, a predominant strategy in learning the shapes was to encode distinctive features that differentiated one shape from another. These features included, for example, distinctively long and short arm combinations, tight corners between arms, and other configurations that made the shapes look like familiar objects. These features however, are specific to a given viewing perspective. A 90° rotation in viewpoint maximally perturbs or destroys them. The fact that 180° views were more easily recognized than 90° views is probably because mirror-reversed forms of these features appear when a shape is viewed from the back (particularly when little or no self-

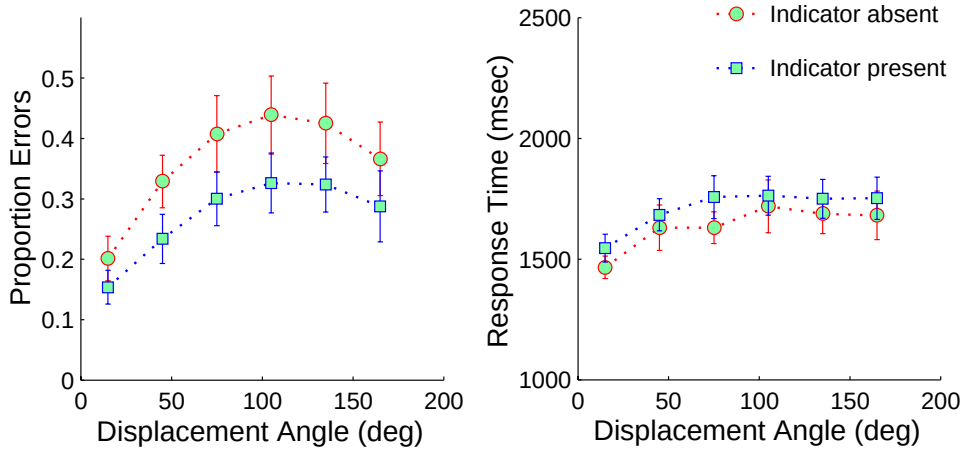

Figure 11. Proportion correct responses and response times for Experiment 4.

occlusion occurs). These results provide evidence that novel shape learning is viewer-centered as reported previously for these kinds of shapes (e.g. Rock, DiVita & Barbeito, 1981).

We observed systematic differences in accuracy as a function of the manipulations of several extrinsic factors; factors that are independent of the objects' shape. The first such factor was the three-dimensional context (a room) in which the shapes were learned. We reasoned that if subjects extract and encode shape independently of the background then subsequent recognition tests should be unaffected by its absence. On the contrary, we observed that a visual context facilitated the accuracy of recognition for all angular displacements. The second experiment showed that this effect was not an artifact of removing the scene from view during recognition. Here, two conditions were used; one in which the room stayed in a fixed spatial relationship to the shapes and another in which the room rotated around the shapes. In the latter case subjects could not rely upon the room providing a fixed reference frame. This appears to be reflected in the results because when subjects were given a fixed room their error rates were always lower than when no such fixed spatial relationship could be relied upon. The benefit of a stable backdrop was most apparent for the oblique views (those outside of the restricted direction used during training) and suggests once again that for familiar views shape recognition was optimal. For oblique views however and especially for 90° views the shapes look very different and error rates were high. The stability of the room-shape relationship appeared to be most useful for these views.

These results agree with the scene-based facilitation in view-dependent object recognition reported by Hinton & Parsons (1988). One possibility for such facilitation is that conjunctions of features between the room and the test shape may provide useful cues for recognition. This latter explanation however cannot explain why we observed a facilitation of context for view displacements in excess of 90 degrees. In such cases, the region of background seen during learning is no longer visible, and subjects would not be able to encode objects with respect to it.

A more plausible explanation for facilitation observed by having a stable backdrop during learning and recognition is that it tells subjects by how much their view has changed and in which direction. This would agree with the findings of Simons & Wang (1998) in the sense that subsidiary, or extrinsic, cues are found to affect spatial cognition judgements (see also Simons, Wang & Rodenberry, 1998). In Simons & Wang's experiments the task involved judgements of spatial layout of several objects and subjects ability to say whether a scene had changed or not was found to be improved by knowing ones movements around the test scene (the test scene was hidden during this time.) However, this facilitation was attributed to spatial updating which is clearly not the case here. Here, the subjects did not actually move and they did not see their motion around the shapes. The significance of these results and those of Simons & Wang's is that subsidiary sources of information contribute to shape recognition and that such information is probably most useful when it allows a mental representation to be 'adjusted' in some way.

Our final two experiments bear this out. In Experiment 3 and 4, we isolated the shapes from the room and used abstract indicators of what we thought was the pertinent information being supplied by the room. Namely, we indicated to subjects what was their original (learned) viewpoint (Experiment 3) and what would be their new (test) viewpoint (Experiment 4). In both cases, we found similar facilitatory effects of this extrinsic information. Moreover, in Experiment 4, the new viewpoint indicator was shown prior to the presentation of each test shape, and subjects therefore could not encode the shapes in conjunction with this indicator or any another extrinsic detail that remained visible during the learning and the test phases. Our results therefore suggests that subjects benefit from being told the extent of viewpoint transformation when asked to identify objects from novel views.

What are the plausible mechanisms for such extrinsic information to facilitate shape recognition? Assuming that observers can differentiate between a set of shapes from a familiar, but limited, range of directions, this differentiation could depend on identifying features unique to each shape, which may be viewpoint specific. From unfamiliar directions, the subject would have to search through a limited space of possibilities to determine if there is a match between the viewed configuration of features and the stored representations. Searching through this space would therefore entail accounting for the unknown viewpoint transformation, which increases uncertainty and introduces a greater potential for making mistakes (e.g. identifying a stimulus with the wrong representation, or not finding a match in the allotted time). In such a scenario, extrinsic information may serve to limit the search space by specifying the viewpoint transformation for the given stimulus. If the subject knows what particular transformations are to be discounted, then a given pattern may be processed and matched against a smaller search space. Since a transformation of the input pattern is still necessary, the response times would be elevated compared to those patterns that do not require transformation. Decision uncertainty however is reduced. Under this mechanism, any extrinsic information that helps limit the search space can reduce error rates but not necessarily reduce viewpoint dependence. Although it is difficult to claim that all manipulations reported here tap into the same mechanism for reducing this search space (and a strict comparison between them is not possible), we do believe that we have shown that this search space can be reduced by extrinsic information.

Appendix A

The Virtual Environment

Virtual environment simulations are becoming an increasingly popular alternative to real scenes in the study of spatial cognition (e.g., Maguire et al., 1998; Christou & Bülthoff, 1999; Bülthoff & van Veen, 2001). For the current experiments we devised a learning and test paradigm based entirely within a simulated, richly decorated familiar environment (Figure 1). It seems reasonable to assume that the contextual richness of an environment enhances its memorization and its ability to serve as a frame of reference. Also, realistic textural detail and illumination enhance 3D cues to shape and spatial layout. The virtual environment 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 produced from several simulated lightsources was modeled using Lightscape (AutoDesk). This software generates realistic shading by physically-based calculations that include not only first-order (direct illumination) effects such as cast shadows but also takes account of second order components produced by the interreflection of light between diffusely reflecting surfaces.

The Geometric Forms

The 3D target shapes were computer generated wire-like shapes as used by Bülthoff & Edelman (1992). These consisted of 10 cylindrical segments starting from a vertical 'stem' (see Figure 1). To test the influence of environmental reference frames in the encoding of geometrical shape that is clearly view-dependent we used forms that were highly self-similar (with no uniquely identifiable features). In order to generate geometrical forms that were similar, the 'arms' of each paperclip were produced by joining, with cylindrical segments, 9 points on the surface of bounding 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 a set of 32 similar looking shapes. Noticeably degenerate instances of these shapes (e.g., that included self-intersection) were excluded from the set.

Interactive Manipulation of View

Self-control of movement was an important feature of these experiments. There is increasing evidence that active control and manipulation of movement improves memory for spatial structure and shape (Christou & Bülthoff 1999; Harman, Humphrey & Goodale, 1999). This may be because active control of movement allows subjects to perform behavioral learning patterns that optimize the perceptual information available to them. Interactive control of movement was facilitated using real-time 3D graphics (using Silicon Graphics IRIS Performer libraries and OpenGL). Observer movements through the scene were input using a Logitech 6 degree of freedom SpaceMouse (Logicad3d) that was used to control the simulated viewing direction and position in the environment. By applying pressure on the SpaceMouse in the direction they wished to move, subjects received the impression of movement through the scene. Pushing the cap forward moved observers' view of the scene 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.

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. Viewing distance for all experiments for both learning and test stages was constant at 80cm producing a visual angle of approximately 34.5° . The video update frequency (i.e., the number of refresh updates of the scene) was approximately 30 Hz, giving the impression of smooth movement.

Acknowledgments

This research was supported by a research scholarship paid to Chris Christou by the Max Planck Society, Germany.

Commercial relationships: none.

References

Bartram, D. J. (1974). The role of visual and semantic codes in object naming. *Cognitive Psychology*, 6, 325-356.

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

Biederman, I., & Cooper, E.E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception*, *20*, 585-593. [PubMed]

Biederman, I., & Cooper, E.E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 121-133.

Biederman, I., & Gerhardstein, P.C. (1993). Recognizing depth-rotated objects: Evidence and conditions for 3D viewpoint invariance. *Journal of Experimental Psychology: Human perception and performance*, *19*, 1162-1182. [PubMed]

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

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex,* 5(3), 247-260. [PubMed]

Bülthoff, H.H., & Christou, C. G. (2000). The perception of spatial layout in a virtual world. In S.W. Lee, H.H. Bülthoff, T. Poggio (Eds.), *Biologically Motivated Computer Vision. Proceedings of the First IEEE International Workshop, BMCV 2000*. Lecture Notes in Computer Science 1811, (10-19), Berlin: Springer.

Bülthoff, H. H., & van Veen, H. A.H.C. (2001). Vision and Action in Virtual Environments: Modern Psychophysics in Spatial Cognition Research. In M. Jenkins, L. Harris (Eds.), *Vision and Attention*, New York: Springer.

Cave, C. B., & Kosslyn S. M. (1993). The role of parts and spatial relations in object identification. *Perception*, *22*, 229-248.

Christou, C. G., & Bülthoff, H. H. (1999). View dependence in scene recognition after active learning. *Memory & Cognition*, *27*, 996-1007. [PubMed]

Christou, C. G., & Bülthoff, H. H. (2000). Using realistic virtual environments in the study of spatial encoding. In C. Freksa, W. Brauer, C. Habel, K.F. Wender. (Eds.), *Spatial Cognition II. Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications*, Lecture Notes in Artificial Intelligence 1849, (317-332) Berlin: Springer.

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics, 62*(3), 425-451. [PubMed]

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

Harman, K. L., Humphrey, G. K. & Goodale M. A. (1999). Active manual control of object views facilitates visual recognition, *Current Biology* 8, 9:1315-1318.

Hinton, G. E., & Parsons, L. M.(1988) Scene-based and viewer-centered representations for comparing shapes. Cognition, *30* (1) ,1-35. [PubMed]

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. [PubMed]

Humphrey, G. K., & Khan, S. C. (1992). Recognizing novel views of three-dimensional objects. *Canadian Journal of Psychology*, 46, 170-190. [PubMed]

Jolicoeur, P., & Humphrey, G. K. (1998). Perception of rotated two-dimensional and three-dimensional objects and visual shapes. In V. Walsh and J. Kulikowski (Eds.), *Perceptual Constancy: Why things look as they do*, (pp. 69-123) Cambridge, UK : Cambridge University Press.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition*, *13*, 289-303. [PubMed]

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

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

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. [PubMed]

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, Series B*, *200*, 269-294.

Palmer, S. E. (1989). Reference frames in the perception of shape and orientation, In B. E. Shepp & S. Ballesteros (Eds.) Object Perception: Structure and Process, (pp. 121-161New Jersey: Lawrence Erlbaum Associates.

Rock, I. (1973). Orientation and Form, Academic Press: London.

Rock, I. & Di Vita, J., & Barbeito, R (1981). The effect on form perception of change of orientation in the third dimension. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 719-732. [PubMed]

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

Shepard, R. N., & Cooper, L. A. (1982). Mental images and their transformations, Cambridge, MA: MIT Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects, Science, *171*, 701-703. [PubMed]

Simons, D.J., & Wang, R.F. (1998) Perceiving real-world viewpoint changes. *Psychological Science*, *9*, 315-320.

Simons, D.J., Wang, R.X.F. & Rodenberry, D (1998) Object recognition is mediated by extraretinal information. *Perception and Psychophysics*, *64*(4), 521-530. [PubMed]

Srinivas, K. (1993). Perceptual specificity in nonverbal priming. *Journal of Experimental Psychology, Learning, Memory and Cognition*, *19*,582-602

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

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, 32, 193-254. [PubMed]