Perception, 1994, volume 23, pages 595-614

# Object recognition under sequential viewing conditions: evidence for viewpoint-specific recognition procedures

Rebecca Lawson, Glyn W Humphreys, Derrick G Watson
Cognitive Science Research Centre, School of Psychology, University of Birmingham, Birmingham B15 2TT, UK

Received 28 June 1993, in revised form 9 December 1993

**Abstract.** In many computational approaches to vision it has been emphasised that object recognition involves the encoding of view-independent descriptions prior to matching to a stored object model, thus enabling objects to be identified across different retinal projections. In contrast, neurophysiological studies suggest that image descriptions are matched to less abstract, view-specific representations, resulting in more efficient access to stored object knowledge for objects presented from a view similar to a stored viewpoint. Evidence favouring a primary role for view-specific object descriptions in object recognition is reported. In a series of experiments employing line drawings of familiar objects, the effects of depth rotation upon the efficiency of object recognition were investigated. Subjects were required to identify an object from a sequence of very briefly presented pictures. The results suggested that object recognition is based upon the matching of image descriptions to view-specific stored representations, and that priming effects under sequential viewing conditions are strongly influenced by the visual similarity of different views of objects.

## 1 Introduction

Human vision enables a large number of everyday objects to be rapidly and accurately identified as belonging to pre-existing perceptual categories. Humans can then access stored functional and associative knowledge about these objects, in order to evaluate their significance for action. Stored view-invariant visual representations of objects are usually thought necessary, to enable semantic and perceptual associations to be learnt and accessed from any view of an object presented (eg Marr 1982). However, such view-invariant representations, if they exist, may themselves be activated via access to stored representations that are less abstract and that are view sensitive. Single, view-invariant representations need not be derived on-line from objects in order for object recognition to occur.

We have been investigating how different retinal images of an object are recognised as having the same structure, and are assigned to the same perceptual category, as measured in (for instance) experiments involving picture-picture matching. This is the problem of visual object constancy. We have also considered how the description of a single, unique retinal image of an object is matched to a stored visual representation of that object, despite very great changes in the projected image as an object is transformed in depth, in picture naming tasks. Humans apparently have few difficulties with picture and object naming in everyday life, except when presented with very unusual views, such as foreshortened views of objects. For example, Biederman and Gerhardstein (1993) investigated long-term priming of object naming. They found equal facilitatory priming for target views identical to the prime view, and for target views rotated by up to 135° in depth from the prime view of the object (where all views minimised changes in the availability of parts and foreshortening). This result is perhaps unexpected. Profoundly different images of an object are projected from different depth-rotated viewpoints—global shape, size, and spatial relations all frequently undergo major changes. In contrast, different plane-rotated views of an object only vary with respect to positional relations relative to the viewer; global shape and size, and the identity and relative positions of component features of the object remain the same. Nevertheless, Jolicoeur (1985) found that naming latencies for plane disoriented objects were up to 200 ms slower than those for normally oriented, upright objects, and naming latencies increased as a linear function of the disorientation of the image. Such results suggest that humans do not find compensating for plane rotation to be simple and automatic; the effects of depth rotation may be expected to be at least as severe.

With respect to depth rotation, views which foreshorten objects seem particularly difficult to recognise, as suggested both from everyday experience, and from studies of neuropsychological patients. A selective disadvantage in matching and recognising foreshortened views of objects has been reported in the neuropsychological literature (Warrington and Taylor 1973, 1978; Humphreys and Riddoch 1984). These authors found that patients with right parietal lesions had particular problems in matching and recognising foreshortened views of objects; matching and recognition of more standard views was relatively unimpaired.

Theories of human object recognition differ in their explanation of how object constancy is achieved, and of the possible effect of foreshortening upon recognition. Some theorists propose that the image description which is matched to a stored representation is completely view-invariant (eg Marr 1982); others, however, assume that matching involves view-specific image descriptions (Tarr and Pinker 1989; Edelman and Bülthoff 1992; Biederman and Gerhardstein 1993). Three different theories of human object recognition are briefly considered below, with particular emphasis being laid upon the predictions of the effects of viewpoint upon recognition efficiency.

### 1.1 *Marr's account*

Marr and coworkers (Marr and Nishihara 1978; Marr 1982) proposed that object representations are stored as three-dimensional object-centred models. An object-centred image description is derived on-line from the image, on the basis of a frame of reference intrinsic to the object (based around the main axis of the object), and independent of the position of the viewer relative to the object. This view-invariant image description can be matched equally rapidly to the stored object model, whatever the viewpoint of the object presented. Since the main axis of the image defines the coordinate system of the image description, the axis must be assigned prior to the matching of the image description to an object model. Hence the axis must be derived without access to stored knowledge about the structure of the object.

Marr was aware of neuropsychological evidence showing that foreshortened views are harder to recognise than other views. Accordingly, he suggested that the process of deriving the main axis of an image is sensitive to viewpoint. He proposed that image cues, in particular elongation, are used to assign the main axis. When the main axis of the object points directly towards the viewer, the object is foreshortened, and appears small in extent, relative to other viewpoints. The longest apparent axis of a foreshortened view might not even correspond to the main axis of the object. This should render recognition less efficient, relative to when the main axis of the object is easy to derive, allowing the coordinate system for the appropriate image description to be established efficiently. From this it follows that viewpoint effects on object recognition should emerge only if the image description is difficult to derive, for instance because the view foreshortens the main axis of the object. In all other cases, different views of an object should access the same stored representation of the object with roughly equal efficiency.

### 1.2 *Biederman's account*

Biederman (1987) has proposed that properties of edges in the image are detected, specified by five nonaccidental relations—parallelism, symmetry, collinearity, curvilinearity and cotermination. The detection of these properties is assumed to be viewpoint-invariant over the range of views for which the edges are visible, and metric variation in the image is not assumed to affect recognition. The view-invariant features define the presence of volumetric components of the object, which Biederman termed geons. Geons are a set of around fifty simple volumetric primitives which roughly correspond to the parts of the object. Image descriptions are based on the small set of geons representing the main parts of an object, together with a coarse, qualitative description of their spatial relations to each other (ie 'top of', 'parallel', etc). This structural description is matched to a stored object representation. Biederman predicts equal efficiency of object recognition for two different views if the same parts are readily available in both views, so that the same structural description is achieved from both views. Different views of an object will also prime each other as much as identical views, provided that the same component parts are present in the same spatial arrangement to construct the same structural description. This proposal has been supported by evidence of equal priming effects on object naming for objects presented from different views, but revealing the same parts across both views, compared with priming of identical views of an object (Biederman and Gerhardstein 1993). However, in other studies, view-specific effects on object priming have been observed (Humphrey and Khan 1992), suggesting some view-specificity in object recognition.

### 1.3 *View-specific representations*

In a third approach to object recognition it is suggested that the image descriptions derived from objects are view-specific, and are matched to one of a number of view-specific representations which are stored for each object. This proposal has been supported by human experimental research by Tarr and Pinker (1989), in which subjects learnt to name novel stimuli at different orientations. They found that naming latencies for stimuli at different, familiar orientations were approximately equal. However, subjects were increasingly slow to name stimuli rotated further from the nearest familiar orientation. Tarr and Pinker suggested that the different, familiar orientations of a stimulus are represented by a number of view-specific representations. An image description must be rotated to align it to the nearest stored representation before matching can occur.

Neurobiological evidence from single-cell recording studies of monkeys by Perrett and coworkers (Perrett et al 1992, 1993) provides further evidence against the view that matching to stored knowledge involves object-centred representations. They investigated face-sensitive cells in the superior temporal sulcus of the monkey. The majority of these cells generalised their response across image size, position, and plane orientation, and across different lighting conditions. However, most cells were still view selective, generally preferring either front or profile views of the face; such cells did not fully achieve object constancy. The results point to the special difficulty in compensating for depth rotation. A small minority of temporal cortex cells did respond to a range of views of their preferred object, but Perrett et al (1992, 1993) suggested that these cells might be activated via view-specific cells, in a hierarchical fashion. This proposal was supported by the finding that response latencies to the onset of stimuli for these view-invariant cells were rather slower than the latencies of the more common view-specific cells.

Our own studies have focused upon the problem of achieving constancy in object recognition when objects undergo depth rotation relative to the viewer. This explores the limits of human object constancy, as the transformation of the two-dimensional image across different depth-rotated views of an object is almost invariably nonlinear and unpredictable. We have used a number of different tasks to assess object constancy across rotations in depth, aiming to tap different levels of representation in vision. Some tasks, such as sequential picture-picture matching, did not require access to structural descriptions of familiar objects (Lawson and Humphreys 1994). Viewpoint was found to play an important role in determining the efficiency of picture-picture matching, suggesting that view-sensitive representations are involved in achieving matching. Other tasks did necessitate access to stored knowledge, for instance word-picture verification (which revealed a specific foreshortened view disadvantage that was much greater for silhouettes than for matched line drawings), picture naming (which again revealed a specific foreshortened view disadvantage, and which provided evidence for view-specific priming effects), and recognition from a rapid temporal sequence of pictures (Lawson 1994). The last task was employed in the series of experiments described below.

In the current series of experiments we investigated how stored knowledge is accessed from a sequence of different views of the same object. These experiments allow a relatively ecologically valid investigation of the processes involved in object recognition, since information had to be combined across a number of different views, which is a common experience in everyday viewing. It is possible that humans learn to associate different but visually similar views of an object which often occur in close temporal succession (Edelman and Weinshall 1991; Seibert and Waxman 1991). We also investigated recognition under conditions of apparent motion. Until now the emphasis of object recognition research has been on the interpretation of static images and on the investigation of long-term priming effects.

Since most theories of object recognition either ignore, or specifically exclude, discussion of moving objects, predictions from these theories are necessarily rather tenuous. However, since the stimuli themselves were static (albeit extremely briefly presented) pictures, visual processing of the individual stimuli is covered by these theories. Apparent motion per se should not eliminate object processing of the type described by the theories.

We used line drawings of different, depth-rotated views of familiar objects (listed in the appendix). The angular difference between two views of the same object was taken as an indirect measure of their visual similarity. This angle is only an approximation to psychological visual similarity (eg mirror-reflected images at 30° and 150° are probably processed and perceived in very similar ways by humans, yet they are separated by a large, 120° rotation in depth), but it nevertheless provides an atheoretical measure. Views rotated by only a small angle from each other are usually visually similar. Two visually dissimilar viewpoints generally involve a large depth rotation.

The main manipulation made in the studies was to contrast the identification of objects after the presentation of a structured sequence of views (as would occur if the objects were actually rotated) with identification after the presentation of the same individual views, but in a random order. We asked several basic questions: (i) Is identification facilitated when a structured sequence of views is presented, relative to when subjects see a random sequence (experiments 1 and 2)? (ii) Is local similarity between consecutive pairs of views crucial for identification, or is maintenance of the global sequential order more important (experiment 3)? (iii) Do effects on identification directly match those on perceived object motion (experiment 4)? (iv) How does identification fare with sequential views of objects relative to when subjects received repeated static views (experiment 5)?

Evidence for facilitated recognition of structured versus random sequences of views would suggest that object recognition gains from the temporal ordering of different views. There may be several reasons for this. One is that structured sequences of views generate a better perception of apparent motion than random sequences. This may affect performance, either because motion provides a direct source of input to the recognition process (eg if there is a recognition system for moving objects separate from that for static objects; see Humphreys et al 1993), or because motion can be used to help recover object descriptions, which are fed into the recognition process that serves for static objects. The latter account seems more likely, in view of evidence suggesting a distinction between a pattern recognition "what" system, and a "where" system that analyses location and motion (cf Ungerleider and Mishkin 1982; Schiller et al 1991). Furthermore, motion perception may be influenced by local similarity between consecutive views or by the global coherence of the sequence. These factors could also influence recognition directly, independent of any perceived motion. For instance, locally similar views might activate the same stored view-specific object representations, facilitating performance relative to when different view-specific object representations are activated. Biederman (1987) and Tarr and Pinker (1989) predict that two similar views presented consecutively could activate the same, perhaps rather coarsely coded, view-specific structural description. For Biederman, the same set of geons would be activated, in the same spatial relations to each other. For Tarr and Pinker, both views would access the same stored object representation. In contrast, if recognition involves view-independent image descriptions and stored object representations, performance should be as good with globally incoherent and locally dissimilar images as with globally coherent and locally similar sequences, providing that the individual views presented are equally recognisable. This is because all the individual views of an object should give rise to the same object-based image description, and they should all access the same view-independent stored representations (cf Marr 1982). Hence any effects impinge on theories of object recognition.

## 2 Experiment 1

In the first experiment we investigated whether the efficiency of object identification varied when consecutive, briefly presented views of an object were presented in a structured order (so that the object appeared to rotate in depth), relative to when the same views appeared in a random order.

The stimulus presentation duration was also manipulated. A benefit for structured sequences might be highly sensitive to picture duration. For instance, if an object appeared to rotate faster than is ever observed in normal, ecological circumstances, a process which has evolved to take advantage of view transitions might be unable to integrate information rapidly enough to use the available information. Conversely, if the picture duration was too long, motion might appear unnatural and jerky. This might affect performance if motion perception contributed to the recognition process (see section 1). Pilot studies in which picture duration was varied revealed two further constraining factors. If the duration was reduced much below the shorter, 30 ms duration used in experiment 1, subjects were unable to identify any objects; conversely, if the duration was increased much beyond the longer, 45 ms duration used in experiment 1, subjects were close to ceiling on object identification, and may have been able to identify the object directly from a single picture. Hence pragmatic constraints focused attention on the two durations used.

### 2.1 *Method*

The method was identical for all experiments, unless otherwise mentioned.

2.1.2 *Subjects.* Forty-eight subjects from the University of Leuven participated in the experiment for course credit. The subjects were aged between 18 and 35 years, and had normal or corrected-to-normal vision. Each participated in only one experiment, and was naive as to the purpose of the experiment.

2.1.2 *Stimuli.* A set of twelve views of thirty-four familiar objects was produced (see figure 1; the full set of objects are listed in the appendix). All the objects possessed an unambiguous main axis of elongation. The angle of view was defined with respect to the line of sight of the viewer relative to the main axis of the object. The 0° view revealed the main axis perpendicular to the line of sight of the viewer. In the 90° foreshortened view, the main axis of elongation pointed directly towards the viewer and revealed the front of the object or the view with the most important feature to the fore. For instance, the 90° view of a camel was depicted facing head on, and the fork was depicted with the prongs closest to the viewer. Each view was separated by a 30° rotation in depth. Objects were rotated about the vertical axis running through their centre point. The set of twelve views covered a full 360° horizontal depth rotation of each object.

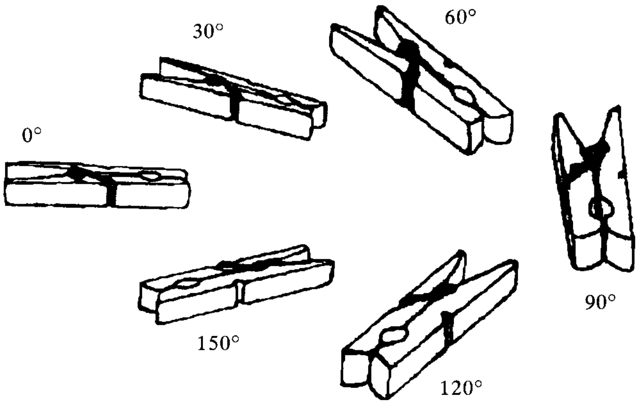

**Figure 1.** The six depth-rotated views of the clothes peg presented in experiment 1.

All stimuli were line drawings, scaled to occupy a square of 5 cm × 5 cm. The pictures were produced by tracing photographs of either the objects or scale models of the objects. Photographs were taken from a slightly elevated angle, which was maintained as a constant during depth rotation of the object. Twelve different pattern masks were also produced (see figure 2). Each occupied a square of 6 cm × 6 cm, and covered an area greater than the extent of any picture. The masks were composed of small, overlapping random sections of the pictures used in the experiment, which were rotated. The masks did not contain any recognisable object parts.

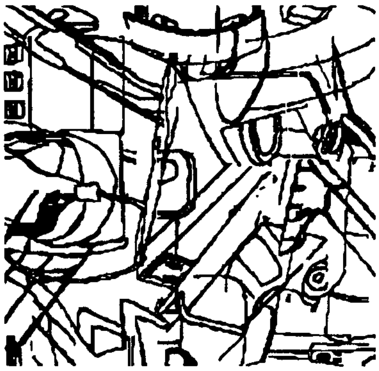

**Figure 2.** One of the twelve pattern masks used in the experiments. Each pattern mask covered an area greater than the extent of any picture.

2.1.3 *Design and procedure.* A computer (NEC Multisync 3D screen, IBM compatible 486) was used to display the stimuli and record responses. On each trial, six views of an object were shown, at 0°, 30°, 60°, 90°, 120°, and 150°. There were two picture presentation sequences. In the structured sequence, the pictures were shown in order, beginning with the 0° view, then 30°, 60°, 90°, 120°, and 150°. The object thus appeared to rotate in depth. In the random sequence, subjects were shown the same six views, but in a randomly generated sequence which was different for each object and for each subject.

The set of thirty-four objects was divided into two sets of seventeen objects, which were presented in different blocks. All subjects completed two blocks of seventeen trials, one random and one structured. The order of presentation of trials in each block was random. The order of presentation of blocks (structured or random), picture duration (30 ms or 45 ms), and the set of objects assigned to each block was balanced across subjects. The experiment lasted about 15 min.

Subjects were given three blocks of practice trials before the start of the experiment, which showed different objects from those used in the experimental blocks. The practice trials involved the same sequence presentation as the first block of experimental trials for that subject. Prior to the second block of experimental trials, subjects were informed of the change in the sequence presentation, but were advised that because the pictures were displayed for such a short time, there would be little visible difference between the two blocks.

The procedure for each trial was as follows. A mask was presented in the centre of the screen; subjects pressed the space bar on the keyboard when they were ready to begin a trial. Six consecutive pictures, each followed by a different mask, were then shown. The mask which followed the final picture remained on the screen until the start of the next trial. Apart from the first and the last mask, all the pictures and masks were shown for the same duration, either 30 ms or 45 ms. When the final mask appeared, subjects pressed the space bar a second time, then typed in the name of the object which they thought that they had seen. When they were satisfied with their response, they pressed the return key, and a new mask appeared, ready for the start of the next trial. Subjects were instructed to try to identify the object depicted by the pictures. They were told that they would see six different views of the same object. Subjects were encouraged to guess the identity of the object if they were unsure, but they could respond "Pass" if they were completely unable to identify the object.

### 2.2 *Results*

In all the experiments reported here, subjects scoring less than 20% or more than 80% correct across the two blocks of trials were replaced (scores of less than seven or more than twenty-seven out of thirty-four objects correct, respectively). This ensured that the subjects included in the analysis were not scoring at floor or at ceiling levels, which would have reduced sensitivity to the experimental manipulations. Nine subjects were replaced by use of these criteria, all of whom scored less than 20% correct (scoring 0%, 3%, 3%, 6%, 6%, 9%, 9%, 15%, and 15% correct; only one subject, scoring 3%, received the longer 45 ms picture duration).

The percentages of correctly identified objects in each condition are given in table 1. An ANOVA was carried out on the mean correct identification responses in each block of trials. For this, and for all the other experiments reported here, the data were log-linear transformed before the ANOVA was performed. For the analysis over subjects, a log-linear transformation was performed on the total number of correctly identified objects from the seventeen objects seen in each block by each subject. For the analysis over items, a log-linear transform was performed on the number of subjects correctly identifying a given object, from the total of twelve subjects who were presented with that object, for a given sequence presentation, at a given picture duration. In this and in all the following experiments, the results for analyses by-subjects and by-items are reported, with $F$-values denoted as $F1$ and $F2$, respectively. There was one within-subjects factor, sequence (whether the picture sequence shown was structured or random), and one between-subjects factor, picture duration (30 ms or 45 ms).

The main effect of sequence was signficant ($F1_{1,46} = 29.235$, MSe = 8.195, $p < 0.001$; $F2_{1,33} = 25.639$, MSe = 14.578, $p < 0.001$). More items were identified with a structured than with a random sequence. The main effect of picture duration was also significant ($F1_{1,46} = 27.909$, MSe = 16.473, $p < 0.001$; $F2_{1,33} = 65.465$, MSe = 39.167, $p < 0.001$). More items were identified at the longer, 45 ms picture duration than at the shorter, 30 ms picture duration. The sequence × picture duration interaction was not significant ($F1_{1,46} = 1.766$, MSe = 0.495, $p > 0.1$; $F2_{1,33} = 0.331$, MSe = 0.174, $p > 0.5$).

**Table 1.** Mean percentage correct responses (with mean number correct out of seventeen trials, and standard deviations in parentheses) for the structured and random blocks in experiments 1 and 2.

| | Structured | Random | Difference |
|---|---|---|---|
| *Experiment 1* | | | |
| 45 ms/picture, 6 pictures/sequence | 61.5 (10.46, 1.9) | 50.7 (8.62, 2.9) | +10.8 |
| 30 ms/picture, 6 pictures/sequence | 44.4 (7.54, 2.7) | 28.9 (4.92, 2.3) | +15.4 |
| *Experiment 2* | | | |
| 30 ms/picture, 12 pictures/sequence | 63.0 (10.7, 2.6) | 41.9 (7.13, 2.4) | +21.1 |

## 2.3 *Discussion*

The results show a consistent benefit for identifying objects from a coherent, structured sequence compared with a random sequence. The magnitude of the benefit was an increase in identification rates of 15% and 11% at the 30 ms and 45 ms picture durations, respectively. Identification was also more accurate overall at the longer, 45 ms picture duration, but there was no interaction between sequence type and picture duration.

The benefit for structured sequences could have arisen for various reasons, which we explore in the following experiments. However, one possible artefactual reason is that the structured sequence always began with the same view of the object, while the random sequences began with different views. Subjects may have identified fewer objects with the random sequences because they could not predict the first view in the sequence (or indeed any other view in the sequence). Note that this factor could affect performance irrespective of whether consecutive views followed a coherent structured sequence. However, were this the case, then structured sequences should always benefit relative to random sequences, even when consecutive views were locally dissimilar. This was investigated in experiment 3. Also, if the predictability of different views was of paramount importance, performance might be expected to benefit if the structured sequence began with a view that was easy to recognise rather than difficult to recognise. In two further experiments, we investigated this, and found no effect of whether the sequence began with a view that was easy or hard to recognise (Lawson 1994). For now, we conclude that objects are easier to identify from structured sequences of views than from random sequences.

## 3 Experiment 2

The structured sequence benefit in experiment 1 was consistent, but not very large. In experiment 2 we attempted to replicate and specify further the effect, by investigating whether an even greater structured sequence benefit would occur if the presentation sequence was longer. Accordingly, a twelve picture sequence was used. There were two reasons for predicting a stronger structured sequence benefit in experiment 2. (i) With a longer sequence, subjects have more opportunity to 'lock onto' the image of the rotating object. In experiment 1, the sequence may have been too short to track a particular feature or the overall shape of the object. (ii) The structured and random sequences in both experiments 1 and 2 included the same set of pictures. Many short random sequences would contain 'accidentally structured' sequences, in which similar views of an object were presented consecutively. Longer random sequences would have fewer such accidentally structured sequences. For the twelve picture sequence, the average difference in depth rotation between two consecutive views in the random sequence would be slightly greater (78°) than for the six picture sequence (70°) used in experiment 1. However longer 'accidental' sequences, with each successive view rotated 30° from the previous view, would be less likely for longer random sequences. Only the shorter, 30 ms picture duration used in experiment 1 was employed, since increasing the number of views in each sequence may increase the overall ease of identification. Ceiling effects might have occurred if a longer picture duration were used.

### 3.1 *Method*

There were twenty-four subjects. All pictures were presented for 30 ms, and twelve views of each object were presented on each trial. In the structured sequence, the 0° view was shown first, then 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, and 330° views, so that each consecutive view was rotated 30° in depth. The same views were shown in the random sequence, but in a random order, which was different for each object, and for each subject.

### 3.2 *Results*

Two subjects were replaced for failing to score between 20% and 80% correct overall (scoring 15% correct and 85% correct). The data from the twenty-four subjects presented with trials at the 30 ms picture duration only, in experiment 1, were included with the data from experiment 2 in an ANOVA, to allow a between-subjects comparison of the relative efficiency of identification of short (six picture) and long (twelve picture) sequences. There was one within-subjects factor, sequence, and one between-subjects factor, sequence length (either six or twelve pictures in a sequence).

The percentages of correctly identified objects in each condition are given in table 1. The main effect of sequence was significant ($F1_{1,46} = 45.722$, MSe = 15.506, $p < 0.001$; $F2_{1,33} = 42.069$, MSe = 30.555, $p < 0.001$). More items were identified with a structured sequence than with a random sequence. The main effect of sequence length was also significant ($F1_{1,46} = 20.295$, MSe = 11.717, $p < 0.001$; $F2_{1,33} = 27.493$, MSe = 26.152, $p < 0.001$). More items were identified with a long compared with a short picture sequence. The sequence × sequence length interaction was not significant ($F1_{1,46} = 0.407$, MSe = 0.138, $p > 0.5$; $F2_{1,33} = 3.293$, MSe = 1.669, $p > 0.07$).

### 3.3 *Discussion*

The structured sequence benefit found in experiment 1 was replicated, with 21% more objects being identified with a (twelve picture) structured sequence, compared with a random sequence. In addition, more pictures were identified overall with the longer, twelve picture sequence, compared with the comparable six picture sequence (and 30 ms picture duration) tested in experiment 1. There was no interaction between sequence type and sequence length, although there was a trend for the structured sequence benefit to be greater for the longer, twelve picture sequence.

## 4 Experiment 3

In experiments 1 and 2 the benefit for object recognition of presenting visually similar views consecutively in a sequence was established. This structured sequence benefit is relatively insensitive both to picture duration and to sequence length. In addition, it is insensitive to whether the structured sequence begins with a view that is easy or hard to recognise (Lawson 1994). Thus the structured sequence benefit is not affected by a number of global manipulations of the sequence—picture duration, sequence length, and the position of good and bad views within the sequence.

In experiment 3, we conducted an explicit comparison of identification performance when the views were globally coherent but locally visually different with performance when the views were locally similar but globally incoherent. Globally coherent (GC) sequences were constructed by rotating each consecutive view in the sequence by 60° in depth. GC sequences reduced the similarity between consecutive views compared with that in the structured sequences in the previous experiments (when 30° depth rotations separated consecutive views), but maintained the overall coherent structure of the sequence. GC sequences were contrasted with locally similar (LS) sequences. LS sequences were constructed by presenting pairs of visually similar views (rotated by 30° in depth), whilst consecutive pairs of views were visually dissimilar (rotated by at least 120° to each other, and the direction of apparent motion between consecutive pairs of views, whether clockwise or anticlockwise, was reversed). These manipulations ensured that the LS sequences were globally incoherent, despite possessing local similarity between alternate consecutive views.

Theorists who hold that object recognition is based on the matching of view-independent image descriptions to object-centred stored representations must maintain that the recognition of GC sequences should be good, given that the views involved are individually identifiable. The only proviso to this would be if the structured sequence benefit is dependent on perceived motion, and the large changes in view between consecutive images in the GC condition disrupts motion perception. We assess this in experiment 4. Performance with LS sequences may also depend on how well motion is perceived, but we can note that it is unlikely to be better then with GC sequences (given the large changes in rotation between consecutive pairs of views and the switches in the direction of rotation).

In contrast to this, it is possible that the structured sequence benefit in experiments 1 and 2 occurs because consecutive views are visually similar in structured sequences. Visually similar views may activate the same view-specific representation, facilitating recognition by increasing the activation of that particular object representation (Tarr and Pinker 1989). Performance would then be predicted to be relatively good for LS sequences. If consecutive pairs of views of an object are visually dissimilar, then they may activate two different view-specific representations, reducing the likelihood that any particular representation is activated above threshold. This may occur in the GC sequences, where there is less visual similarity between consecutive views, relative to the LS sequences.

### 4.1 *Method*

There were twelve subjects. All pictures were presented for 30 ms. Each subject saw two blocks of sequences: the GC sequence (0° then 60°, 120°, 180°, 240°, 300°, 30°, 90°, 150°, 210°, 270°, 330°), and the LS sequence (0° then 30°, 210°, 180°, 60°, 90°, 270°, 240°, 120°, 150°, 330°, 300°).

### 4.2 *Results*

No subjects were replaced, since they all scored between 20% and 80% correct overall. The percentages of correctly identified objects in each condition are given in table 2. An ANOVA was carried out with one within-subjects factor, sequence (either LS or GC). The main effect of sequence was significant ($F1_{1,11} = 24.403$, MSe = 9.946, $p < 0.001$; $F2_{1,33} = 44.341$, MSe = 29.904, $p < 0.001$). More items were identified with the LS compared with the GC sequence.

**Table 2.** Mean percentage correct responses (with mean number correct out of seventeen trials, and standard deviations in parentheses) for the locally similar (LS) and globally coherent (GC) blocks in experiment 3.

| | LS | GC | Difference |
|---|---|---|---|
| 30 ms/pictures,<br>12 pictures/sequence | 68.6 (11.67, 3.4) | 41.7 (7.08, 3.3) | +27.0 |

### 4.3 *Discussion*

Experiment 3 revealed a very clear benefit for LS sequences (in which consecutive views were visually similar, although the sequence as a whole was globally incoherent) compared with GC sequences (which were globally coherent, but consecutive views were visually dissimilar). In fact, the size of the benefit for LS sequences was at least as large as the advantage for the globally structured (and locally similar) sequences used in experiments 1 and 2, relative to the random baseline. This suggests that local similarity between consecutive views may be sufficient to generate the structured sequence advantage. A further experiment (Lawson 1994) was undertaken to investigate whether there was any benefit for GC sequences, with global structure but low local visual similarity, compared with random sequences, lacking both global structure and local visual similarity. There was no evidence whatsoever to suggest that objects were more readily identified when presented in the GC sequence relative to the random sequence. This allows us to dismiss any account of the structured-sequence benefit that holds that the global coherence and familiarity of the sequence is crucial: it is not. In fact the benefit appears to be wholly due to local visual similarity between pairs of consecutive views in a sequence.

The present data are difficult to account for in terms of view-independent approaches to object recognition (cf Marr 1982). The same individual views were used in the GC and LS sequences, and it is probable that the perception of motion was stronger for GC compared with LS sequences, since only in the GC sequences was there a single, coherent direction and rate of motion (and see also experiment 4). There is thus no reason for LS sequences to be advantaged. Instead, the data point to the importance of local similarity between consecutive views of objects for object identification. Indeed, these similarity effects also seem highly specific—effects are apparent with 30° depth rotations (with LS sequences) but not with 60° depth rotations (with GC sequences). The variation in similarity between 30° and 60° rotations is difficult to capture in terms of a componential approach to object recognition (Biederman 1987), since, for the stimuli used, the same major components were generally apparent with both rotations (this point is examined in more detail in section 7).

## 5 Experiment 4

Our argument in favour of local similarity between consecutive views rests on there not being an advantage in the strength of perceived motion with LS relative to GC sequences. If perceived motion is important for the structured sequence benefit, and if perceived motion is weaker with GC sequences, this could account for performance with GC sequences being impaired relative to that with LS sequences. We assessed this directly in experiment 4, in which we had subjects rate the degree of perceived motion with four different types of sequence presentation. These were the structured sequence presented in experiment 2 (with each view rotated by 30° in depth from each other, in a structured, coherent manner), the GC sequence (the same as the structured sequence, but with a 60° depth rotation between consecutive views), the LS sequence (with a 30° rotation in depth between alternate pairs of consecutive views, but no overall structure or coherence to the sequence), and a pseudorandom sequence (with at least a 90° depth rotation between consecutive views, and no overall structure or coherence to the sequence). Subjects rated the degree of coherence of perceived motion for these four types of sequence presentation, in three different blocks of trials.

In the first block of trials, subjects rated masked sequences with a short, 30 ms picture duration, identical to the sequences presented in experiments 2 and 3. In two further blocks, the sequences were presented, first at the same picture duration as in block 1 but without masking, then at a longer picture duration, again without masking. The two final blocks were included to ensure that subjects could accurately perceive the object and its motion, since neither was clearly perceived in the first block of trials.

### 5.1 *Method*

There were twelve subjects. All the subjects had previously participated in another sequential identification experiment with the same stimuli. Each subject rated three blocks of trials. In the first block of trials, the sequences were presented with a 30 ms picture duration, and masked. The second block of trials repeated the first, but the mask between each picture was replaced by a blank screen. In the third block of trials, the picture duration was increased to 60 ms, and the pictures were again ummasked.

Each block of trials presented thirty-six objects (those listed in the appendix, plus a racquet and a leek, so that nine objects were shown at each of the four different sequence presentations in each block). Four different types of sequence presentation were shown in the same order in each block. The first trial presented the structured sequence (0° then 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, and 330°). The second trial presented the LS sequence, (0° then 30°, 210°, 180°, 60°, 90°, 270°, 240°, 120°, 150°, 330°, 300°), the third trial presented the pseudorandom sequence (0° then 120°, 270°, 180°, 30°, 240°, 90°, 330°, 210°, 60°, 300°, and 150°), and the fourth trial presented the GC sequence (0° then 60°, 120°, 180°, 240°, 300°, 30°, 90°, 150°, 210°, 270°, 330°). The fifth trial presented the structured sequence again, and so on. The order of presentation of objects in each block was random, and was different for each subject.

The subjects were instructed to rate the degree to which they perceived coherent, apparent motion of the object on each trial. Subjects were asked to use a rating scale of 1 to 5, where 1 indicated completely incoherent, random motion, and 5 indicated very coherent perceived motion. After each trial, the subject typed in a number between 1 and 5 inclusive. Subjects were told that the identification of the object was not required.

### 5.2 *Results*

The mean ratings of coherence of perceived motion in each condition are given in table 3. An ANOVA was carried out with two within-subjects factors, sequence (either structured, GC, LS, or pseudorandom) and block (1, 2, or 3). The main effect of sequence was significant ($F1_{3,33} = 66.476$, MSe = 24.414, $p < 0.001$;

$F2_{3,32} = 34.806$, MSe = 18.310, $p < 0.001$) but not the main effect of block ($F1_{2,22} = 1.156$, MSe = 0.330, $p > 0.3$; $F2_{2,64} = 1.476$, MSe = 0.247, $p > 0.2$). The interaction between sequence and block was significant ($F1_{6,66} = 23.198$, MSe = 2.902, $p < 0.001$; $F2_{6,64} = 13.002$, MSe = 2.177, $p < 0.001$). A posteriori Neuman-Keuls comparisons showed that for the first block of trials (with a masked, 30 ms picture duration) there was no difference in the perception of coherent motion for the structured, GC, and pseudorandom sequences. However, these sequences were rated as having more coherent motion than the LS sequence ($p < 0.01$ for subjects, $p < 0.05$ for items). For the second (unmasked, 30 ms picture duration) and third (unmasked, 60 ms picture duration) blocks of trials, there was no difference between the perceived coherence of motion for the structured and the GC sequences, both of which were rated as having more coherent motion than the pseudorandom sequence ($p < 0.01$). The pseudorandom sequence, in turn, was rated as having more coherent motion than the LS sequence ($p < 0.01$ for subjects, $p < 0.05$ for items).

**Table 3.** Mean ratings of coherence of perceived motion (on a scale from 1, completely incoherent motion, to 5, very coherent motion) in block 1 (masked pictures at a 30 ms picture duration), block 2 (unmasked pictures at a 30 ms picture duration) and block 3 (unmasked pictures at a 60 ms picture duration), for the structured, the globally coherent (GC), the locally similar (LS), and the pseudorandom sequences in experiment 4. Standard deviations are given in parentheses.

| | Structured | GC | LS | Pseudorandom |
|---|---|---|---|---|
| Block 1 | 3.68 (0.66) | 3.38 (0.69) | 2.87 (0.64) | 3.35 (0.50) |
| Block 2 | 4.16 (0.47) | 4.19 (0.58) | 2.23 (0.47) | 2.70 (0.43) |
| Block 3 | 4.24 (0.41) | 4.11 (0.58) | 1.86 (0.50) | 2.49 (0.39) |

### 5.3 *Discussion*

In the first block of trials, which were identical to the comparable trials in experiments 2 and 3, subjects reported only a weak perception of motion. Nevertheless, even for the first block of trials, subjects rated the LS sequence as being less coherent than the other three sequences. This difference increased over the two unmasked blocks of trials, with the structured and GC sequences being rated as much more coherent than both the pseudorandom and the LS sequences. An unexpected result was that the pseudorandom sequence was consistently rated as being more coherent than the LS sequence. One possible explanation for this is that the reversals in the apparent direction of motion (clockwise or anticlockwise) in the LS sequence strongly disrupted the perceived motion of the object. In the pseudorandom sequence, consecutive views were rotated by such large angles that it is unlikely that any obvious direction of motion was perceived. Over all three blocks, there was no difference between the ratings of the coherence of perceived motion for the structured and the GC sequences, indicating that the global structure of the motion of objects in the GC sequence was clearly apparent to subjects.

These rating results dissociate from the identification results of experiments 2 and 3. Objects presented in a structured sequence were easier to identify than objects presented in a random sequence in experiment 2, and objects presented in a LS sequence were easier to identify than objects presented in a GC sequence in experiment 3. However, highly coherent motion is perceived for the structured and the GC sequences, but not for the pseudorandom and the LS sequences. The degree to which coherent motion is perceived is independent of the ease of identification of a picture sequence.

However, estimates of the coherence of perceived motion may not correspond to the robustness with which structure can be extracted from stimulus motion (Ullman 1979). It may be that LS sequences do facilitate the recovery of the structure of an object from its motion, owing to the small spatial separations between corresponding features on consecutive views, whereas for the GC sequences, large separations may preclude the recovery of structure-from-motion information. Further research with random-dot stimuli (for which only motion information is available with which to recover object structure) will be necessary to provide a definitive answer to whether the results reported here are due to the effects of structure-from-motion or the nature of stored object representations. However, we believe that there are a number of reasons why an account in terms of view-specific stored representations is more plausible. First, each view of an object presented was immediately masked, each view and each mask were presented for the same duration, and the depth rotation between consecutive views was large (30°) relative to the depth rotation between successive frames in standard structure-from-motion displays. These conditions would be predicted to disrupt strongly any process which extracted structure from motion, in particular because corresponding features across consecutive views of an object were interleaved by an irrelevant mask, which was, at a feature level, very similar to the views of the object. Second, there was no effect of increasing the length of the sequence of views, indeed a sequence of just two consecutive views (separated by a mask) was sufficient to produce a benefit for LS over GC sequences in experiment 3 that was as great as the benefit for structured over random sequences in experiment 2. Third, the stimuli presented were sparse line drawings, and few features would be available to a process which extracted structure from motion by locating common features across different views. However, these points do not provide conclusive evidence against the involvement of a structure-from-motion process in the current experiments, and this issue is to be addressed in future research.

## 6 Experiment 5

We showed in experiment 3 that the global structure of a sequence of object views does not affect identification when the views differ by 60° depth rotations. What is important is the local similarity between consecutive pairs of views. In experiment 5 we assessed the limits of these similarity effects. For instance, how does object recognition vary when structured sequences of views are presented, relative to when the same view of an object is presented repeatedly? Identical views would benefit from image descriptions remaining the same, for example producing enhanced detection from contour summation. If structured sequences were as good as identical sequences as far as identification is concerned, this would provide evidence for similarity effects based on view-specific representations abstracted away from purely retinal coordinates.

In experiment 5, subjects saw structured sequences, and either identical sequences of the hardest view to recognise, the foreshortened 270° view, or identical sequences of the easiest to recognise, 60° view. Identical sequences with either the hardest or the easiest view were tested, in order to examine the full range of levels of identification efficiency for identical sequences of pictures, since the goodness of a view should influence the ease of recognition (Palmer et al 1981). The hardest and easiest views were selected on the basis of a ratings study, in which twenty independent subjects rated which was the best, most revealing view of the objects used in the experiment. A set of independent subjects were shown the full set of twelve views of each object used in the current experiments, from 0° to 330°. Subjects consistently preferred the 60° and 120° views, which accounted for 48% of preferred choices. The 30° and 150° views were next most popular, whilst only 4% of subjects selected foreshortened views, and 5% chose views at 0° or 180°. The choice of views was also supported by converging experimental evidence from naming and verification experiments (Lawson 1994). The structured sequence contained views with different ratings of goodness, both hard and easy to recognise. On the basis of the presence of such views, identification performance with the structured sequence should fall somewhere between that with hard and that with easy identical sequences, if view-specific but nonretinotopic representations are used to achieve identification.

### 6.1 *Method*

There were twenty-four subjects. All pictures were presented for 30 ms. Each subject saw two of three picture sequences. All subjects saw a structured sequence (60°, 30°, 0°, 330°, 300°, and 270° views); half also saw the 270° identical sequence (the 270° view shown six times) and the other half saw the 60° identical sequence. The practice trials prior to an identical sequence block showed a sequence of six 270° or 60° pictures, as appropriate. There was an extra practice block before the second experimental block, which repeated the third practice block given before the start of the experiment, but with the appropriate sequence presentation (identical or structured).

### 6.2 *Results*

Four subjects were replaced for failing to score between 20% and 80% correct overall (scoring 85%, 85%, 94%, and 97% correct; all four received the 60° identical sequence). Separate ANOVAs were carried out for the subjects presented with the two different identical sequences. There was one within-subjects factor, sequence (either structured or identical).

The percentages of correctly identified objects in each condition are given in table 4. For the twelve subjects receiving the 270° identical sequence, the main effect of sequence was not significant ($F1_{1,11} = 3.906$, MSe = 1.472, $p < 0.08$; $F2_{1,33} = 2.735$, MSe = 7.258, $p > 0.1$), although for subjects there was a tendency for more items to be identified with the structured compared with the 270° identical sequence. For the twelve subjects receiving the 60° identical sequence, the main effect of sequence was significant ($F1_{1,11} = 70.394$, MSe = 8.205, $p < 0.001$; $F2_{1,33} = 24.529$, MSe = 30.189, $p < 0.001$). More items were identified with the 60° identical sequence compared with the structured sequence.

**Table 4.** Mean percentage correct responses (with mean number correct out of seventeen trials, and standard deviations in parentheses) for the structured and identical blocks in experiment 5.

| | Structured | Identical | Difference |
|---|---|---|---|
| 270° identical view | 55.4 (9.42, 1.8) | 43.1 (7.33, 3.8) | +12.3 |
| 60° identical view | 52.0 (8.83, 1.8) | 78.4 (13.33, 1.3) | −26.5 |

### 6.3 *Discussion*

The structured sequence tended to be easier to identify than the sequence of identical 270° views, but it was harder to identify than the sequence of identical 60° views. The structured sequence included the 60° (easy) and the 270° (hard) views used in the identical sequences, together with three relatively good views of the object, and the hard, 0° view. From the rated goodness of each view, the ease of identification of objects presented with the structured sequence should fall somewhere between those for the 270° and the 60° view identical sequences. This was what we observed. The data suggest that the visual similarity important for the structured sequence benefit is coded nonretinotopically. Performance with the structured sequence, containing some good views of objects, tended to be better than that with the identical sequence of poor views, despite the poor views gaining through contour summation.

## 7 General discussion

Experiments 1 to 5 revealed the following:

(i) subjects were better able to identify an object from a structured sequence of briefly presented depth-rotated views than from a randomly ordered sequence of such views (experiments 1 and 2);

(ii) this held, provided that consecutive views were not rotated by as much as 60° (the GC sequences, experiment 3);

(iii) the identification benefit was as large for globally incoherent sequences with locally similar views (LS sequences compared with GC sequences, experiment 3) as for sequences that were both globally coherent and locally similar (structured sequences compared with random sequences, experiment 2);

(iv) these effects were unrelated to perceived motion between consecutive views, since coherent motion was more strongly perceived for the GC sequences than for the LS and pseudorandom sequences (experiment 4);

(v) the identification benefit for structured sequences of rotated views tended to be larger than that for identical sequences of 'poor' (foreshortened) views of objects (though less than that for identical sequences of 'good' views) (experiment 5).

Object identification could benefit for several reasons when a structured sequence of depth-rotated views is presented, relative to when a random sequence of the same views is presented. The present experiments rule out a number of accounts for this benefit. For instance, experiments 3 and 4 showed a double dissociation between rated perceived motion and the structured sequence identification benefit: in experiment 4, coherent motion was relatively easy to perceive for GC sequences and difficult to perceive for LS sequences; in contrast, in experiment 3, identification was substantially easier for LS than for GC sequences. The perception of coherent motion seems to play little part in the structured sequence identification benefit, and indeed seems to depend on a separate system from that mediating object recognition. This is consistent with the neurophysiological distinction between a ventral object recognition system and a dorsal system concerned with spatial localisation and motion perception (Ungerleider and Mishkin 1982; Mishkin and Appenzeller 1987; Schiller et al 1991). We suggest that the object identification benefit for structured sequences reflects activation within the ventral object recognition system. Neuropsychological evidence suggests that motion information may sometimes be used to classify certain types of visual stimuli, even when the processing of static information is impaired [eg for classifying facial expression (Humphreys et al 1993)]. However, the stimuli which can be classified in this way may be relatively limited, and may have characteristic patterns of motion which serve for classification purposes. Motion per se may play little direct role in object identification.

Also, the benefit cannot be due to simple contour-summation processes or to similarity between retinotopic representations. Sequentially depth-rotated (and retinotopically nonequivalent) views tended to be easier to identify than identical but 'poor' views of objects, even though the latter benefited from contour summation (experiment 5).

Perhaps most importantly, the results of experiment 3 showed that local similarity between pairs of consecutive images (with 30° depth rotations between consecutive views) was sufficient to produce a substantial structured sequence advantage. Globally coherent sequences with locally more dissimilar pairings (with 60° depth rotations) were insufficient to elicit the structured sequence benefit. This last result is difficult to account for within any theory that holds that object recognition depends on the matching of view-invariant image descriptions to object-centred stored representations (eg Marr 1982). All the images in the GC sequences in experiment 3 were identifiable, and ought to map onto the same stored representation. The GC sequence was also globally coherent and familiar. Identification should have been facilitated or equal to LS and random sequence identification, owing to activation of the same stored representation. It might be argued that the priming here does not reflect stored object representations, but rather the similarity of consecutive view-specific representations, derived prior to the generation of an on-line representation that is object-centred and view-independent. However, the present effects were derived under conditions in which consecutive stimuli were interpolated with pattern masks, which ought to disrupt any early on-line representations of objects (eg see Ellis et al 1989). Priming based on the activation of stored representations seems more plausible under these conditions.

Relatively simple objects were presented in the current experiments, depicted from an elevated angle, so that views rotated by 60° in depth usually revealed the same main components. Facilitated or equal GC identification relative to LS sequences should then have occurred if stored object representations were componential (eg Biederman 1987), since Biederman predicts viewpoint-invariant recognition, providing that parts are not accreted or deleted, or spatial relations between parts changed, as an object rotates in depth.

However, for some objects there was accretion and deletion of parts, even for pairs of views rotated by as little as 60° in depth. A rating study was therefore conducted to test the prediction that changes in the visibility of parts across different views of an object is a major factor determining whether or not a structured sequence benefit is observed (eg Biederman and Gerhardstein 1993). Ten independent subjects were shown the full set of twelve views of each of the thirty-four objects used in the current experiments, and were asked to divide the objects equally into two sets of seventeen objects. In the same parts set, the same parts of the object were visible across all views, whereas in the changing parts set, there was accretion and deletion of parts across the different views of the object, ie some parts were visible for only a subset of the twelve views of an object. On the initial sort, all subjects placed more objects in the same parts set, suggesting that for most objects there was little variation in the visibility of parts across the different views. The thirty-four items were divided equally into same parts and changing parts sets, on the basis of the ratings of the ten subjects. The mean number of subjects selecting a same parts item as belonging to that set was 8.59 (SD = 1.46), and for changing parts items, the mean number was 8.59 (SD = 1.62), indicating a high degree of consistency in subjects' choices. The items analyses for experiments 1, 2, and 3 were then repeated, included parts (same parts or changing parts) as a between-items factor. In no analysis was there a significant main effect of parts, or a significant interaction involving parts. In no case was there a trend towards a larger structured sequence benefit for the changing parts items, as Biederman and colleagues would have predicted (see table 5).

The data therefore suggest that the activation of highly view-specific representations is necessary to generate the structured sequence benefit. The benefit is eliminated if consecutive views are rotated by more than 30° in depth. It is not changes in the visibility of parts, but metric changes in the position of features in the image as objects are rotated, that minimise the combined activation of common representations and so eliminate the structured sequence benefit.

In conclusion, the results presented here suggest that object recognition is most efficient if the same stored view-specific object representation is activated by two successive pictures. Picture priming is maximised when two pictures reveal very similar views of the object. This occurs even under conditions where consecutive views are interpolated with pattern masks.

These results are difficult to explain by theorists such as Marr (1982), who assume that a single stored object-centred representation is accessed whatever the view of the object presented. In particular, if a single view-invariant representation is the primary representation for object recognition, then no difference would be predicted in the recognition efficiency of structured and random sequences, since the same static information is available in both cases, and only the temporal order of presentation differs.

**Table 5.** Mean percentage correct responses for the structured/LS and random/GC blocks for the seventeen items rated as revealing the same parts across all views and the seventeen items rated as having changing parts (LS locally similar, GC globally coherent).

| | Structured/LS | | Random/GC | | Difference | |
|---|---|---|---|---|---|---|
| | changing | same | changing | same | changing | same |
| 45 ms/picture (experiment 1) | 63.2 | 59.8 | 54.9 | 46.6 | +8.3 | +13.2 |
| 30 ms/picture (experiment 1) | 42.2 | 46.6 | 28.4 | 29.4 | +13.8 | +17.2 |
| 30 ms/picture (experiment 2) | 67.7 | 58.3 | 44.6 | 39.2 | +23.1 | +19.1 |
| LS/GC (experiment 3) | 68.6 | 68.6 | 45.1 | 38.2 | +23.5 | +30.4 |

The results suggest instead that views of objects rotated 30° in depth from each other will often activate a common view-specific representation. However, there is a limit to this tolerance to differences in viewpoint; the sequential identification experiments indicate that views rotated just 60° in depth from each other are no more likely to activate the same view specific representation than random pairs of views rotated by at least 90° in depth. More research is needed to clarify why such highly view-specific effects emerge for the short-term priming tasks presented here, in contrast to the results of Biederman and Gerhardstein (1993), which suggested a greater degree of view-invariance in object recognition for a longer-term priming task. In addition, in the current experiments, visual similarity was assumed to covary with the physical measurement of difference in viewpoint of two pictures of an object. Further research will be necessary to elucidate the factors actually determining visual similarity, by considering whether metric changes in features are important. It is hoped that a more principled approach to expressing visual similarity can be adopted, for instance, the use of aspect graphs (Koenderink 1990; but see also Biederman and Gerhardstein 1993). Finally, experiments should also contrast effects with familiar and with novel objects, to allow discrimination between priming at the level of the temporary structural description of the image and priming of the stored object representation, rather than rely on indirect argument concerning whether effects are sustained under masking conditions. Irrespective of this last point, the present results demonstrate that a highly view-specific representation of an object survives masking, and primes the recognition of the same object shown in a similar view, independent of motion perception between the stimuli.

**Acknowledgements.** This work was supported by a grant from the Science and Engineering Research Council of Great Britain (Grant No 9031212X) to Rebecca Lawson. We are grateful to Mike Harris, Toby Lloyd-Jones, and Alexander Verstaen for comments on an earlier version of this paper. We thank Michael Tarr, Johan Wagemans, and an anonymous reviewer for their comments.

## References

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

Biederman I, Gerhardstein P C, 1993 "Recognizing depth-rotated objects: Evidence for 3-D viewpoint invariance" *Journal of Experimental Psychology: Human Perception and Performance* **19** 1162-1182

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

Edelman S, Weinshall D, 1991 "A self-organizing multiple-view representation of 3D objects" *Biological Cybernetics* **64** 209-219

Ellis R, Allport D A, Humphreys G W, Collis J, 1989 "Varieties of object constancy" *Quarterly Journal of Experimental Psychology A* **41** 775-796

Humphrey G K, Khan S C, 1992 "Recognizing novel views of three-dimensional objects" *Canadian Journal of Psychology* **46** 170-190

Humphreys G W, Donnelly N, Riddoch M J, 1993 "Expression is computed separately from facial identity, and it is computed separately for moving and static faces: neuropsychological evidence" *Neuropsychologia* **31** 173-181

Humphreys G W, Riddoch M J, 1984 "Routes to object constancy: Implications from neurological impairments of object constancy" *Quarterly Journal of Experimental Psychology A* **36** 385-415

Jolicoeur P, 1985 "The time to name disoriented natural objects" *Memory and Cognition* **13** 289-303

Koenderink J J, 1990 *Solid Shape* (Boston, MA: MIT Press)

Lawson R, 1994 *The Effects of Viewpoint on Object Recognition* unpublished PhD thesis, Birmingham University, Birmingham, UK

Lawson R, Humphreys G W, 1994 "View-specificity in object processing: Evidence from picture matching" submitted for publication

Marr D, 1982 *Vision* (San Francisco, CA: W H 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

Mishkin M, Appenzeller T, 1987 "The anatomy of memory" *Scientific American* **256**(6) 62-71

Palmer S, Rosch E, Chase P, 1981 "Canonical perspective and the perception of objects", in *Attention and Performance* Volume IX, Eds J Long, A Baddeley (Hillsdale, NJ: Lawrence Erlbaum Associates)

Perrett D I, Hietanen J K, Oram M W, Benson P J, 1992 "The organization and function of cells responsive to faces in the temporal cortex" *Philosophical Transactions of the Royal Society of London, Series B* **335** 23-30

Perrett D I, Oram M W, Hietanen J K, Benson P J, 1994 "Issues of representation in object vision", in *The Neuropsychology of Higher Vision: Collated Tutorial Essays* Eds M J Farah, G Ratcliff (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 33-62

Schiller P H, Logothetis N K, Charles E R, 1991 "Parallel pathways in the visual system: their role in perception at isoluminance" *Neuropsychologia* **29** 433-441

Seibert M, Waxman A M, 1991 "Learning aspect graph representations from view sequences", in *Advances in Neural Network Information Processing Systems* Volume 2, Ed. D S Touretzky (San Mateo, CA: Morgan Kaufman) pp 258-265

Tarr M J, Pinker S, 1989 "Mental rotation and orientation dependence in shape recognition" *Cognitive Psychology* **21** 233-283

Ullman S, 1979 *The Interpretation of Visual Motion* (Cambridge, MA: MIT Press)

Ungerleider L G, Mishkin M, 1982 "Two cortical visual systems", in *Analysis of Visual Behaviour* Eds D J Ingle, M A Goodale, R J W Mansfield (Cambridge, MA: MIT Press) pp 549-586

Warrington E K, Taylor A M, 1973 "The contribution of the right parietal lobe to object recognition" *Cortex* **9** 152-164

Warrington E K, Taylor A M, 1978 "Two categorical stages of object recognition" *Perception* **7** 695-705

## APPENDIX

**The thirty-four items presented in the experiments**

Banana
Bone
Camel
Can opener
Car
Clothes peg
Coathanger
Comb
Corkscrew
Fork
Glasses
Hammer
Iron
Kangaroo
Key
Knife
Loaf
Paper clip
Pen
Pencil
Razor
Ruler
Saw
Scissors
Screwdriver
Shoe
Spanner
Spoon
Stapler
Telephone
Toothbrush
Torch
Train
Whisk