Perceptual recognition of familiar objects in different orientations

Newell, Fiona N.

Perceptual Recognition of Familiar Objects in Different Orientations

Fiona N. Newell B.A.

Thesis submitted to the University of Durham,
Department of Psychology
in candidature for the degree of
Doctor of Philosophy,
December,
1992

Abstract

Recent approaches to object recognition have suggested that representations are view-dependent and not object-centred as was previously asserted by Marr (Marr and Nishihara, 1978). The exact nature of these view-centred representations however does not concord across the different theories. Palmer suggested that a single canonical view represents an object in memory (Palmer et al., 1981) whereas other studies have shown that each object may have more than one view-point representation (Tarr and Pinker 1989).

A set of experiments were run to determine the nature of the visual representation of rigid, familiar objects in memory that were presented foveally and in peripheral vision. In the initial set of experiments recognition times were measured to a selection of common, elongated objects rotated in increments of 30° degrees in the 3 different axes and their combinations. Significant main effects of orientation were found in all experiments. This effect was attributed to the delay in recognising objects when foreshortened. Objects with strong gravitational uprights yielded the same orientation effects as objects without gravitational uprights. Recognition times to objects rotated around the picture plane were found to be independent of orientation. The results were not dependent on practice with the objects. There was no benefit found for shaded objects over silhouetted objects. The findings were highly consistent across the experiments.

Four experiments were also carried out which tested the detectability of objects presented foveally among a set of similar objects. The subjects viewed an object picture (target) surrounded by eight search pictures arranged in a circular array. The task was to locate the picture-match of the target object (which was sometimes absent) as fast as possible. All of the objects had prominent elongated axes and were viewed perpendicular to this axis. When the object was present in the search array, it could appear in one of five orientations: in its original orientation, rotated in the picture plane by 30° or 60° , or rotated by 30° or 60° in depth. Highly consistent results were found across the four experiments. It was found that objects rotated in depth by 60° took longer to find and were less likely to be found in the first saccade than all other orientations. These findings were independent of the type of display (i.e. randomly rotated distractors or aligned distractors) and also of the task (matching to a picture or a name of an object). It was concluded that there was no evidence that an abstract 3–dimensional representation was used in searching for an object.

The results from these experiments are compatible with the notion of multiple-view representations of objects in memory. There was no evidence found that objects were stored as single, object-centred representations. It was found that representations are initially based on the familiar views of the objects but with practice on other views, those views which hold the maximum information about the object are stored. Novel views of objects are transformed to match these stored views and different candidates for the transformation process are discussed.

Statement of Copyright

Declaration

The work contained in this thesis was carried out by the author between January, 1990 and December, 1992 at the University of Durham. I declare that the work contained in this thesis is my own and that no part has been previously submitted in candidature for any other degree.

In Press

Part of this thesis has been published;

Newell, F.N. and Findlay, J.M. (In Press); *Viewpoint Invariance in Object Recognition*. Irish Journal of Psychology.

Findlay, J.M., Newell, F.N. and Scott, D.W. (1992); *How Much of the Visual Periphery is Monitored during Visual Search?* Proceedings of 3rd International Conference on Visual Search. Nottingham, 1992.

To all My Family especially My Parents,
with Love and Thanks for all Your Support.
Go raibh míle maith agaibh.

Acknowledgements

I am extremely grateful my supervisor, John Findlay, for guiding me throughout the course of this research, for his enthusiasm, encouragement and his seemingly unlimited source of knowledge on a vast number of issues in visual perception. Thanks also go to Bob Kentridge for toiling with the Macintosh to produce a program for stimulus presentation and data recording. Particular thanks go to all those people who volunteered as subjects in the eye movement experiments especially Ala, Andy, Bob, John and Robin. Thanks also to all of the subjects, who are too numerous to mention, who participated in the reaction time experiments. I am also grateful to the technicians for maintaining the laboratory equipment. Finally, I thank Roland for helping me through the final stages of writing up this thesis.

Fiona N. Newell

I would like to acknowledge the financial support provided by the Science and Engineering Research Council over the past three years.

Table of Contents

List of Figures

Chapter One

Approaches to Object Recognition

In the natural environment an object's visual characteristics can change from scene to scene and from moment to moment. Different changes in the light throughout the day change the luminosity and shading of the object's surface. Its retinal size changes with position in the visual array and viewing distance. The orientation of the object may change relative to the environment and its visual orientation changes with the viewing position of the observer. The object may also be in the presence of a number of other objects that share similar characteristics and locating the correct object means ignoring all of the irrelevant objects in a scene. Finally, the object may be occluded by other objects. Despite this variability people generally seem to have very little difficulty in detecting and identifying familiar objects.

Theories of object recognition deal with how the visual system uses information available from the retinal image to access a representation of that object in memory in order to associate the relevant semantic information with it, such as its name for example. However, the goal of representation is not to reconstruct an image in memory as such but to represent enough information to allow an animal or person to interact with their environment. The two most important properties of an object that an animal needs to know are the spatial locations of the object and the specific characteristics of the objects itself. These have often been referred to as the "what" and "where" visual systems (Ungerleider and Mishkin, 1982). These properties therefore need to be recognised and represented by the animal in such a way that the animal can interact with the surrounding world. The recognition process would seem to be more complicated for man in that human recognition permits verbal labelling. An intriguing problem for visual scientists to solve is how humans can recognise an object across a variety of different conditions such as changes in the ambient light or changes in orientation. If the goal of visual perception is to create a representation of an object that includes enough information about the object in order that the animal or person can interact with the object, then how is this goal achieved?

This thesis investigates the processes involved in the recognition of objects across different orientations using an experimental approach. The aim of the thesis is threefold; to identify the information accessed from the 2-dimensional, retinal image for recognition purposes, to account for the process involved in matching the inputted image to a stored representation and to provide a description of the stored representations in visual memory. The effects of different views on recognition is studied for both foveally presented stimuli and stimuli presented in peripheral vision. Two experimental approaches are used in this investigation; reaction time studies and visual search studies. The thesis is structured around

these two experimental approaches. The remainder of this chapter introduces the different approaches to object recognition. Chapter 2 introduces the experimental evidence for the different theoretical approaches to object recognition. Chapter 3 includes an experimental investigation into the effects of different views of objects on recognition times. An analysis into the nature of the information accessed from the retinal image for recognition is discussed in Chapter 4. Chapter 5 includes an investigation into the effects of different views of unfamiliar objects on recognition. In the second part of this experimental investigation of object recognition, a visual search paradigm is used. Chapter 6 outlines the various contributions to object recognition from visual search studies. Chapter 7 includes a test of the effect of different orientations of objects presented in peripheral vision on detection times. In conclusion, Chapter 8 attempts to relate the findings from the experimental work included in the thesis to the models outlined in Chapter 1.

The present chapter begins with an introduction to the theoretical contributions of the experimental and computational approaches to object recognition. This section looks at the various models of recognition in detail. In the next section a discussion of the contributions from animal studies to object recognition and of the underlying neural substrates in object recognition is included. The contributions from single-unit recording studies to the nature of visual memory is discussed in the following section. Finally, the evidence from neuropsychological studies on the effects of brain injury on the ability to recognise shapes and objects is outlined. Neuropsychological studies have proven to be important in highlighting how visual information is processed and stored in the visual cortex and how selective damage along the visual pathways can disrupt the recognition processes in different ways.

1.1 Experimental and Computational Approaches

In general, computational and experimental approaches to object recognition can be divided into two broad classes; object-centred and view-centred approaches. Theorists that support the object-centred approach to recognition claim that the recognition of objects is orientation invariant (Marr, 1982; Biederman, 1987). On the other hand, theorists that support the view-centred approach claim that the recognition of objects is dependent on the view of the object being observed and that some views are more recognisable than others (Jolicoeur, 1992; Ullman, 1989).

This section is introduced with an outline of the early approaches to object recognition, followed by a section on the object-centred approaches and the view-centred approaches. The section on the view-centred approaches is divided according to the nature of the transformation process hypothesised that matches the image to a stored view. Chapter 2 discusses the experimental evidence for these different approaches.

1.1.1 Traditional Theories of Object Recognition

Traditionally the problem of shape recognition has been dealt with in three different ways: template models, feature models and structural descriptions.

1.1.1.1 Template Models

Template models propose that an image-like representation of an object is stored in memory. If a separate representation is stored for each view, an inputted view might be matched with these stored views until the correct match indicates the identity of the object. Such a model has obvious limitations in that there would be large demands on the storage capacity of the brain if a topographic replica of the retina were stored for each possible retinal image of an object. A possible improvement is to 'normalise' the retinal stimulation by a process of size scaling and rotation so that only one standard template is used for each object. This early template model did not adequately account for the third dimension but subsequent improvements have involved 3-D mental rotation (see Pinker 1984).

1.1.1.2 Feature Models

Feature models such as the Selfridge 'Pandemonium' model (Selfridge, 1959) propose that the visual system responds to the presence of features in the retinal image. These features are integrated and the object is identified by matching the activity of the features against the weights associated with the object's features as represented in memory. It was hoped that features might be found which were invariant over different views etc., but this hope has not been fulfilled (Ullman, 1989). Also features have never been well defined. One way in which the term is used is to describe the obvious components or 'geons' into which some objects can be partitioned (Marr, 1982 and Biederman, 1987). Warrington and James (1986) proposed that the features represented in memory corresponded to clusters of visual contours which were unique for each object. Another definition of features describes the properties of the object such as its colour for example. Treisman et al. (1990) postulated that the visual system extracts

> "....features that specify textures, surfaces, and their spatial layout, features that specify events (movement, change), and finally features that define the shapes and structures of objects," (Treisman et al. 1990).

Feature models in general, do not explicitly propose how the spatial relations between features are accounted for. Take, for example, a mug and a bucket. A mug's features may consist of a handle and a conical shape truncated at the bottom and open at the top. The handle is placed on the side of the mug. However, if the handle was placed on the top of the cylinder, the identity of the object has changed and it now resembles a bucket more than a mug (see Figure 1). This illustrates the importance of including the spatial relations between the features of an object if the theory is built around partitioning the object.

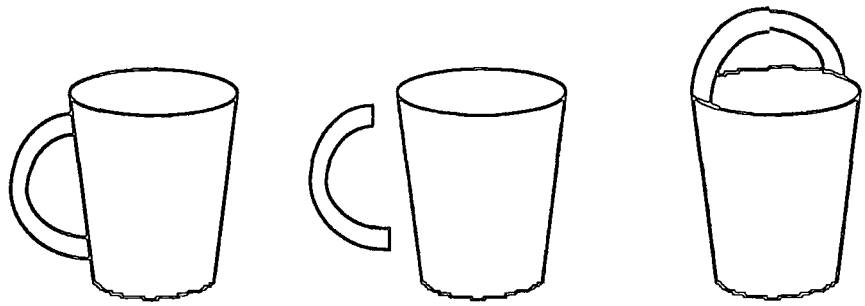

Figure 1: Illustration of how different spatial arrangements of the same parts or features can create different object descriptions. Placing the 'handle' shape to the left of a truncated cone produces a mug whereas the same handle placed on top of the truncated cone produces a bucket (after Biederman, 1987).

1.1.1.3 Structural Descriptions

Structural descriptions consist of a set of abstract or propositional descriptions about an object. This notion has been derived from theories of imagery, particularly the view proposed by Pylyshyn (1973) that images are stored as propositions. Propositions are abstract, language-like representations of objects and are therefore in stark contrast to the picture-like representations of the template or feature models described above. In early work these descriptions have been used to describe the configurations between the structures of alphanumeric symbols. For example a description of a 'T' may indicate that it has two parts, one horizontal and the other vertical. The vertical line supports and bisects the horizontal line. The relative lengths of the lines are not important and are therefore not specified. This notion is an appealing one and indeed it has been incorporated into some of the more recent theories on object recognition (see discussion below on Marr and Nishihara, 1978 and Biederman, 1987).

Structural descriptions however are not considered to explain adequately the nature of representations in memory. They can not account for the findings on the effect of rotations of objects on reaction times. For example Shepard and Metzler (1971) found that the time taken to compare two rotated stimuli was proportional to the difference between their angles of rotation. According to Kosslyn (1980), structural descriptions would not predict a linear function in this case. He argues that a difference of 180° would mean changing the description of 'top' to 'bottom' and therefore yield short reaction times. In fact rotations of 180° resulted in the longest reaction times. In sum, structural descriptions were not considered adequate on their own but may be useful if incorporated into a more general account of object representation.

Recent theoretical proposals on how the visual system recognises an object in different orientations, despite different visual information from the different views, fall into two broad classes; object-centred or invariant properties models and view-centred models. These two classes differ fundamentally in that the object-centred models assume that an object's representation in memory describes the object in coordinates relative to an intrinsic

reference point and recognition is therefore invariant over different view points. The view-centred models on the other hand propose that the objects representation describes it relative to the viewer and that a transformation process is involved to align the input with the stored representation of the object in memory. Theorists of the view-centred class differ as to whether there is one single representation in memory for each object or whether there are multiple representations. Another contentious issue is the nature of the alignment process. Possible transformations are discussed below.

1.1.2 Object-Centred Approaches

In terms of the object-centred approach, efforts have concentrated on trying to extract the invariant properties of an object in order that a single, object-centred representation is stored (Hinton and Parsons, 1981; Marr, 1982; Marr and Nishihara, 1978; Jolicoeur and Kosslyn, 1983). Such models propose that properties or features of objects are extracted that are invariant over a number of transformations of the objects such as orientation for example. An example of features that are invariant over orientations are non-accidental properties of the edges of an image such as parallel edges, symmetry and co-termination. If these properties are present in the image then they are assumed to reflect the objects true properties in the real world (Marr and Hildreth, 1980 and Biederman, 1987).

Many computational models that have been proposed seek to draw insights to the workings of the human visual system. One of the most influential computational approaches in vision is that of David Marr (1982). Marr proposed that the visual system was organised in a modular fashion with information processed sequentially from simple edge detection to more abstract representations of whole objects and their properties (Marr and Hildreth, 1980; Marr and Nishihara, 1978). According to Marr, information about an object proceeds along the visual system from the initial information in the raw primal sketch, to an intermediate, view-dependent description called the $2^1/_2$-D sketch until it is ultimately represented as a 3–D model which is invariant over view point. This notion of an object–centred description as a representation set a precedent for subsequent work on object recognition which carried on well into the 1980s.

Marr and Nishihara (1978) postulated that the object description in memory holds information which is invariant over various transformations such as rotation or displacement. Marr (1982) produced an outline for a computational model of the whole visual process from retinal image to object representation. A retinotopic representation is initially generated from the properties of the image on the retina. The edges of the image are extracted by looking for the zero-crossings in the second derivative of the patterns of light intensity in the image. From this information the raw primal sketch is generated (Marr and Hildreth, 1980). This primal sketch contains information about the changes in light intensity across the image. The next stage of the visual process is to construct a viewer-centred representation of the image. Marr referred to this representation as the $2^1/_2$-D sketch and considered its

construction to be the goal of early visual processing. The $2^1/_2$-D sketch holds a description of the surfaces of the image with respect to the viewer. All of the information available about the image merges to form this sketch e.g. information about texture, shading and motion. However, representations based on a retinotopic frame are early representations of objects and further processing is required in order to represent the object in terms of coordinates that are independent of viewpoint and are intrinsic to the object itself.

Marr and Nishiharas (1978) idea of representation rests on the notion that a *single*, object-centred description is stored to enable the recognition of the object in its many views. They argued that this object–centred description is built up from information about the principal axis of the object (for example the axis of elongation or symmetry). This view–independent description is in the form of a structural description and is built through spatial arrangements between the object and its parts. For more complex objects the description is hierarchical in its format going from the global properties of the object to information about the component axes of the principal axes. The model therefore adequately accounts for the spatial relations between the parts of an object by including hierarchical descriptions of the object. This work was soon recognised as of great potential. Models based on three dimensional structural descriptions could explain how the visual system recognises objects despite being occluded, rotated in depth or degraded. It has also been argued that 3–dimensional models are constructed and used in problem solving tasks and that results from such studies support Marr's ideas that objects are represented as 3-D models (Shepard and Metzler, 1971 and Cooper, 1990).

The main problem for an object-centred model is to account for how the same object-centred description is derived from different views of the one object. Marr (1982) proposed that if the principal axes that are intrinsic to the object are derived from the image then the same object centred description can be created. However his explanation of this process is vague. It has also been argued that the automatic learning of each 3-D model is problematic (see Poggio and Edelman, 1990).

Biederman tried to account for the process which derives the same object-centred description from different views by arguing that, for visual purposes, all 3-D objects are made up of a set of 3-D component parts which are invariant over different views (Biederman 1987; Biederman and Gerhardstein, 1992). He argues that there are a limited number of primitive components or "geons" (geometric ions) that make up most of the objects that we know. The total number of geons is less than 50 and more than 10. Volumetric primitives include cylinders, cones, wedges and blocks and these components can be derived from the properties of the 2–dimensional image. Figure 2 below illustrates a number of geons that make up different 3–dimensional objects due to the different structural relations between them. Representations are therefore extracted from information about edges in the visual scene. Other information may contribute to the representation such as diagnostic features (e.g. colour) but is not as advantageous in creating a representation as edge information (Biederman and Ju, 1988).

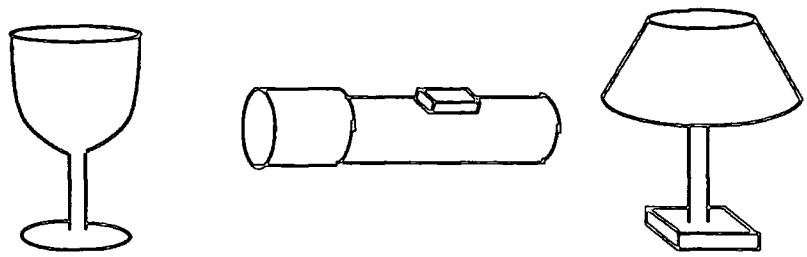

Figure 2: An illustration of three objects made up of three different components or 'geons'.

The significant properties available in edge-based descriptions are termed non-accidental properties in that if they are present in the image then they must reflect the properties of the object in the real world. For example, if parallel lines are present in the image then, almost inevitably, they must be a property of the edges of the object in the world. These properties are generally invariant over viewpoint thereby giving access to the stored representation in memory. He terms this his Recognition-By-Components (RBC) theory. Objects are partitioned at regions of deep concavity (see Hoffman and Richards, 1984) and these parts are matched against the stored objects in memory. Recognition comes about when the geons are arranged to match the stored representations in visual memory, therefore, objects are recognised as spatial arrangements between these component parts.

There are some constraints however as to how the visual system can achieve these viewpoint invariant descriptions. Firstly, Biederman argues that the objects must be capable of being readily decomposed into different geons. The invariant descriptions break down if the objects cannot be partitioned or if the parts are highly irregular, corresponding to texture regions rather than volumetric primitives. He gives crumpled paper or irregular lumps of clay as examples of objects which cannot be readily partitioned into geons and therefore do not have viewpoint-invariant descriptions. Secondly, each geon-based object representation in memory must be sufficiently unique in order for viewpoint invariance to be achieved. His theory assumes that a view-point invariant structural description can be created from a single view of any 3-dimensional object. Recent developments of RBC have argued that representations of objects as geons are not only invariant over orientation in depth but also with respect to retinal size and position (Biederman and Cooper, 1992).

Hummel and Biederman (1992) generated a neural network model of 3-D object recognition based on Biedermans Recognition-By-Components theory. Briefly his theory states that all objects are parsed into a limited set of geons and objects are represented by the structural descriptions or spatial relations between these parts. The Hummel and Biederman model successfully recognises 3-D objects made up of different geons by declaring the spatial relations between the geons. In order that the appropriate edge or vertex associates with other such features of the same geon the model needed to solve the 'binding' problem. The

question of how independently coded features are bound together into integrated descriptions of objects presents a major problem for current vision research. It is important to solve this problem for any model of object recognition that uses structural descriptions to represent the objects. Figure 3 below illustrates this problem. To state it more simply; how does the system know which features conjoin together to represent the whole of the part or object? Hummel and Biederman (1992) provide a solution to this problem by synchronising the out-puts that respond to features of the same geon. These outputs are thus phase-locked (see Engel et al., 1992). However, only a limited number of outputs can be synchronised before the system becomes confused. His model therefore can only recognise objects made up of a maximum of 2 or three geons. It could not, for example, recognise multi-component objects such as hands. It has also been demonstrated that humans often make errors when asked to integrate features of objects shown in rapid succession and even create illusory conjunctions between features shown in a single display (see Treisman 1986). The evidence that temporal binding is used by the visual system to conjoin features is therefore equivocal.

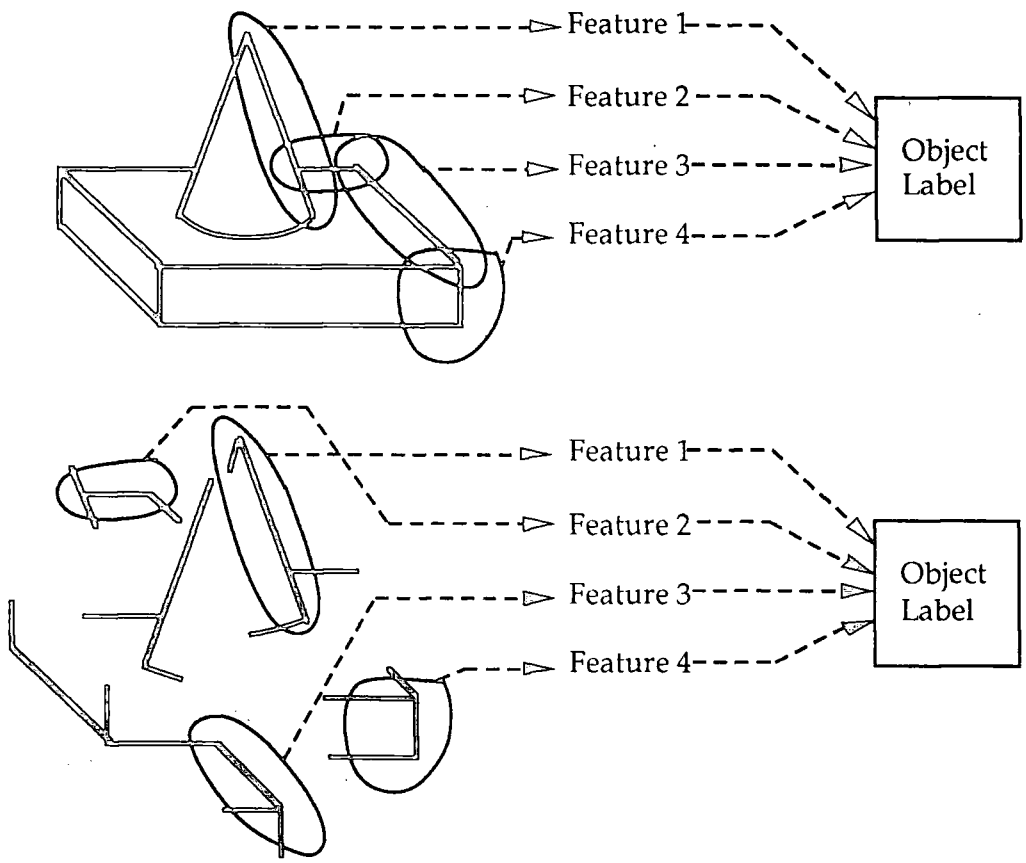

Figure 3: Illustration of the 'binding' problem. Without an explicit spatial representation of the parts of objects, illusory recognition of the same object will occur to both images because they both contain the same vertex complexes (after Hummel and Biederman, 1992).

The notion that objects are partitioned is supported by work on the conceptual categorisation of objects (Rosch, 1973 and 1977). Tversky and Hemenway (1984) showed that members of the same basic level category share the same parts and this consistency is not found between members of sub- or super-ordinate categories. For example, the same parts of a table or a chair were consistently found across subjects but this consistency was not evident for the super-ordinate category such as 'furniture' or the sub-ordinate category such as card table or kitchen chair. It is not clear, however, whether this correlation between conceptual categories and the number of parts in common between its members reflects the processes of a concept forming system or whether it is simply a coincidence in the real world. In other words, it is not conclusive whether the visual and concept forming system exploits this information in order to make interpretations about objects in a visual scene. Murphy (1991) argues that parts are neither necessary nor sufficient for establishing basic-level categories. He ran a set of experiments in which he tested the nature of the information used in order to create basic level categories. He found that parts are not necessary aspects of basic category structure contradicting Biederman's assertions on the importance of parts.

Both the Marr and Nishihara (1978) and the Biederman (1987) models are the most influential contemporary models of 3-dimensional shape recognition and they are not afflicted by the problems associated with the earlier models of shape recognition outlined above. Nevertheless, the object-centred approach does have a number of problems (some of which have already been discussed in the text above) which need to be addressed. For example, both of these models propose that stored object-models are based on a spatial arrangement of a number of shape primitives such as generalised cones (see Marr, 1982) or geons (Biederman, 1987). Although most classes of objects can be parsed into primitive components, there are a number of other classes which cannot be easily described as a collection of 3-dimensional components such as faces, shoes and trees (see Hoffman and Richards, 1984). A further problem for Marr's object-centred model is that there is no general procedure for deriving the object-centred model from the $2^1/_2$-D sketch. Although Biederman did outline such a procedure (Hummel and Biederman, 1992), it has limitations and the procedure could not apply to objects made up of 3 or more components. More recent approaches have assumed that representations are not object-centred but view-centred and that recognition proceeds by transforming the image to match a stored view. The viewer-centred approaches are discussed below.

1.1.3 View-Centred Approaches

More recent object recognition models have proposed that representations are in fact view-centred and the inputted images are matched to a view of the object in memory that corresponds to the shape of the image. Most view-centred models are based on the early template models discussed above. However, current view-centred models incorporate more elaborate transformation processes in order that the problem of matching across disoriented shapes (in any dimension) is alleviated. Such models have proposed that images are either

directly matched to a stored view or that some sort of transformation process aligns the image with the stored view. The models differ on the nature of this transformation process. View-centred models have proposed that novel views are aligned either by a simple transformation of the image, or by aligning a limited number of features of the image or by interpolating between the stored views of objects. These transformation procedures are discussed below.

1.1.3.1 Simple Transformation Models

In one of the most recent theoretical approaches to recognition, Jolicoeur (1992) proposed that objects are stored in memory as orientation-specific representations that are referenced to the retinal upright. He argued that one of two systems can be used to identify disoriented objects; a mental-rotation system and a feature-based system, both of which work in parallel. For the purposes of this section only the mental rotation model will be discussed here. The feature-based system is discussed in detail later on in the thesis.

The mental-rotation system postulated by Jolicoeur is analogous to the mental rotation effects found in investigations of pattern matching across rotated images (see Shepard and Metzler, 1971). This mental-rotation system is able to rotate an image along the shortest path of rotation until the image is aligned with the upright. Although Jolicoeur does not make the processes involved in this computation explicit he does suggest that the process could be supported by general heuristics such as aligning the longest axis with either the vertical or the horizontal because most elongated objects are either vertical or horizontal due to the constraints of gravity. Another heuristic is that most animals have their feet closer to the ground than to the sky. Thus mental rotation would align an inverted image of an animal to the upright by rotating the shape by 180°.

Tarr and Pinker (1989) also proposed that images are mentally rotated to match a stored view. However they proposed that objects are represented as a collection of stored views rather than a single stored view. According to Tarr and Pinker, images are mentally rotated in an analogue fashion along the shortest path such that the image is aligned to the nearest stored view of the object.

There is, however, a drawback to the simple transformation type of model. It is difficult to imagine how the visual system decides what is the correct object representation to align the input to and to determine the correct transformation of the input. In other words, it assumes some level of recognition before the correct transformation is applied. Corballis (1988) pointed out that "it is hard to understand how one could mentally rotate an unrecognised shape to a canonical or upright orientation, because in the absence of recognition one could hardly know what its canonical orientation was". For example, according to the heuristics proposed by Jolicoeur (1992) an image of an animal is mentally rotated by 180° if the feet are in the opposite direction to the ground. However, there is some level of recognition involved in order that the feet are identified. Jolicoeur (1992) argues that top-down processing can affect recognition but that the image needs to be rotated in order that the

specific example of the animal can be identified. Tarr and Pinker suggest that certain characteristics of the object can be extracted in order that the correct representation is chosen to match against the inputted image. These 'characteristics' were not made explicit in their model. In order for the model to be a comprehensive one on the workings of the visual system then all of the processes involved in recognition need to be explained.

It could be suggested that the salient characteristics of the orientated objects may be extracted from the 2-D retinal image and that a process of mental rotation is applied which ultimately leads to a direct matching of the transformed input with a stored representation. These characteristics may for example include global characteristics of the object such as its principal axis or more local salient features such as the trunk of an elephant. A transformation process may be applied to the extracted characteristics in order that they may be aligned with a standard orientation in the visual system. Once oriented to a standard then a matching process between the input and all potential stored representations may proceed. The next problem for the visual system is to choose the correct corresponding representation in memory in order that recognition can occur. This is a difficult process to account for. The representations need to be sufficiently unique in order that two shapes are not confused with each other. However, the representations should not be so unique that novel examples of an object are not recognised as different examples of the same object.

1.1.3.2 Alignment Models

Ullman (1989) recognised this problem and proposed that the representations of the objects in memory are based on pictorial descriptions rather than more abstract descriptions such as structural descriptions previously used by Marr and Biederman. He does however suggest that abstract descriptions can be used in representations but that they are used in a pictorial manner. He gives as an example of an abstract pictorial description a verbal description of a "wiggly" line on the top of a chicken's head. In a pure alignment method every instance of an image of a chicken would be aligned to match the internal representation of a chicken. The exact shape of the top of a chicken's head is not important for the recognition of a chicken therefore if the top is described in some sort of abstract description which denotes its wiggliness, then the need to transform the crown to match the internal description is reduced. In other words, a general description of the crown as a wiggly bit on the top of a chickens head would suffice as a representation of that feature. Other more important features of objects would be represented internally as pictorial descriptions. This alleviates the constraint that the shape description of the input needs to directly match a representation for recognition to occur.

In another similar approach, Lowe (1985) proposed an alignment scheme based on matching a perspective view of an object to a 3-dimensional model. Lowe argued that spatial information is "the dominant source of information for verification in most tasks" and therefore images are matched to their stored inputs on the basis of the spatial information alone. Other information may be redundant for the purpose of recognition or may be

occasionally used. For example, colour may be important for the recognition of fruit and other vegetation. He developed a computer program that uses inferences from the contours of the 2–D image to derive 3-D relations from the features of objects. Properties of the 2-D image can include: co–linearity, curvilinearity, terminations at a common point, terminations at a continuous curve, crossing of continuous curves, parallelism, lines converging at a common point, equal spacing, virtual lines and points, and shadows creating parallel virtual lines. From these properties found in the retinal image, inferences can be made on the properties of the 3–D object observed. A parallel model (as opposed to a modular account) of the visual system is proposed based on the spatial information of the visual scene because Lowe argues that it is more in keeping with our knowledge of how the human visual system operates. In this model perceptual transformations or alignments are applied uniformly to the whole image.

The difference between the models proposed by Ullman and Lowe is that Ullman's model used pictorial alignment whereas Lowe's model used alignment of the object contours only. However, both of these models can recognise novel views of objects without any recourse to top down processing.

The issue of corresponding between the image and the representation has already been mentioned above and it was argued that theories of recognition need to specify how the visual system ultimately chooses the correct corresponding representation without having to search through all representations. Lowe (1985) argued that this process did not proceed in an analogue fashion but that choosing the correct representation was based on the statistical probability of finding the representation. Ullman (1989) on the other hand postulated that a measure of the degree of match between the aligned input and the object representation (which is in pictorial form) is required in order to decide which of the representations resembles the input most closely.

1.1.3.3 Interpolation Models

An alternative approach to the simple transformation models and the alignment models has recently emerged (Edelman and Weinshall 1991; Cutzu and Edelman 1992; Poggio and Edelman 1990; Intrator et al, 1991). Like Tarr and Pinker (1989), these workers also suggest that representations of objects are multiple and not singular. According to this approach, an object is represented by a few of its 2–dimensional views, encoded as clusters in a representational space. They proposed that the human visual system relies not on linear transformations or normalisation between the image and a stored representation but on an interpolation process between these representations. With this method the representation 'best fitting' the description of the input is chosen as a match to indicate the input's identity. Because the visual system represents objects in a number of views which are close enough to each other in multidimensional space, the need to transform inputs to match a stored representation is avoided. Instead the visual system interpolates a disoriented object between the stored views of that object. This model is also sensitive to viewpoint in that the further the novel view is from the stored views the more difficult it is to recognise. Poggio

and Edelman(1990) argue that "having enough 2-D views of an object is equivalent to having its 3-D structure specified".

Based on Poggio and Edelman's (1990) algorithm, Edelman and Weinshall (1991) show how a view interpolation model could be implemented in a neural network which uses non-linear interpolation. Their self-organising network model, called conjunctions of localised features (CLF) model, was trained on a few views of novel 3–dimensional objects which resulted in compact representations of the specific trained views. These compact representations are referred to as clusters of stored views and each cluster has its own representational space. Each cluster corresponds to a specific object and the centre of each cluster represents the standard or prototypical view of the object. A centre can be updated according to the amount of information that is available about the object e.g. the number of different views that have been inputted. Learning of a new view of an object results in the addition of a new unit to the representational cluster.

Recognition occurs by applying a function to the input in order that the appropriate stored object cluster is accessed. A multi–variate function can be derived for each object from a small number of views of the object such that each function is specific to each object. The differences between functions therefore correspond to the differences between objects so that when an inappropriate function is applied to the wrong object then the wrong representational view will be chosen and this can be easily detected as being incorrect. These functions are termed radial basis functions and are applied to each inputted view according to the distance between the input view and the centre of the basis unit. The distance between the novel view and the centre is measured by the distance between the extracted features of the image and the representation (Intrator et al, 1990). In the CLF model these features are extracted at an early stage in the process. Other objects and novel views of the correct object are rejected if the Euclidean distance between the features of the input and the centre of the cluster is above threshold (see also Cutzu and Edelman, 1992). Their model shows a lack of generalisability to recognise novel views of objects that differ more than 30° from the nearest stored views which according to Poggio and Edelman (1990) is compatible with the finding that people have difficulty in recognising novel objects when viewed 30° from the trained views (Rock and DiVita, 1987).

In sum these approaches assert that objects are represented by a number of different viewpoints. The views of objects are encoded as retinotopically organised features and are constructed to form complete view-specific object representations. Views are stored in clusters in representational space. Practice with novel views of objects results in the construction of a new representational unit. Matching between an input and a stored representation is determined by the distance between the features of the input and the stored representations and the application of a generalised radial basis function to this distance. Novel views are therefore interpolated between the stored representations. An appropriate match will be made between the input and the stored representations when the distances between them are

below threshold.

Important characteristics about the view interpolation approach are a) that stored views of objects are not only view dependent but are 2–dimensional, b) there is no need for any sort of transformational or normalisation processes and finally, c) an extraction of the features of the observed views is fundamental to the construction of representations.

The following sections of this chapter show how the issues outlined above have become significant in different approaches to object recognition. The contributions from animal physiology, single unit studies and agnosic studies are discussed and an attempt is made to relate the different findings from these diverse areas. To date, the hypotheses proposed to link behaviour to the physiological processes are tentative at best but more recent computational models are built using constraints which have physiological foundations.

1.2 Animal Psychophysiology

Physiological studies with animals have attempted to determine the neural substrates that underlie object recognition processes. Recent work has concentrated on the discovery of multiple visual areas in the cortex which may be involved in the recognition processes in humans. Cortical areas, that perform extensive analysis of the visual image beyond that carried out by the primary visual cortex or V1, are being discovered at a rate of about one every two years. A major task for neurophysiological investigators is to understand how these 20 or more visual areas contribute to visual perception and to visually guided behaviour.

1.2.1 Inferotemporal Lesion Studies

It has been found in a number of different studies that the inferotemporal cortex plays a central role in the discrimination of shapes and objects (Weiskrantz and Saunders, 1984; Gross, 1978; Holmes and Gross, 1984). In one such study, Weiskrantz and Saunders (1984) were interested in determining the brain regions that were involved in the transformation of objects that preceded recognition and whether these regions were separable from the regions which acquired an initial representation of a novel object. They found that the posterior inferotemporal lobe addresses viewer-centred information and that the anterior inferotemporal lobe was concerned with the storage of an object-centred model. In their study monkeys were initially trained to discriminate objects from a set of distractors. Having reached criterion, the monkeys were then tested on a discrimination task which occasionally included a transformed target object. The transformations of the objects included size, orientation or shadow configurations. Weiskrantz and Saunders found that in the initial learning stage, prior to the transform tests, monkeys with inferotemporal (IT) lesions (especially anterior inferotemporal lesions (AIT)) were selectively impaired compared to

monkeys with parietal, superior temporal sulcus (STS) lesions or unoperated monkeys. In the transformation condition, the prestriate and IT groups were impaired to the same extent compared to the parietal group. The IT group were also slower at reaching criterion on the orientation transformation condition than the other groups. It was revealed that the AIT group took longer to learn postoperatively than preoperatively which suggests that the AIT is more important for visual discrimination learning than other regions of association cortex. In fact, in a discrimination task where the animals had to choose a foodstuff from non food items the IT group were 8 times worse than any other group. Therefore it could be argued that the IT is not involved in the acquisition of a new object but in the retention of already stored representations. There was no effect of lesions to the STS or the posterior parietal lobe on the discrimination tasks.

Weiskrantz and Saunders suggested that these results indicate increasingly severe discrimination learning deficits the further forward the lesion extends in the AIT region. Transformation deficits increase as the lesion extends into the prestriate cortex particularly to the posterior IT region. They concluded that the AIT may be involved in the storage of a prototype of an object which is invariant to different transforms. Lesions in this area therefore result in impaired acquisition of this type of representation and even when it is acquired the representation is more flawed than in the control group. Furthermore, the posterior regions of the IT cortex may process transformations necessary to match the input with the stored prototypical view. They suggest that lesions in the posterior IT area causes an impairment in adjusting a viewer-centred representation to match the object-centred prototype in the AIT. However, this model is speculative and their data is not conclusive on the nature of the representations in these areas. Suffice it to say that the data supports the notion of a storage mechanism in the AIT area and a transformation mechanism in the posterior IT area. Further research is needed however to indicate the nature of the stored representation in the AIT cortex.

Many other studies have supported the role of the inferior temporal cortex in learning to discriminate objects and shapes (Gross, 1978) particularly in different orientations (Holmes and Gross, 1984a; Holmes and Gross, 1984b; Gaffan, Harrison and Gaffan, 1986). Holmes and Gross (1984a) reported an interesting finding that animals with IT lesions were not impaired at learning to discriminate stimuli which differed in orientation by 60° or more. In a previous study Gross (1978) found that learning to discriminate between two identical patterns that were rotated from each other by 90° or 180° was not impaired in monkeys with IT lesions. The IT group were found to have impaired retention of the preoperatively learned discriminanda but their performance after relearning the discriminations was comparable to the control groups.

In both studies discrimination between different patterns was impaired. The Holmes and Gross (1984a) study replicated these findings but also found that IT monkeys were impaired at discriminating between two rotated identical shapes or rotated 3–dimensional

objects provided the rotation difference was small i.e. less than 60°. As the IT group were not impaired at discriminating between identical shapes when sufficiently rotated then this may suggest that the control group viewed the rotated patterns as equivalent to each other whereas the IT group did not which made discriminations more difficult for the control group and resulted in little difference between these groups. Under small rotational differences between the identical shapes both groups may have viewed the stimuli as equivalent. The IT group were severely impaired at learning to discriminate between slightly rotated patterns compared to the control group because, according to Holmes and Gross, the IT group have an impaired ability to perceive shape constancy.

The common pattern in the data reported by Holmes and Gross (1984 a; 1984b) is that the more difficult the task is for the control group, the more impaired the IT group will be at learning and discriminating between two shapes. A consistent marked deficit in learning to discriminate between different shapes or objects was found for the IT group compared to a control group therefore this suggests that the IT cortex is required for discriminating different visual, complex patterns, but not the same patterns that differ in orientation (Gross, 1978).

There is, therefore, overwhelming evidence that the inferotemporal cortex is involved in perceptual categorisation. This conclusion however, could be challenged on the grounds that monkeys with inferotemporal lesions may lose the ability to associate a reward with a particular stimulus rather than the ability to discriminate between two objects. Mishkin ruled out this possibility by developing a nonmatching-to-sample task where the monkey had to choose the object other than the previously learned object in order to obtain the food reward (see Mishkin and Appenzeller, 1987). A new object is encountered in each trial and because the food is always associated with a novel object, the monkey has to rely on discrimination in order to obtain the reward. The results from this task were the same as those on the match-to-sample task for monkeys with inferotemporal lesions suggesting that the inferotemporal cortex is involved in discrimination.

Other work on IT cells have found that their receptive fields are not specific in the sense that the cells are not tied to the locus in visual space of the object being recognised. More abstract information is therefore represented in this area (see Mishkin and Appenzeller, 1987).

1.2.2 Parietal Lesion studies

Other physiological work has confirmed that the parietal cortex plays an important role in visual processing. Ungerleider and Mishkin (1982) reported finding a visual pathway which emerges from the striate cortex and connects to the parietal cortex through a series of stations. They argued that spatial relations are analysed along this pathway, whereas the visual pathway leading to the inferotemporal cortex is more involved with identifying and discriminating objects. These pathways are referred to as the 'where' and

'what' systems respectively. Figure 4 below illustrates these different pathways in the cortex.

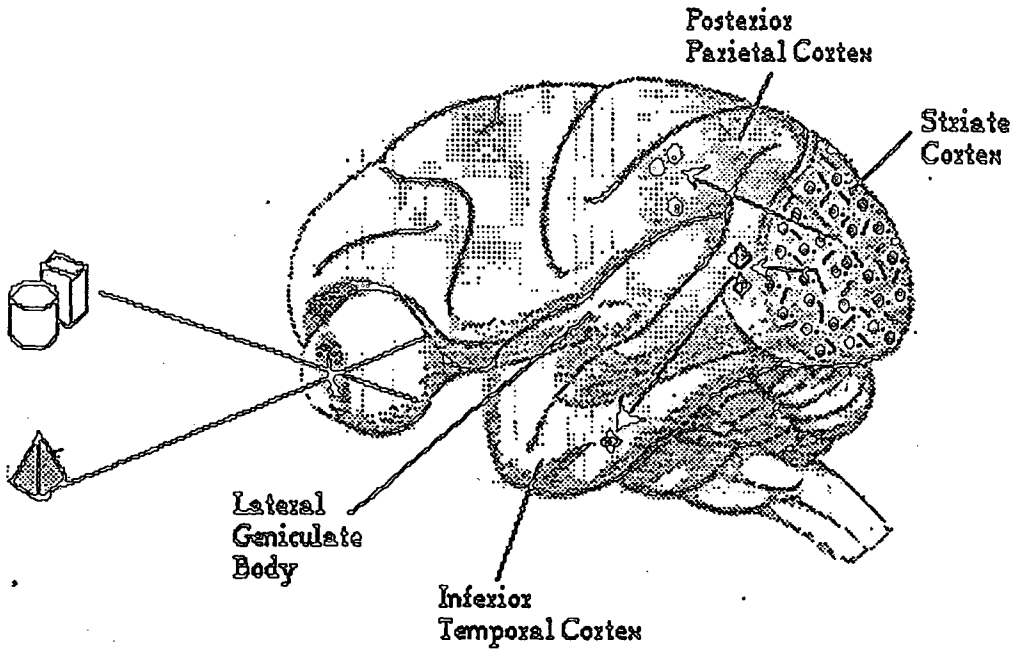

Figure 4: An illustration of the two pathways in the visual cortex which yield separate analyses of visual information. The pathway leading to the inferior temporal cortex deals with the identity and discrimination of objects whereas the pathway leading to the parietal cortex deals with the spatial relations between the objects (after Ungerleider and Mishkin, 1982).

In another study, Pohl (1973) found that monkeys with lesions in the parietal lobe were capable of discriminating between different objects but were impaired at perceiving the spatial relations between objects. The monkeys were presented with a task in which one of two covered wells contained a food reward. A cylindrical object was placed between the two wells and the well with the food reward was indicated by the proximity of the object to it. This proximity varied from trial to trial. Animals with inferotemporal lesions found the task relatively easy but the parietal lesion group were severely impaired. This results again supports the notion that the parietal cortex is involved in a spatial analysis between items in a visual scene.

1.2.3 Convergence of the Two Streams

The visual cortex is a highly interactive system and recent evidence has suggested

that the two visual pathways from the striate cortex converge. Young, (1992) has argued that information from the parietal and inferotemporal regions converge in the region of the principal sulcus (area 46) and in the superior temporal polysensory (STP) areas. Before this reconvergence there are very few connections between the parietal and temporal pathways. Spatial information is carried direct from V1 or via the mid-temporal (MT) to the parietal regions and other information in the visual scene is carried via V4 to the inferotemporal regions. This has led others to suggest that V4 is not only concerned with colour information but that it is also concerned with shape information (Walsh, Butler, Carden and Kulikowski, 1991). Finally, the information gleaned from both streams projects to area 46 and STP. These areas will therefore contain information about what an object is (from IT), where it is (from the parietal areas), its movement in space (from MT) and its colour (from V4).

1.3 Single Unit Recording Studies

Another approach to the functions of substructures in the prestriate cortex is to monitor the responses from single cells or groups of cells to different types of stimuli.

One of the earliest studies on cell recordings revealed that cells in the striate cortex responded most strongly to simple stimuli such as lines with specific orientations and bars with specific positions in the visual field (Hubel and Wiesel, 1962). Subsequent research revealed that visual information is processed in the prestriate areas of the cortex and that these areas typically deal with more complex or abstract visual information (see Cowey, 1985 for a review). Gross et al. (1972) recorded the responses of cells in the inferotemporal area to small shapes presented to the monkeys. They found that these cells responded to complex shapes (e.g. monkeys hands) within an area of 20° to 30° in the visual field. Gross et al. postulated that the areas along the visual pathway deal with progressively more information in the visual field with higher areas processing information about all of an objects physical properties leading to a full representation of the object in the anterior regions of the inferotemporal area.

Other single-unit recording studies have found that cells in the superior temporal sulcus respond more to faces than to a variety of other stimuli (Perrett, Rolls and Caan, 1982; Yamane, Kaji and Kawano, 1988). Perrett et al. found that these cells responded more to faces than other stimuli such as lines or gratings or more complex, potentially arousing stimuli such as hands, bananas or snakes. These cells were shown to respond to a variety of faces regardless of their size, orientation or position. They also found that some cells were sensitive to configuration in that they responded less to faces with jumbled features than to normal faces. However, it is not clear what sort of information the cells use to encode faces or which configural dimensions are important although Perrett argues that coding is for facial parts and their configurations. Yamane et al. (1988) found that single neurons in the gyrus of the IT in monkeys trained to discriminate 3 human faces from a large number of other faces were not

responsive to non-face stimuli but were responsive to familiar faces. A correlational analysis between the responses made to the faces and the quantified facial features revealed that the face neurons (sic) detected combinations of distances between facial parts such as eyes, mouth, hairline and so on. Differences between these distances may reflect differences between individual faces.

If the face neurons detect distances between face parts, how sensitive are they to different views of the same face? Perrett et al. (1991) found limitations in the cells ability to generalise over different perspective views of faces and found that the distribution of cells shows a clustering around four prototypical views; front view, left and right profile and back views. Some cells have also been found which are selective for identity but which are also sensitive to the view of that particular face (see Perrett et al., 1989, for a review). Perrett concludes that these cells store view-centred information about a specific face. These cells therefore seem to have some of the properties of the hypothetical 'grandmother' cells although nothing in Perrett's work suggests that single cells are uniquely responsive, rather that a network of cells store information about a single, familiar face.

One of the most important questions that needs to be asked of this research is whether these face neurons are truly selective for faces or whether they respond to some property of faces that could be found in other objects. As already mentioned Perrett et al. (1982) failed to elicit a response to face neurons from other stimuli that might be expected to be important to the monkey such as bananas or snakes. They also failed to elicit a response from the cells to pictures of jumbled up faces. Other studies have shown that face neurons respond to faces shown in a variety of mediums such as plastic faces or photographs (see Desimone, 1991). All in all, the data strongly favours the notion that there are cells in the inferotemporal cortex that are specifically tuned to face recognition.

1.4 Neuropsychological Evidence

1.4.1 An Introduction to Visual Agnosia

Visual agnosia is the selective impairment of visual object recognition. The term visual agnosia was coined by Lissauer (1890) who documented one of the first cases of object recognition impairment. Lissauer's patient, an 80-year-old salesman GL, received a blow to his head and lost the ability to recognise objects. His visual acuity was almost normal for his age. He also retained the ability to describe objects in conversation and could recognise objects from tactile information or by a characteristic sound they might make (e.g. a whistle). However, he mistook his jacket for a pair of trousers and thought that pictures in his room were boxes full of objects.

Lissauer diagnosed his patient as having visual agnosia. The term visual agnosia is

used to imply that the patient has a disorder that renders him unable to recognise things that he can see. No other deficits such as loss of language skills or general intellectual ability or visual acuity are associated with this impairment. Agnosias in the visual modality often occur for different classes of stimuli such as colours, faces or objects but most patients show significant impairments in all three classes.

There are two different kinds of visual agnosia documented in the literature; apperceptive and associative agnosias. Lissauer suggested that visual recognition can be separated into apperceptive and associative stages with the associative stage a higher visual process than the apperceptive stage. Consequently more cognitive or semantic information about the object is involved in the associative stage and the apperceptive stage involves purely visual or perceptual information. If the patient was impaired at copying drawings of objects or matching two pictures of the same object she/he is considered to have apperceptive agnosia (Warrington, 1985). Apperceptive agnosia occurs mostly in patients with right posterior lesions. These patients are unable to identify overlapping familiar objects and the Gollin's degraded picture test and degraded letters. Apperceptive agnosia is therefore the failure to organise a coherent percept. Warrington and Taylor (1973) tested patients on matching two different views of objects, one prototypical and one unusual view. Warrington(1987) argues that patients who are unable to efficiently allocate two stimuli to the same perceptual category are apperceptive agnosics. This deficit is more marked in patients with right hemisphere lesions than patients with left lesions (Warrington, 1982).

On the other hand, if the ability to copy or match objects is intact but the patient could not name the object then she/he is considered to have associative agnosia. Warrington (1985) refers to associative agnosia as a deficit of semantic categorisation and that it is normally associated with patients with left hemisphere lesions. Patients with a loss of semantic categorisation can show impaired performance at matching visually similar objects according their functions. They are also impaired at demonstrating the use of the objects by action or by mime.

Recent developments in the neuropsychological literature have suggested that this classification of agnosias is too simplistic and that a more sophisticated classification system is needed to incorporate all the neuropsychological evidence (Humphreys and Riddoch, 1987 and Farah, 1991).

Ellis and Young (1988) presented a functional model of object recognition and naming which served to highlight how agnosias can effect different visual systems (see Figure 5 below). They argued that most of the impairments associated with agnosia can be explained using a modified version of the Marr (1982) model on how the visual recognition system works. Marr's model is based on a modular organisation of the brain. They attribute the cause of apperceptive visual agnosia to being an impairment in the construction of the viewer-centred representation. Patients who are unable to match prototypical views of objects with unusual

or foreshortened views are considered to have impaired object-centred representations because these patients can typically recognise objects in the prototypical view suggesting that the viewer-centred representation remains intact. Finally, associative agnosia was considered to be an impairment of access to semantic information about an object.

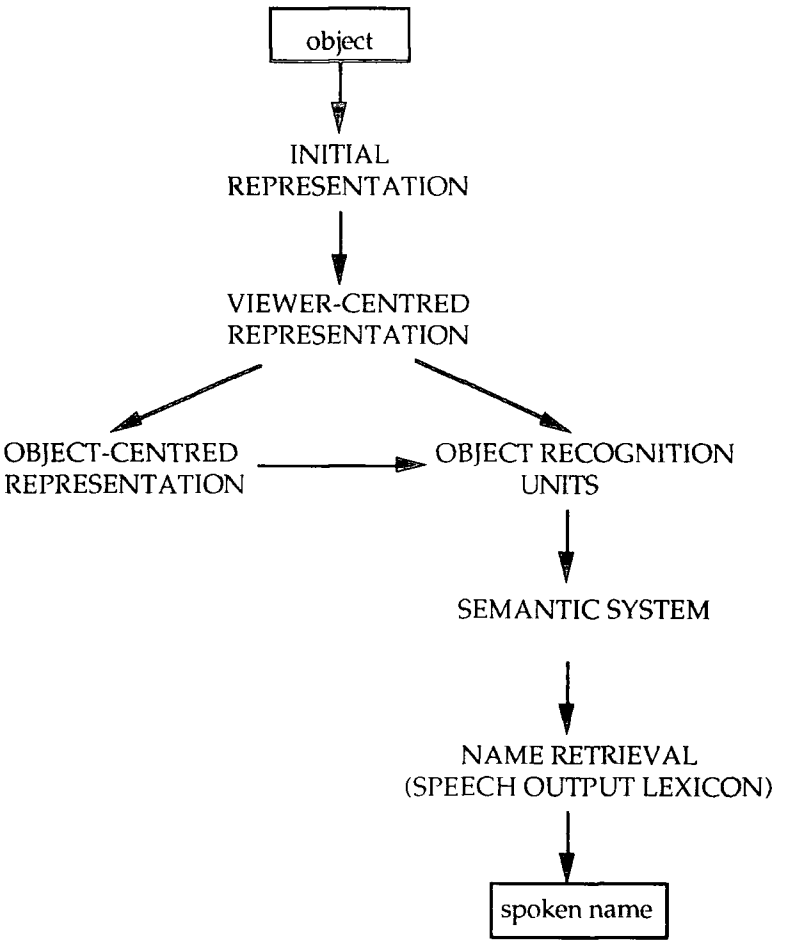

Figure 5; The Ellis and Young (1988) functional model of object recognition which serves to explain many of the deficits found in agnosic patients.

Within the visual modality agnosias are generally object specific in that there is often a dissociation between face recognition, word and object recognition when it comes to visual impairments. For example, a patient with impaired face recognition (e.g. prosopagnosia) can have an intact object recognition system (DeRenzi, 1986) and vice versa (see McCarthy and Warrington, 1990).

It could be argued that different functional components are involved in face recognition and object recognition and that the dissociations encountered between the two types of processing could be attributed to the different task demands (Farah, 1991). For

example, a task where the subject has to recognise a series of faces is a within category task. Object recognition tests are often between category tasks. Therefore, the dissociation may not reflect different storage areas in the brain for faces and separate areas for storing objects but that each makes different demands on the subject. Conversely, the evidence for dissociation between face and object recognition may reflect the different ways these classes of objects are represented in memory (Perrett, Rolls and Caan, 1982). Furthermore, faces may have different descriptions than objects in memory. Faces may be represented by the configural information between the features in each face and identifying a face over all other possible faces is based on differences between the configural information across faces (Young, Hellawell and Hay, 1987). It could be argued that objects on the other hand need not, in general, rely on configural information for representational purposes.

Theories differ on the nature of the object description in memory. Marr (1982) for example suggested that an object's description in memory is based on the principal axis of the object. Indeed Ellis and Young (1988) have adopted Marr's approach to object recognition to explain the deficits encountered in the agnosic literature. However, other models of object recognition have also been proposed and these models may well prove to be as instructive as the Marr approach in explaining object agnosia. For example, an alternative approach to object recognition proposes that objects are represented as they are viewed and a multiple set of different views in memory describe each object (Jolicoeur, 1992; Edelman and Bülthoff, 1990). The different classes of agnosia may reflect an impairment to transform across different views of simultaneously presented shapes (apperceptive agnosia) or between an inputted image and a stored view (associative agnosia). Support for the different models of agnosia proposed in the literature is discussed below.

1.4.2 An Outline of an Agnosic Case Study

Humphreys and Riddoch (1984, 1987) carefully investigated the visual abilities of patient HJA after he had suffered a stroke which produced a small bilateral lesion in the occipito-temporal region of the brain. He had severe problems in recognising faces and objects. His visual acuity was normal and his ability to discriminate length, orientation and position was intact. He was also susceptible to visual illusions. His perception of depth through stereopsis was also normal.

His agnosia could not be termed apperceptive because he was capable of making good copies of drawings of objects and matching prototypical views of objects to foreshortened views. He was also capable of drawing objects from memory. According to the Ellis and Young (1988) model, his viewer-centred representation would seem to be intact (see Figure 5 above). HJA was however, severely impaired in recognising objects. His ability to recognise objects seemed to be hierarchical according to the amount of information available about the object. For example, he was much better at recognising real objects (21/32) than photographs of the objects from prototypical views (12/32) or line drawing of the objects although it took HJA 25

seconds to correctly identify an object. Identification followed a laboured feature-by-feature description of the object and mistakes were usually made because he had either failed to identify or missed a feature.

In contrast to his good perceptual abilities (matching and drawing from memory) HJA's performance on accessing stored knowledge of objects from vision was poor. Humphreys and Riddoch argued that this impairment was not due to a disruption of his stored knowledge of objects because he was capable of describing the features and functions and other particulars of objects on request. His impairment therefore must be due to an inability to access this stored knowledge from visual information alone.

HJA's performance on an object constancy test was compared with the performance of four other patients with right hemisphere lesions (Humphreys and Riddoch, 1984). Humphreys and Riddoch tested the patients ability to match objects shown in a prototypical view with one of two other objects shown. The correct match object was shown in one of two conditions, in a foreshortened view or a minimal feature view. The foreshortened view had the effect of reducing the information about the principal axis of the object. In the minimal feature condition, the saliency of a distinctive feature of an object was reduced. Performance on naming objects in a foreshortened view was the same for all patients. HJA was significantly worse at naming and matching objects in a minimal feature view than in a prototypical view. This trend was not observed for the other patients. Humphreys and Riddoch argued that these findings suggest a dissociation between axis-based and feature-based descriptions of objects. The fact that HJA recognises objects after careful identification of each feature and cannot recognise objects if their salient feature is reduced suggests that his ability to represent objects in terms of their global characteristics is impaired. Indeed, in a subsequent study, HJA's performance was impaired when asked to identify letters which made up another larger letter (Humphreys, Riddoch and Quinlan, 1985). Humphreys et al. argued that his representations of objects were based on the local characteristics of the shape independently of the global characteristics and that he typically uses local features to identify objects because of this difficulty he has in segmenting features appropriately.

According to the Ellis and Young (1988) model of object recognition, HJA is impaired at accessing the object recognition units from the viewer-centred representations (see Figure 5). However, their model does not account for the different routes to object constancy suggested by Humphreys and Riddoch (1984) study outlined above. For example, there is no indication from the Ellis and Young model that either local visual information (as exploited by HJA) or global visual information (as exploited by the other patients) can be used to build representations of objects in memory.

Other findings outlined below have argued that a division of agnosia into two classes is too simplistic and that models of object recognition need to incorporate such findings.

1.4.3 Support for Apperceptive and Associative Agnosia

Warrington (1982, 1985) argued that the two stage classification system is a useful model of agnosia. She proposed a model based on Lissauer's original classification which states that pre-categorical and post-sensory information comes from both visual cortices to the right hemisphere where perceptual categorisation takes place. The output from the perceptual categorisation stage goes to the semantic categorisation system in the left hemisphere. Warrington (1982) also argues that deficits such as impaired shape perception, achromatopsia and visual disorientation are not deficits of object recognition per se and are therefore termed pseudo-agnosic syndromes. Because these impairments affect the pre-categorical stage of the perceptual system, then any subsequent processing is impaired as a consequence and the patient has impaired recognition. However, Warrington argues that because the recognition system itself is not affected but that inputs are defective then these impairments are pseudo-agnosic.

Warrington argues that a serial model of the post sensory categorical stages of object recognition best explains the deficits found in agnosic patients. The first categorical stage, the perceptual system, is said to receive input from both visual cortices and is therefore post sensory and pre-semantic. This system is lateralised to the right posterior cortex. Damage to this system results in apperceptive agnosia. The second categorical stage, the semantic categorisation stage is lateralised to the left posterior cortex and receives input from the perceptual system. Damage of this system results in associative agnosia.

A more recent case study however, has suggested that the serial model of apperceptive and associative agnosia is an inadequate one. Goodale et al. (1991) studied a patient DF who showed all the classic signs of apperceptive or pseudo-agnosia. However, despite not being able to match shapes or match the orientation of a card with the orientation of a slot, DF was capable of posting the card skillfully through the slot. When asked to reach out for objects, DF made the appropriate grasping actions and positioned her fingers normally when asked to pick up the objects. This suggested to the authors that DF was capable of covert recognition and that her perception of orientation and shape was intact enough to allow her to grasp objects. Cowey (1991) also argues that DF's apperceptive agnosia is not due to an inability to create representations of objects or shapes. This finding also suggests that shape perception can occur without awareness in parallel with processing that leads to conscious awareness.

1.4.4 Other models of Visual Agnosia

Humphreys and Riddoch (1987) however, questioned the classical distinction between apperceptive and associative agnosia by arguing that this classification is too simplistic. Also, in most of the reported cases of apperceptive and associative agnosia, there does not seem to be any correlation between the agnosic classification and the damaged

anatomical substrate or area in the brain. There seems to be at least five areas in the brain that when damaged cause some form of agnosia. Therefore, neither apperceptive nor associative agnosia can be associated with a particular lesion site. They argued that a model of object recognition based on the deficits found in agnosic patients should account for recognition in normal subjects.

Humphreys and Riddoch propose a model of object recognition which categorises the deficits reported in the literature into seven different classes; impaired shape processing, impaired transformation processes, impaired integration processes, loss of stereoscopic vision, impaired access to form knowledge, impaired access to semantics and impaired semantic knowledge. Deficits can occur within a stage (e.g. semantic information) or between stages in visual processing (e.g. transformation processes). They argue that all of the case studies reported can be accounted for in this model and that neuropsychological evidence supports this fractionation. Finally, they illustrate how the recognition system that could account for the deficits in agnosic patients might work. They believe that their model has an advantage over the classical apperceptive/associative model in that it is open to amendments based on future evidence from agnosic patients whereas the apperceptive/associative model is not.

Following Humphreys and Riddoch, Farah (1991) also argues that a two stage model of agnosia is too broad a classification after reviewing the literature on agnosia. She argued that there was too much heterogeneity within each class. For example,the term associative agnosia has been applied to patients with optic aphasia (an inability to name an object whilst retaining other semantic information such as its function) and patients with a general loss of semantic information not related to the visual modality.

Instead Farah proposes that deficits of the general recognition system (i.e. recognition of faces, objects and words) do not fall into discrete categories but that they reflect different levels of perceptual abilities. Deficits at the highest level of shape processing underlie associative agnosia. She noted that many of the cases reported in the literature described impairments that were overlapping across the different categories rather than being associated with a single agnosic category. Prosopagnosics for example may also be impaired at recognising man made objects such as buildings or public monuments. Patients with such deficits cannot be termed associative object agnosics because they are capable of recognising a myriad of other objects. There seems to be evidence for a double dissociation between face and object recognition but occasionally the type of visual processing associated with face recognition may apply to a number of different objects. The clue to the workings of the visual system does not therefore solely come from the different types of deficits observed but also from the nature of the stimuli which are being tested.

In reference to the example given above, there may be some common processing of faces and buildings. Other classes of objects may require parsing and subsequent recognition

relies on the structural descriptions between these parts. Other more simple shapes may not require parsing and are therefore processed differently and will subsequently not be affected by any impairment of the decomposition process. Farah argues that impairments of these processes underlie the range of associative agnosias. A mild impairment of the part decomposition system allows more objects to be recognised but objects with more complex parts are difficult to recognise. A severe impairment of this recognition system would result in impaired recognition of many more objects but recognition of simple, perhaps single-part, objects remains intact. She argues therefore that face recognition, object recognition and word recognition do not involve unique processing but that they involve different levels of the same process i.e. the process of decomposing objects into their component parts and representing objects as spatial relations between these parts.

Prosopagnosia is an example of this impairment because faces, Farah argues, are represented as single complex parts. That faces are represented as single parts is supported by the fact that familiar faces can be recognised at very short exposure times and also within a single saccade. Valentine (1991) supports this argument by stating that "decomposition into parts may be intrinsically an inappropriate approach to describing faces" because he suggests that the spatial relationships between the parts would have to be sufficiently accurate in order for individual faces to be recognised. Also, prosopagnosia does not in itself provide evidence for the uniqueness of face recognition but could indicate a selective impairment in a more general recognition system.

Valentine (1988) also questioned whether face recognition was qualitatively different from other forms of recognition and concluded that the evidence for face recognition as a unique process was equivocal. The effect of inverting faces on recognition has previously been used to support the hypothesis that faces are special (Yin, 1969), however recent studies have argued that the inversion effect may be due to expertise and familiarity of the stimulus class rather than to the unique processing of faces (Diamond and Carey, 1986; Valentine, 1988).

1.4.5 Object Recognition models based on Agnosia

Evidence for the modular organisation of the brain from studies of visual agnosia is equivocal. The Ellis and Young (1988) model of object recognition postulates that brain injury can cause an impairment in building any one of the object representations (i.e.view-centred or object-centred) or in accessing stored knowledge about the object. This model however, does not account for the different types of information used to build representations of objects. Their model is based on the Marr model of visual recognition but other interpretations based on other models of visual recognition could explain the different manifestations of agnosia without recourse to Marr's ideas on representation. For example, support for the notion of an object-centred description in memory is not forthcoming from the neuropsychological literature and that viewer-centred models of object recognition could also apply to the

different impairments found.

More recent interpretations of the deficits in visual agnosia emphasise the importance of integrating the parts of objects in order to process more global information about objects (Humphreys and Riddoch, 1987; Farah, 1991) suggesting that brain damage causes an impairment in the decomposition and integration of parts rather than specific impairments in the processing of specific classes of objects. Valentine (1991) on the other hand suggests that structural descriptions may not at all reflect the processing of the visual system and that some sort of pictorial alignment model (Ullman, 1989) could account for the representations of objects.

1.5 Linking the Different Approaches

An attempt to link the major findings from the different areas in object recognition is included in the sections below. An initial comparison between the findings from the animal physiological literature and the human neuropsychological literature to the experimental approaches is discussed.

1.5.1 Are there Parallels between Human and Monkey Brains?

There are some drawbacks in studying animal brains in order to provide more information about the workings of the human brain. For a start the brain of the macaque monkey is about one-sixteenth the size of the human brain. Also the human brain is a lot more complex than the macaque brain. The structures in the macaque brain all have counterparts in the human brain but it is difficult to tell what processes these structures are involved with in the human brain. We need to consider the fact that the functions of these structures may well be different in the human brain due to the course of evolution. An obvious example of a function peculiar to the human brain is the capacity for language.

Nevertheless some parallels can still be drawn between the macaque and the human brain that are useful in guiding an understanding of the functions of human brain structures. For example, Warrington (1982) found that patients with lesions in the right posterior hemisphere were impaired at perceptual categorisation. The same conclusion was met by Weiskrantz and Saunders (1984) to explain the impairment of learning to discriminate transformed objects in monkeys with inferotemporal lesions. Both studies concluded that these lesions removed the capacity to develop object-centred descriptions of the visual scene. Other common deficits have been found between patients and monkeys (see Cowey, 1985 for a review) but such direct comparisons are rare. One of the problems in looking for common functioning in the monkey and human brain is that patients with brain injuries do not usually have damage confined to a specific structure in the brain and damage most often affects many structures. Monkeys on the other hand can have very localised lesions which makes it easier

to monitor the effects of damage to that specific structure. The recent development of PET scans may prove to be more indicative of the locus of functioning in the human brain by monitoring the blood concentration in the different structures. However, it is too early to draw any conclusions about the workings of the human prestriate cortex from similar lesions in the monkey.

A parallel between studies on animals with lesions in the parietal lobe which results in impaired judging of spatial relations between objects can be found in the agnosic literature. In a review of object agnosia, Farah (1991) argued that two kinds of representational processes underlie the recognition of faces, objects and words. She hypothesised that these representational processes may involve the processing of complex parts and processing of multiple parts, and spatial relations are fundamental to the latter processes. Humphreys and Riddoch (1984) also found that HJA's injury rendered him reliant on local information of objects, particularly to the spatial relations between features, because his global information processing was impaired. They concluded that there are two routes to object constancy, one based on global descriptions relative to the principal axis of an object and the other based on structural descriptions between the local features of the objects. Animal studies of the separate functions of the parietal and inferotemporal lobes lends credence to the notion of two representational systems that characterise the deficits found in agnosic patients although this link may be tenuous.

Mishkin and Appenzeller (1987) have argued that there are two types of memory pathways, one that deals with recognition based on distinguishing between two objects and the other based on spatial relationships. Figure 4 above illustrates the workings of the parietal and IT areas in analysing visual information. Cowey (1985) argues that information in these pathways can proceed in parallel. The Humphreys and Riddoch (1984) study certainly supports the notion that when one route is damaged, the other can be relied upon for visual purposes. However, as already mentioned the anatomical link may be tenuous because the Humphreys and Riddoch study may be open to another interpretation: the results may reflect the differences in accessing the representations of the stimuli not the differences between two memory pathways. The spatial relations between the features might not need to be processed in order to recognise the object, the patient may simply have lost the stored representation of that object in memory or the ability to transform a novel input to match to the nearest stored view. For example, HJA may find it more difficult to transform an image of an object that is rotated so that a feature is reduced in salience and then match it to its appropriate representation in memory than to transform an image of an object with its axis foreshortened. This notion of transforming an input to match a stored view was already discussed (see 'View-centred Approaches' above).

More recently, Biederman and his co-workers found that the AIT is not necessary for object recognition (Biederman, Gerhardstein, Cooper and Nelson, 1992). Seven patients with unilateral temporal lobectomies in which the anterior and medial regions of the inferior

temporal lobe were removed were not found to be impaired at an object– matching task. Two pictures of objects, which were either the same object (same object shown from different viewpoints) or different named but visually similar objects, were presented in succession to the group of patients and a control group. Line drawings of different objects, including familiar and nonsense objects, were projected to the lobectomised hemisphere at durations too brief to make a fixation. It was expected that performance on the same/different task would have been much worse for the patients when the images were presented to the visual field contralateral to the lobectomised hemisphere. This prediction was generated from previous studies in which animals were found to be impaired at learning to discriminate between two shapes (Weiskrantz and Saunders, 1984). Weiskrantz and Saunders argued that the inferotemporal cortex is involved in generating 3-D object-centred descriptions which are invariant to viewpoint. Biederman et al. therefore tested whether the IT was involved in high level recognition in humans. They found no difference in performance for images projected to the lobectomised and normal hemispheres. The effects of the rotated objects was the same for the lobectomised group and the control group. Moreover, when the patients were asked to name the objects presented, there was no difference found between naming pictures that were projected to the lobectomised and the normal hemisphere. There was an overall difference found between the patient group and the control group in the time taken to match or name objects. However, Biederman et al. argue that this result is open to different interpretations, for example, it may reflect differences between the groups that existed prior to surgery. Also, after a unilateral temporal lesion, visual information can still access the intact contra-lateral temporal lobe via the cerebral commisures. Another problem with this study is that the number of different rotations used was quite small. The differences between the control group and the patient group may well reflect an impairment due to the lobectomy. Holmes and Gross (1984) found that monkeys with inferotemporal lesions were impaired at learning to discriminate between objects that were rotated less than 60° from each other. Larger rotations reduced this impairment. As Biederman et al. used rotation differences of up to 60°, it would have been interesting to find whether the reaction time differences reduced with larger rotations. The results may then show a more direct correspondence to the lesion studies with monkeys and it could well be concluded that an impairment was present in the lobectomised group.

It has been argued that the work on single-unit recordings reveals a property of the visual system that resembles a parallel distributed network for the coding of general object features (Desimone, 1992). Desimone argues that there is no evidence to suggest that face neurons respond exclusively to the face of one individual (c.f. Perrett et al., 1987) and that these cells are more likely to respond to different facial features and therefore collectively respond to a single face (Perrett et al., 1989). This suggests that the interconnections between the cells are important in determining the identity of the individual observed. This notion of storing icons of different parts of objects or faces has also been supported by Nakayama (1989). He suggested that information proceeds along the visual system in the form of a processing pyramid. The bottom of this pyramid holds the finest detail of the visual scene and the top

holds more coarse information. This feature pyramid is subsequently connected to stored information. This stored information is in the form of low-resolution icons and recognition corresponds to an aggregate of icons responding simultaneously. According to Nakayama, these icons are view-centred and responses depend therefore on the nature of the input. This model seems to support well the data from the single-unit recordings (Perrett, Rolls and Caan, 1982).

However, the neuropsychological data, particularly the studies on prosopagnosia also suggests that a loss of face recognition is often coupled with a loss of recognition of some objects. Perrett et al. (1982) found that some of the face cells respond only twice as well to faces as to nonface stimuli suggesting that these cells respond to some general feature of faces such as the configuration of the features and that this configuration may also apply to objects such as buildings. Indeed, prosopagnosics are often impaired at recognising buildings (see Farah, 1991). If we apply this data to Nakayama's model it seems that this model has a major shortcoming in that the spatial relations between the icons or features of an object are not specified. It would seem that this information is necessary particularly for objects such as faces or buildings. In proposing a model of object representation and recognition that relies on parts, geons or icons, then it should be a fundamental property of this model to incorporate how the spatial relations between the parts are encoded. This has proved to be difficult (see section on Biederman's model above). Models proposed that involve linking features that are processed simultaneously have also been criticised as being unrelated to human vision (Parker, 1989). Parker argued that features are not processed simultaneously but that low spatial frequency information is processed more rapidly than high spatial frequency information. It may be that the high level features are nested in the coarse, low-level image and that recognition of the object proceeds in this fashion. The parallells to the Nakayama model are obvious. Nevertheless, an alternative model that does not parse the object but uses a pictorial description as a representation may be more appropriate to the workings of the human visual system. Such a model has been proposed by Ullman (1989), Tarr and Pinker, (1989) and Edelman and Weinshall (1990) and was discussed in earlier sections of this chapter.

1.6 Conclusions

The focus of this chapter was on the different approaches to object recognition in psychology. A discussion of the experimental and computational approaches concluded that models of object recognition can be divided into two broad classes, object-centred approaches and view-centred approaches. The various models proposed within these classes were outlined. Secondly, the findings from animal physiological studies were discussed in the context of the underlying neural substrates to visual recognition. Most animal studies have found support for two separate pathways from the striate cortex; the infero-temporal pathway which deals with object discrimination and the parietal pathway which deals

with the spatial relations between objects. A review of single-cell recording studies revealed support for the notion of view-centred representations stored in the superior temporal sulcus. Finally, the neuropsychological literature revealed some support for dual-processing of visual information in the human recognition system. The findings in the neuropsychological literature were discussed in terms of the different models of object recognition outlined in the initial sections of the chapter. Finally, a link between the findings from these different approaches was discussed.

In general, the links between the animal and human studies are very tentative although there seems to be some obvious parallels such as the role of the prestriate cortex in the recognition and discrimination of complex images. The receptive fields of the cells along the visual pathway from area V1 to the prestriate cortex become larger and consequently more complex shape information is represented such as faces for example. Some computational models have proposed that the information stored to represent objects is more abstract in the higher visual areas. Such representations may therefore include 3-D object models which are invariant over different transformations of the object in the environment. The neurophysiological evidence for this type of representation has not been forthcoming. On the contrary, the evidence favours more view-centred models of object representation that propose that a collection of views serve to represent an object in visual memory (see Perrett et al., 1982, 1989). However, as Perrett et al (1991) note, view-independent descriptions can be created by combining the output of several view-dependent descriptions.

The next chapter reviews the experimental literature on object recognition with reference to the evidence for the different models of object recognition discussed in this chapter.

Chapter Two

Recognising different Views of Objects

This chapter concentrates on the experimental evidence for each of the different theoretical approaches to the recognition of objects in different viewpoints discussed in the previous chapter. This area of research is a relatively new one and therefore not much work has been done that unequivocally supports one theory over another. Initial theories concentrated on shape and pattern matching. More direct theories on object recognition stemmed from the results of experiments on the recognition times to disorientated line drawings of objects. Recent theories are based on evidence from experiments using 3-dimensional, wire frame objects that resemble paper clips. These experimental findings are obviously removed from the recognition of real 3–dimensional objects and it is difficult to tell whether the pattern of results observed in the literature can be extended to explain the processes behind the recognition of familiar, 3–dimensional objects.

In summary, in reviewing some of the experimental evidence on the different approaches to the recognition of objects in different orientations, this chapter is structured around two general sections. In the first, the evidence supporting the object-centred approach to object recognition is reviewed. Such a model was proposed by Marr (1982) and an account of his theoretical approach was given in chapter 1. A review of the experimental evidence for the object-centred approach espoused by Biederman (1987) is also discussed. In the second section, the empirical evidence for a view-centred approach is discussed. Recent theories on object recognition have argued for a view-dependent rather than an object-centred approach. Such theories such as those of Edelman and Weinshall (1989) and Jolicoeur (1992) have been outlined in chapter 1. This section also includes a review of the evidence for the multiple stored view approach of object recognition. Towards the end of this chapter a discussion of evidence that the mapping of an input to a stored view is mediated by either a simple transformation or that novel views are interpolated between stored views is included. Finally a discussion of the shortcomings in the experimental literature will be given including a brief introduction to the chapters on the recognition of disoriented objects presented foveally.

2.1 Early Approaches to Object Recognition

In one of the earliest studies on object recognition, Arnoult (1954) reported that shape discrimination was view dependent. He tested the effect of varying the angular difference between stimuli in a shape discrimination task. He found that reaction times and

errors increased as a function of the angular distance between a pair of shapes[1]. He concluded that the angular distance between shapes resulted in the judgment of two rotated shapes as being the same shape more difficult. In other words, the ability to discriminate between shapes was found to be dependent on the view of the shapes.

Arnoult's study on shape discrimination used arbitrary, silhouetted shapes as stimuli. In a later study on shape discrimination, Bartram (1976) studied the effects of matching across different views of pictures of objects. He tested the effects of three viewing conditions on the matching times of line drawings of objects. The three conditions were; identical pictures of an object, different views of the same object and different pictures of objects having the same name. Like Arnoult, Bartram found that matching times depended on the angular distance between views of the same object, in that identical pictures were matched more readily than different views of the same object. Different objects with the same name were slowest to match. However, when the stimuli were photographs of objects, as opposed to line drawings, then there was no difference found between the identical picture condition and the different view condition for highly familiar objects. The same effects observed for line drawings of objects were found for less familiar photographs of objects. He concluded that at least three levels of coding are involved in shape discrimination tasks; a picture–code level, an object-code level and a non-visual semantic code. The picture–code level is involved in matching across similar shapes or objects when the stimuli are unfamiliar (such as line drawings or photographs of unfamiliar objects), whereas an object-code is involved in matching across more familiar shapes, such that two views of an object are readily perceived as being of the same object.

However, as Quinlan (1991) argued, there is a subtle difference between tasks that involve pattern classification or distinction and pattern identification. There may be different processes involved in assigning a shape to a particular category and being able to identify a pattern. The latter involves accessing knowledge about a particular shape from memory. Therefore, investigations of naming latencies of different views of objects may be more appropriate to the study of the recognition of disoriented objects than shape matching.

Other early work on the recognition of objects did in fact use naming latencies as the dependent variable. Bartram (1974) tested the effect of naming latencies across different views of objects. He found that subjects could name pictures of objects more rapidly when preceded by a trial containing an identical picture of the object. They were slower at naming an object when it was preceded by a different view of the same object and slower still when preceded by a different object with the same name. He also found that practice reduced the naming latencies across the same view of an object and that this effect transferred to the different views of the same object. Bartram concluded that two different visual codes are

[1] Later studies on discrimination between rotated pairs of shapes attributed the function between reaction times and angular distance to mental rotation (Shepard and Metzler, 1971; Cooper and Shepard, 1973).

involved in naming across different views of objects; a 2-D picture code and a 3-D object code and that information may be encoded using either of these codes. He also postulated that a name code and a non-visual semantic code are involved in accessing information about an object.

Both of Bartrams' studies discussed above set a precedent for the empirical work on object recognition, particularly the effect on the recognition of different views of objects. His findings were important in highlighting the issue that recognising objects from different views can depend on the view of the object observed.

However, the studies reported above have limited implications. The effects observed were relative to the differences across two views of the shapes or objects used as stimuli in the experiments. As such, the findings are not conclusive evidence that object classification or identification is absolutely view-dependent but is only view dependent when simultaneuosly compared to one other version of the same stimulus. The findings therefore do not discount the notion that recognition may be view–independent if given all possible orientations. Indeed, Bartram (1976) found that matching objects was view–independent for highly familiar objects. Familiarity may indeed be important for storing view-independent information. This issue will be discussed in more detail later on in the chapter. The following section reviews the empirical support for the theory that recognition is view-independent or object-centred.

2.2 Evidence for Object-Centred Approaches

A number of different studies have reported finding experimental evidence that objects are represented as object-centred models and that recognition is therefore invariant over different orientations (Biederman, 1987; Biederman et al. 1991, 1992; Ellis et al., 1989).

Biederman included experimental evidence to support his idea that objects are represented as parts or geons (Biederman, 1987). According to Biederman, objects are parsed at regions of sharp concavity in the edges of the object and objects are represented as spatial relations between these parts (see Figure 2, Chapter 1). He found that subjects made more errors to depictions of objects with some of their parts removed. However, for complex objects shown under brief exposures, the error rate was low when only half the number of the object's geons were depicted. He also tested the effects of degraded images of objects on recognition. Errors were measured in naming objects with deleted contours. He compared recognition accuracy between two sets of degraded objects on recognition; in one set the information needed to recover the object's geons was intact whereas in the second set this information was removed. He found that subjects could recognise degraded objects more accurately when there was enough information to recover the geons. This result was found to be independent of object occlusion. Biederman argued that an underlying principle of recognising objects by their

components can account for his findings.

Biederman also used a priming technique to test his theory (Biederman and Cooper, 1991; Biederman and Gerhardstein, 1992). Biederman and Cooper examined whether priming effects in object recognition were due to the prior presentation of an objects features (i.e. edges), the object model (i.e. what the object is) or the objects components or parts. The subjects initially saw a set of familiar objects with either every second feature from each part removed or half of the components (or parts) removed. Biederman and Cooper then tested the effects on recognition by priming the subjects with either the identical image they saw in the previous block, the complement image or a different exemplar of the target object. They found that recognition was slower and less accurate for objects primed by a different exemplar than both the other priming conditions. Performance was identical for objects primed with either half of their identical features or the other complementary features, which suggested that object priming effects were not due to a repetition of the objects features. However, objects primed by half of their identical components were faster to be recognised than objects primed by their complementary components. This finding suggested to Biederman that the visual priming of an object is through the activation of a representation of the object based on the object's components and their spatial relations.

In another priming study, Biederman and Gerhardstein (1992) found that the time to name familiar objects previously primed by the same object were not affected by the orientation in depth of the priming object. There was an advantage for objects primed with the same object over objects primed with the same named but visually different object indicating that the effects observed were due to visual rather than semantic priming. Biederman argued that the results indicate that there is no difference in the recognition of objects in different orientations when the geon structural descriptions between two images of the same object are the same. He repeated these findings for a set of nonsense objects that conformed to the recognition–by–components (RBC) constraints on the nature of object representations in memory. In other words, priming between two different orientations of unfamiliar objects was the same as priming between identical orientations of the objects only if the orientations allowed the objects to be readily partitioned into geons. Biederman argued that the results found in previous studies which reported that recognition times are slower for objects primed with a different view of the same object than objects primed with an identical view (e.g. Bartram, 1974) was because the same 'geon structural description' was not readily available in the depth rotated views. He argued, therefore, that different views of objects are equally recognisable provided that views yielded the same geon structural description. A comparison of this approach to object recognition and other current approaches is given below.

The notion that recognition is facilitated when regions of sharp concavity are left intact in a fragmented image, because the geons are more recoverable from such an image, than one where the regions of concavity are removed (Biederman, 1987) can nevertheless be questioned. In a careful examination of the effects of priming with fragmented line drawings on the recognition of objects, Snodgrass and Feenan (1990) found that priming is optimum when

just enough information is available in the image to support perceptual closure (i.e. the Gestalt principle of filling in gaps in contours so as to perceive the most meaningful forms from the image). Biederman's findings reported above (Biederman, 1987) may have been affected by this perceptual closure hypothesis. In other words, the full object may be more readily perceived from a fragmented image that includes regions of sharp concavity because perceptual closure creates a more veridical representation of the object and not because the object is represented as geons.

Also, the notion that geons or volumetric primitives are the primitive features of objects has not always been supported. Using a visual search paradigm, Brown, Weisstein and May (1992) predicted that if geons are simple then they should pop-out of a visual array. They found no evidence that these geons were processed preattentively which should have been expected if they were primitive features. In fact, conditions where pop-out was exhibited could be explained by the difference in the 2-D features between the two volumetric primitives used in the search task.

Other recent experimental evidence on the effects of orientation on recognising shapes has supported the notion of their being two independent representations in memory; a view–centred description and an object–centred description. Ellis, Allport, Humphreys and Collis (1989) investigated the effects of exposure times on matching successive stimuli. They compared three types of matching; objects that were identical, rotated identical objects and objects that shared the same name. They found that for a stimulus onset asynchrony (SOA) of 200 milliseconds or more, the time to match two identical objects, even when rotated, was reduced and that matching times were fastest for objects that were identical. These effects were apparent for different sizes of the images and for different locations. For shorter SOAs (100 milliseconds) the benefit found in matching two rotated, identical objects was reduced but matching identical objects was unaffected by shorter SOAs. In this condition, size changes affected the identical views condition but not the rotated views condition. They argued that differential effects between identical objects and rotated objects with different SOAs supports the notion of the existence of two separate visual codes; one non-retinotopic but view–dependent code and the other object–centred and that building an object–centred description is a slower process than the construction of a view–centred description.

However, the results reported by Ellis et al. (1989) are open to other interpretations. It may be the case that the longer the SOA the more familiar the object becomes and that orientation becomes broadly tuned around the familiar view (Koriat and Norman, 1985). The results may also suggest that objects are represented through a set of multiple, view-centred representations and not a single object centred representation: The long SOAs increase the time available to access the representations of these familiar objects in memory and therefore no effects on recognition times for slight deviations in orientation will be observed if the object is already represented in the view observed. Jolicoeur (1985) also argued that novel depictions of a known set of objects may initially need to be transformed (which takes time) to

match a stored view in order to be recognised. Familiarity of these objects, however, diminishes the time needed to recognise them in different orientations. Section 2.4 below includes a discussion of the effects of familiarity on recognition.

2.3 Evidence for View-Centred Approaches

In contrast to the object-centred theories, more recent developments in the area of object recognition have proposed that recognition is a process where the retinal image is matched to a stored representation of that object through a process of alignment (Ullman, 1989; Tarr and Pinker, 1989; Jolicoeur, 1992). Representations of objects are view-centred (or view-dependent) in that the speed of recognition of an object depends on the orientation of that object. Fastest recognition times therefore occur to objects that directly match the orientation of the representation in memory.

The evidence cited for the view-centred approach differs in the number of views that are found to represent each object. For example, some studies have found that recognition times are fastest to one single view of an object (Palmer et al.,1981; Jolicoeur, 1985) whereas others have found that recognition times are fastest to a number of views (Tarr and Pinker, 1989; Edelman and Bülthoff, 1990; Cutzu and Edelman, 1992). A common finding across studies in view-specific effects was that the familiarity of the view plays an important role in determining the representation of the object, in that, the most familiar views of the objects are the views most likely to be represented. This section reviews the experimental evidence documented for the view-centred approach to recognition.

2.3.1 *Single Views as Representations*

Palmer, Rosch and Chase (1981) proposed that objects are perceived relative to a single view of an object which they termed the 'canonical' view. They defined the term 'canonical' as the view that maximises the salient information about the object. In their study, subjects were asked to rate how good or how typical a presented photograph of an object was of that object. There were 12 different photographs of each object, each corresponding to a different view of the object. They found high degrees of consistency between subjects when asked to report the best photographed view of each object. In a second task, subjects were asked to report the amounts of different surfaces visible in an imagined view of an object. An imagined view referred to the view that was most easily called to mind after experience with all the views. Results were found to correlate highly with the amount of surfaces visible in the best photographed view of the objects. Subjects were also required to photograph the view of the object that was imagined in the imagery task. The subjects' reported 'best' views were found to be highly consistent across tasks. In a subsequent naming time experiment, recognition times to different views of objects corresponded to the subjective reports of the most imagined view etc.. The results conformed to a benefit for a $^3/_4$ view of

most objects and recognition times were found to be a monotonically decreasing function of angle of rotation from this 3/4 view. A 3/4 is the view where the principal axis of the object lies at 45° to the line of sight. They termed this 3/4 view the canonical view and it corresponded to the view that maximises the amount of salient information of the object such as the visibility of information about the object (Palmer et al., 1981). Figure 6 illustrates the canonical view of the set of objects used by Palmer et al. in their studies. From their results they argue that "peoples concepts of objects contain at least implicit aspects of perspective". However, they do not commit themselves further on the nature of the representations in memory but suggest that representations of a particular view followed by some sort of transformation are more easily reconciled with the close relationship between object recognition and imagery.

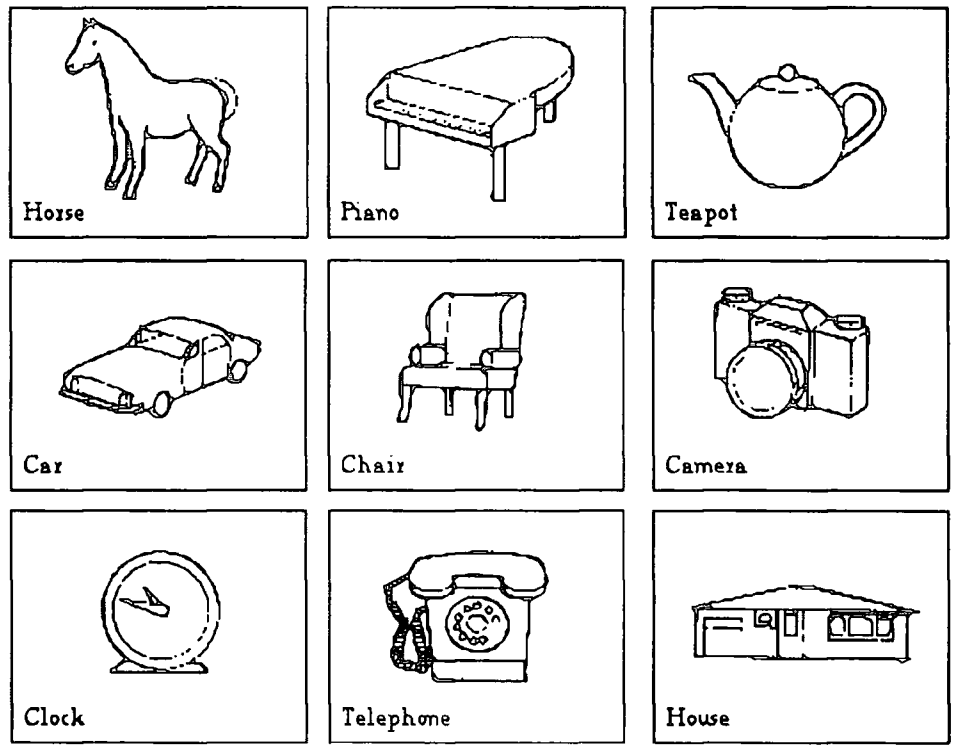

Figure 6; An illustration of the canonical views of a set of objects which Palmer et al. argued were the views which maximised the salient information the objects. A 3/4 view was found to be the canonical view of most objects e.g. the horse, piano, car, chair and camera.

Palmer's model successfully solves the problem of storage capacity encountered by the early template approach because only one view per object is stored and each new input is transformed and matched to this canonical view. However, this approach makes high computational demands on the visual system to achieve normalisation. Also there does not seem to be any underlying rules as to how the visual system decides what is the best or

canonical aspect of an object to store. These rules seem to be object-specific and are therefore not helpful in highlighting the default strategy used by the visual system in storing views of objects.

It could be argued therefore that the notion of a canonical view is not a useful one because it is difficult to define a canonical view in a general way. It also does not seem to offer a solution to the problem of variability across shapes as examples of the same object. According to this approach, each new shape or example of an object would have to be stored as a separate representation.

Others have found a facilitation effect in recognising disorientated objects when those objects are shown in the upright view (Jolicoeur, 1985; Diamond and Carey, 1986 and Yin, 1969) suggesting that the upright may be the canonical view of many objects. Jolicoeur (1985) investigated the effects on recognition times of objects shown in different orientations in the picture plane. He used both water–coloured drawings and line drawings of common, natural objects in different categories such as furniture, clothing and cars. He found that subjects' reaction times to name disoriented objects initially increased monotonically as the orientation increased from the upright. He claimed this provides evidence against models that involve the extraction of invariant properties.

However, the orientation effects diminished with practice and were found to be more pronounced with unfamiliar stimuli which suggests that familiarity can attenuate the effects of orientation on object recognition and consequently diminish the effect of canonicality. The diminished orientation effects due to practice on a certain set of objects did not transfer to another set of familiar objects indicating that general practice on mental rotation does not account for the effects observed (see Koriat and Norman, 1985). He argued that his findings resemble the effects observed from mental rotation experiments (Shepard and Metzler, 1971) and concluded that a process of mental rotation is involved before a novel view of a previously known object is recognised. However, Jolicoeur found that practice produces non-linear effects on naming time (objects rotated 180° away from the upright are recognised faster than 120° away from the upright). He argued that mental rotation effects do not explain these non-linear effects, nor the diminished orientation effects that occur with practice and that some other process may serve to explain these findings. The notion of a single canonical view however, also fails to account for the diminished effect of orientation with practice.

The literature on single views as representations fails to give adequate evidence that a single view is stored as a representation. There is a discrepancy in the literature on single view accounts, in that, practice effects the recognition of objects in different orientations such that it reduces the initial facilitation effect observed for a single orientation. Although most of the studies reported have looked at orientation in the picture plane, it would seem that the single view privilege reduces with familiarity. This may be

due to an increase in the number of stored views of the objects and that these stored views reflect the most familiar views of the objects found in the environment.

2.3.2 *Multiple Views as Representations*

In general, theories that assume multiple, view-dependent representations make predictions about the time required to recognise objects in different orientations. For example, if a specific instance of recognition involves a transformation, then provided that the transformation does not overlap with other stages, the process must take longer than when no transformation is required. Conversely, if instead of having a single, canonical view per object stored in memory, a set of different views of objects are stored, then the time taken to recognise objects would therefore be fastest for orientations that directly match these stored views. Objects that are disorientated from the stored views would require a transformation in order to align it with the stored orientation. This makes the prediction that the time taken to match a disoriented object would be directly proportional to the angular disparity between the nearest stored view and the inputted view.

Tarr and Pinker (1989) presented experimental evidence to support such a model. Subjects were initially trained on a set of novel, paper-clip type objects shown in a single orientation. The subjects were allowed to study these objects shown in a particular view and they were given extensive practice at naming and discriminating between the objects. Tarr and Pinker found that response times increased with increasing orientation away from the practice views. With practice however, all views were equally recognisable. At this stage the subjects were probed with a novel orientation of the object. A large differential effect between the time to recognise trained views and novel views was found. Their findings prompted them to suggest that novel orientations of objects are rotated to match the nearest stored view. Views chosen to represent novel shapes in memory were affected by the initial views presented to the subjects and the familiarity of these views. Their results suggested that the normalisation of novel views occurs by mentally rotating the novel views to match the nearest stored view.

Tarr and Pinker (1989) suggest that the number of stored views per object is limited and that the visual system can take the shortest path of rotation that will align the input to its counterpart. They argued that his process might be the well known mental rotation process described by Shepard (Shepard and Metzler, 1971 and Shepard and Cooper, 1982) on the basis of experimental studies (see section 2.3.3 below).

Further evidence that objects are represented as a collection of multiple views was reported by Edelman et al. (1989, 1990). Edelman, Bülthoff and Weinshall (1989) trained subjects on a set of views of novel, wire-frame objects. They then tested the subjects on the recognition of these views by presenting them with trials which either displayed a view of the target object or a distractor object. The subject was asked to indicate whether the object

viewed was the current target object. The dependent variables were reaction times and error rates. They found that recognition times varied with view-point of the objects and that a number of different views were found to minimise recognition times. These views however, did not correlate with the views chosen by the subjects as the 'best' views of the objects (see Palmer et al., 1981). They also found that the initial facilitation effect for a number of different views decreased with practice. Their findings were independent of stimulus complexity and familiarity. Their results lend support to the idea that the representations of 3–dimensional objects are view-specific and that more than one view is stored to represent an object.

In a subsequent study, Edelman and Bülthoff (1990) confirmed their previous conclusions that objects are stored as view–specific representations as opposed to object–centred representations and that representations include at most, partial depth information. Subjects were initially trained on a limited number of views of novel objects shown in motion ±13° around a reference view in order to give an impression of their 3–dimensional structure. All of the subjects reported perceiving the stimuli as 3-dimensional objects. The subjects were then tested on the recognition of static versions of the trained views. The authors reported finding a difference in the recognition rates between these familiar views. This result suggests that the emergence of canonical views cannot be attributed solely to the number of times the views have previously been seen. However, practice causes the effect that all views are equally recognised. Subject's recognition times to novel views of the objects were then tested. Edelman and Bülthoff found that recognition accuracy decreased with orientation away from the familiar or previously trained views. The authors concluded that objects are represented by a number of 2–dimensional views of the objects and that novel views are interpolated between these views. The evidence that views are interpolated rather than mentally rotated is discussed below (see section 2.3.4).

Tarr and Pinker's (1989) study concentrated on 2-D novel stick figures and orientations in the picture-plane or 180° flips in the depth plane. The information available in the orientations remained constant and it would be interesting to know how orientations in depth where information about 3-D objects is reduced, affect the results. Jolicoeur's (1985) results are consistent with those found by Tarr and Pinker. However, Jolicoeurs orientations were again only in the picture plane where information about the object remained constant. What is difficult to tell is whether these results can generalise to 3-D familiar objects and to gradual rotations in depth.

Biederman and Gerhardstein (1992) argued that the type of stimuli used in the Tarr and Pinker experiments are peculiar in that their representations are not sufficiently unique in order to identify between the different shapes. Biederman and Gerhardstein suggest that although the shapes used can be readily partitioned into geons but that the ultimate representations differ by way of a highly complex descriptor and that the entire set of shapes could activate the same representation in memory. These data therefore do not offer a

challenge to the RBC theory because the stimuli do not meet the criteria necessary to create view-point independent representations. They therefore argue that representations where there is no need to rotate the input is a more apt model of the visual system than storing multiple representations and transforming novel inputs. Biederman argues that representations are invariant over viewpoint only if the object can be readily partitioned into its component geons and that the representation is sufficiently discrete to avoid confusion in identifying the object. He provides experimental evidence for this (see Object–Centred Approaches above). However, in some orientations the object is more difficult to partition due to occlusion or accretion of the parts which makes it more difficult to access the correct representation. He does not explicitly explain how objects are recognised despite the different effects on the parts due to orientation. It may well be that some sort of transformation process is needed in order to correctly identify the object.

The evidence supporting the different transformation processes implied by the studies reporting object representations as view–specific descriptions of objects (Jolicoeur, 1985; Tarr and Pinker, 1989; Edelman et al., 1989, 1990) is outlined below.

2.3.3 *Evidence for Simple Transformations*

Mental rotation is often used to explain the transformation process involved in matching simultaneous shapes (Shepard and Metzler, 1971) or in naming disoriented, familiar shapes (Jolicoeur, 1985; Cooper and Shepard, 1973; Koriat and Norman, 1985). In these studies, recognition times or matching times are found to be a linear function of the amount of orientation away from a reference point, such as the orientation of another shape or the upright orientation. One of the earliest accounts of mental rotation was reported by Shepard and Metzler (1971).

When subjects were asked to compare two rotated shapes, Shepard and Metzler (1971) found that subjects reaction times increased with increasing difference between the angle of the shapes. Figure 7 below illustrates the task that the subjects were required to perform in the Shepard and Metzler experiment.

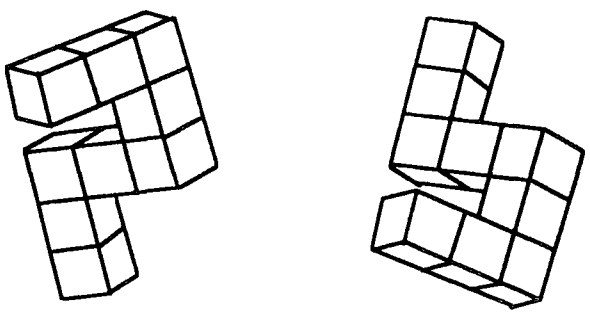

Figure 7; An example of a trial in which subjects had to decide whether two shapes were the rotated versions of the same shape or mirror-images of each other (after Shepard and Metzler, 1971).

They argued that the reaction times reflected the time taken to mentally rotate one shape in order to directly match it with the other and that more time was needed to mentally rotate the shapes as the difference between their orientations becomes greater. They found that shapes were mentally rotated at a speed of about 55° per second and similar rotation rates were found for rotations in depth and rotations in the picture plane. Shepard suggested that mental rotation may well be the process used by the visual system to match input views with stored views of objects (Shepard and Cooper, 1982). However, visual recognition seems to operate more rapidly than the speed of mental rotation processes.

Koriat and Norman (1985) on the other hand found that overlearning of one single orientation causes broad orientation tuning (see Shepard and Hurwitz, 1985 for discussion). Subjects were asked to respond as fast as possible to whether orientated letters were normal or reflected. The orientations were in increments of 60° in the picture plane. Instead of finding a linear function of reaction times to orientations from the upright for normal letters, they found a more curvilinear trend. The linear function was present for the reflected letters. They argued that the curvilinear trend found for normal letters could be due to one of three things; 1) that orientation becomes broadly tuned as a result of the way the stimuli are normally encountered in the environment and familiar stimuli are often encountered tilted from the upright and is therefore a consequence of familiar orientations in the environment, 2) that practice with upright stimuli automatically results in broad tuning and that it is therefore a characteristic of the visual system or 3) that practice with mentally rotating stimuli results in broad tuning. They found an increase in the curvilinear trend with the number of experimental blocks. Although Cooper and Shepard (1973) found non-linear effects with rotated letters, Shepard and Metzler (1971) found strong linear effects with rotated 3-D novel objects.

These effects may indeed reflect the different types of tasks that the subjects were required to perform. In the Cooper and Shepard (1973) task, as in the Koriat and Norman task, the subjects had to recognise single disoriented letters which would involve a matching process with a representation in memory. However, the Shepard and Metzler task was different in that the subjects had to compare two simultaneously presented objects that were rotated away from each other. This task did not involve matching to a stored description of the shapes. The different results found across these different tasks may reflect the differences between them. Jolicoeur (1985) also argued that the reduction in the orientation effect around the upright requires less time to take effect if the stimuli used are relatively simple. Indeed, Kubovy and Podgorny (1981) argued that the pattern of data found for experiments with a single oriented familiar letter would break down in simultaneous matching conditions. The curvilinear trend may then be a consequence of the nature of representations in memory. However, these data do not indicate whether the curvilinear effects are due to general practice with mental rotation or whether familiarity automatically increases the number of representations that are stored in the visual system.

In sum, work on mental rotation strongly suggests that a process of normalisation is used to match either simultaneously presented stimuli (Shepard and Metzler, 1971, and Larsen, 1985) or successively presented stimuli (Cooper and Shepard, 1973 and Koriat and Norman, 1985).

2.3.4 *Evidence for Interpolation Methods*

The view interpolation model has received support from both psychophysical and computational studies. From their findings on the effects of recognising novel views of objects, (see section 2.3.2) Edelman and Bülthoff (1990) argued that the effects observed in matching a novel view to a stored view could not be explained by a mental rotation transformation or an alignment process (Ullman, 1989) but that an explanation in terms of the interpolation approach would seem to be more parsimonious. In their investigation, subjects showed great difficulty in generalising to novel views of objects the further these views were to the original, trained views. Edelman and Bülthoff argued that the recognition rates seemed to be linked to the 2-D deformations of the images rather than the distance between novel and trained views. In other words, the deformations of the features in the 2–dimensional image affect delays in the recognition times from the familiar views of the object and not the distance between the orientation of the novel view and the stored view. Recognition of different views of objects proceeds by measuring the 2–dimensional, Euclidean distances between the features of the inputted image and the stored representations. A computer model of the recognition process based on non–linear interpolation between stored views (Edelman and Weinshall, 1991) simulated exactly the results found from the psychophysical study reported by Edelman and Bülthoff.

The conclusion that novel views of objects are interpolated between the stored representations has recently received further support from psychophysical studies. Cutzu and Edelman (1992) postulated that the recognition of objects is done by comparing the sum of the Euclidean distances between the image-based features from the input view to the stored views. This would predict the effect that the recognition time of a disoriented object would be positively correlated with the summed distances between the features of the input and the best stored view. Instead of comparing with the nearest stored view, the visual system takes the best view from among a 'cluster' of stored representations of the object. Matching occurs when the feature-space distance between the input and the stored view is minimised. By measuring the reaction times to a group of tube-like objects, Cutzu and Edelman found no evidence for a single, canonical view. Secondly, they found that reaction times were not linearly dependent on the distance between a novel view and a stored view as predicted by Tarr and Pinker's model. Instead, reaction times were correlated with the summed feature distances between the novel view and the best (shortest reaction time) stored views. They argue that linear transformation models suggested by Tarr and Pinker, (1989) and Ullman, (1989) are not representative of the workings of the visual system and they propose that a model measuring non-linear deformations between features is a more appropriate model of the

human visual system.

However, for real objects this still leaves the question of how the relevant features of a 3–dimensional common object are extracted. It is not clear what constitutes a relevant feature in common objects which makes the view interpolation model difficult to apply to everyday objects. It could be suggested that the ideas proposed by the partition theorists (Hoffman and Richards, 1985) may suggest ways of determining these features. Alternatively, experiments measuring the effect on recognition time of partially occluded objects may give hints as to the features of the objects.

2.4 The Effect of Familiarity of Viewpoint on Recognition

The significant point to be made about object–centred theories discussed above is that they do not assume any level of familiarity with the object before a representation can be made. Representations can be created from a single view of the 3–dimensional object. These theories are therefore called object-centred theories because representations in memory are independent of the view of the object observed.

On the other hand, view–centred approaches specify that representations are view–dependent and are not invariant over viewing position. One of the assumptions of the view dependent approaches is that representations are built around the familiar views of the objects. In other words the familiarity of the view determines the representation of the object. This means that no preferred view should exist for familiar objects that are equally likely to be seen in any orientation. Indeed it does seem to be the case that orientation effects diminish with practice with a variety of 2-D stimuli such as line-drawings of common objects (Jolicoeur, 1985), nonsense characters (Koriat and Norman, 1985) stick-like objects (Tarr and Pinker, 1989) and wire objects (Rock and DiVita, 1987 and Edelman, Bülthoff and Weinshall, 1989).

Rock and DiVita (1987), for example, found that objects were recognised fastest when the projected retinal image of the object remained the same as that seen in a training block despite changes in the position of the object. Subjects were initially trained on a single view of a set of 3-dimensional, wire-frame objects shown in a nearby position to the viewer. In the test stage, subjects recognition times were tested to different displacements of the objects. The objects were either shown in the same position as in the training block, displaced laterally to upper left or right and to the lower left or right, or displaced and rotated so that the projected image of the object was the same as that shown in the training stage. Rock and DiVita found that recognition times were fastest to the test condition where the projected image of the object was the same as that viewed in the training block. They concluded that the recognition of novel, 3–dimensional objects is viewer-centred and that recognition is facilitated by the most familiar view of the object.

Studies on face recognition have found that the recognition of inverted faces is very difficult (Yin, 1969, Diamond and Carey, 1986). Yin compared the effects on recognition times of oriented faces to other classes of familiar objects. He found that more errors were made in general to objects that were oriented away from their normal upright position but that faces were disproportionally affected. He argued that the familiar orientation of faces, i.e. the upright orientation, constrains the representation and this view is therefore the most recognisable view. He also suggests that the differential effects found between recognition of faces and other mono-oriented objects may be due to a special factor related only to faces. This factor may be the nature of the representation of faces. Young, Hellawell and Hay (1987) argued that faces are represented through the configural information between the features. This specific type of representation may explain the lack of generalisability to different orientations of faces over other types of objects that are not represented in terms of the configural information. However, Harries et al. (1989) found that faces had two preferred view-points for inspection purposes when the face is rotated in the Y plane, i.e. around the vertical axis. These views correspond to a view close to profile and a full frontal view of the face. Their results suggests that faces are not represented by a single representation but that multiple representations of each face in different orientations in depth may be stored. Diamond and Carey (1986) demonstrated that faces are not a specific category of representations and that any class of objects that are over learned in one orientation cause the same lack of generalisability to different orientations (e.g. dogs). They asserted that expertise in one orientation causes constraints on the nature of the representation of that object.

Larsen (1985) investigated the effect of familiarity on mental rotation across different orientations and sizes of 2-dimensional shapes. He looked at the effects on mental rotation of unfamiliar stimuli, i.e. he presented a new pair of rotated stimuli in every trial. The results showed strong linear effects with increasing angular difference between the two stimuli (he also found linear effects with increasing size difference). His findings supported the conclusions asserted above that familiarity with the stimuli causes differential effects on mental rotation times (see Koriat and Norman, 1985 and Kubovy and Podgorny, 1981).

In sum, familiarity with the stimuli reduces the linear effect on matching times suggesting that some other process is also in operation (Cooper and Shepard, 1973 and Koriat and Norman, 1985). It could be argued that as different orientations become familiar, then these orientations may be used as representations to which novel orientations are matched. The number of representations therefore increases with familiarity of the stimulus (Tarr and Pinker, 1989).

Studies on mental rotation effects also suggest that representations of familiar stimuli are view dependent and not invariant to view point. This work therefore supports the interpretations of the proponents of the alignment approach (Tarr and Pinker, 1989, Jolicoeur, 1985, Ullman, 1989).

2.5 Principal Axes in Object Recognition

The question of how the visual system decides which transformation or function to apply to the input must be raised. Tarr and Pinker (1989) argued that certain characteristics of the image are extracted and that transformation processes are applied without recognition of the object. As was already argued, these characteristics may be object specific. A potential candidate for an early extracted characteristic of familiar objects is the principal axis of that object.

Marr (1982) argued that the function of the recognition system was to create 3–D object models of objects based on the principal axis of the objects. These axes could include the elongated axis or the axis of symmetry of the object. This axis is initially extracted from the information given in the retinal images of the occluding contours of the object (Marr and Nishihara, 1978). Although Marr did not give a detailed account on how the information extracted from the retinal image to create the $2^1/_2$–D sketch could be used to develop a 3–D model of that object, his ideas that the principal axis of objects is important for the purposes of representation still hold.

Quinlan (1991) studied the effects of different principal axes on the speed of recognition. He argued that Marr's explanation of what constitutes a principal axis is vague and misleading. He argued that if two axes were pitted against each other in the same object e.g. if the elongated axis was perpendicular to the axis of symmetry, then this would cause problems for Marr's model. How would the visual system decide which axis is the more salient? Using novel, 2–D line drawings of various shapes, he looked at the effects of these two types of axes pitted against each other on response times to a match/mismatch design and found that, for recognition purposes, when the elongated axis is conjoined with the axis of symmetry then this axis is more salient than when they are disjoint. He also found that when the principal axis was explicitly included in the drawing of the shape that this caused a decrease in recognition times relative to other stimuli which had either a minor axis, the gravitational upright axis or the horizontal axis included. Finally he found that by comparing the recognition effects to disoriented shapes with elongated axes and disoriented shapes with conjoint axes (i.e elongated axis equal to symmetrical axis) recognition times were shorter when the conjoint axes were vertical but not when the elongated axis was vertical. He argues that the vertical axis may be important for some shapes but not for others. These findings do not fit in with Marr's axis-based account.

There has been some suggestion that orientation causes a differential effect on recognition times not because the principal axis is rotated and resolving it becomes more difficult with orientations away from a standard view, but that the positions of the focal features such as the top and bottom of the object changes and they need to be re-aligned (Rock, DiVita and Barbeito, 1981). However, Humphreys (1984) examined the time taken to judge whether two rotated elongated shapes presented sequentially had the same structure and

found that the time taken to match the shapes was affected by the orientation of the principal axis of the shape rather than the position of the focal features. When the subjects could predict the location of the stimulus however, this effect was reversed and matching times depended on the location of the focal features.

Other theorists have also considered that principal axes are coincidental with the representation and that the visual system does not rely on them in order to create a representation of the object (Lowe, 1985). Others still have argued that Marr's idea on axis–based descriptions is correct and that they are fundamentally important to the representation (Pentland, 1986). Pentland claims that complex natural images are made up of a set of superquadratic components rather like Marr's' generalised cones or Biedermans' geons. However, these components preserve their smoothness and do not rely on concavities but are nevertheless segmented. The components are derived from a natural scene from information about the surface tilt. This tilt is a consequence of the orientation of the major axis of each component part.

The issue of the role of principal axes in the representation of objects is still unresolved. It seems from the literature that principal axes may be important for the description of some shapes and not for others. It is also difficult to apply findings from rotated unfamiliar, 2-D stimuli to 3-D common objects. Perhaps it is the case that for common objects, the principal axis is an important basis of the description of the object in memory. Indeed it can be observed that most natural objects have a major axis of symmetry which is usually coincidental with an axis of elongation in that they are the same axis. Further research on the differential effects of principal axes in familiar objects is therefore needed.

Following from research into the effects of principal axes, many researchers have argued that it is not the objects intrinsic axis that is transformed to match the stored representation but that a description of the object's image is built relative to a reference frame. There are a number of reference frames to which the object can be described. A discussion on the evidence for the different reference frames is included below.

2.5.1 Reference Frames

A recent interest in reference frames has developed from computational approaches to object recognition. Marr and Nishihara (1978) proposed that an objects intrinsic reference frame is resolved from the $2^1/_2$-D sketch, or viewer-centred description in order to create an object–centred description. Objects are therefore recognised by describing the view-centred image in terms of co-ordinates that are relative to the objects principal axis or intrinsic reference frame. The notion of intrinsic reference frames is central to theories of object recognition that use structural descriptions as representations. According to these models, the object properties must be described in some way relative to the object itself in order for recognition to occur. One of the most popular methods of describing an object's properties is

relative to a salient axis or reference frame (Marr and Nishihara, 1978; Marr, 1982; Hinton and Parsons, 1981).

The problem with the notion of reference frames is that it is difficult to determine a salient axis of a lot of objects. Humphreys (1983) found that objects with ambiguous elongated axes such as squares and hexagons were more difficult to match when transformed than objects with salient principal axes such as isosceles triangles or elongated pentagons. He concluded that shapes with ambiguous elongated axes can change their structural descriptions depending on which axis is aligned with the gravitational upright. For example, a square is recognised as such only when the vertical is perpendicular to two edges but it is recognised as a diamond when the vertical intersects the junction of two edges. Thus it could be argued that objects with non-salient principal axes can use extrinsic reference frames as a basis of the structural descriptions. Other studies have also shown that the upright or vertical reference frame is important in describing an object. Many investigations have found that not only do people use the upright orientation as a primary reference direction to which objects are interpreted or mentally rotated (Jolicoeur, 1985; Diamond and Carey, 1986; Yin 1969 and Koriat and Norman, 1985) but that the upright also plays a role in how the surrounding environment is represented. Palmer et al. (1988) however argued that the effects observed in the Humphreys (1983) study were due to the fact that all of the shapes were viewed as coplanar. With 3-D depth cues, such as perspective, included in the display subjects find matching and identifying shapes easier than when the shapes are viewed without depth information.

Feldman (1985) postulated that there are at least 4 reference frames with which a shape can be described. A shape can be described relative to retinotopic co-ordinates, head position, a gravitational frame or an object centred frame. Retinal and head position based frames are said to create data driven or bottom-up reference frames whereas gravitational or object-centred frames are said to be conceptually driven frames (Corballis, 1988). Shepard and Hurwitz (1984) classify these frames as egocentric, object centred and environmental respectively. They argue that different mental rotation processes are specific to each reference frame. Both the Marr (1981) and the Hinton and Parsons (1981) models of object recognition postulate that the conceptually driven frames are the most important for object recognition and that these frames must be resolved in order for an object-centred representation of the object, which is invariant over view point, to be stored.

It has also been asserted that objects are recognised independently of a reference frame or any coordinate system (Corballis, 1988). For example, objects can be defined as a set of locally based features that constitute a representation of objects that is essentially orientation independent. Moreover, it has been postulated that intrinsic reference frames are not needed for recognition to occur and that view-centred information is sufficient for recognition (see alignment models in Chapter 2). Robertson et al. (1987) tested the notion of whether reference frames or images were transformed in mental rotation tasks. The subjects

were asked to respond whether successively presented letters were normal or mirror-image reflections. The investigators were interested in whether the subjects rotated the second letter presentation relative to the first presentation or relative to the upright. The results suggested that either of these transformations can operate and that a hybrid model of relative rotations and rotations to the upright best fits the data.

2.6 Future Directions in Object Recognition

The obvious gap in the literature on the effects of orientations on the recognition of objects is the lack of experimental work on 3-dimensional, familiar objects. The following three chapters hopes to address this gap by presenting evidence on the effects of orientation on the recognition of familiar objects such as a frying pan, a lamp, a bottle and a light bulb. Chapter 3 presents evidence on the recognition of common objects in different orientations in the X, Y and Z axes and in the picture plane. These effects are shown to be independent of the most familiar orientations that the objects are found in the environment and also independent of practice. Chapter 4 includes results from experiments where the effects of rotating silhouetted versions of the objects were measured. Results from an experiment where the orientations were primed are also included. Chapter 5 includes evidence that the effects observed with familiar objects can also be generated by training subjects on unfamiliar, 3-D objects.

Chapter Three

Effect of Orientation on Recognition

3.1 General Introduction

As was already mentioned in the previous chapter, there has been little evidence documented on the recognition of familiar, 3-dimensional objects in different orientations. This chapter presents experimental evidence on the recognition of familiar, elongated 3–dimensional objects shown in different orientations. Recent advances in computer graphics have allowed us to generate 3–dimensional drawings of objects and to present them on a screen in controlled orientations.

It has been argued that although the visual system may construct a representation based on the axis of the object, the stored representation of the object already in memory may not necessarily be a view-independent description as Marr (1982) suggested but may in fact be in some sort of view-dependent, canonical format (Palmer, Rosch and Chase 1981 and Humphreys and Riddoch 1984). If this were so one would expect to find a facilitation effect for the recognition of objects that conform to the characteristics of the canonical view in memory, including the orientation of the axis or the presence of salient features. Indeed Palmer found that subjects recognise the canonical view faster than other views. Palmer argued that this canonical view is the view that maximises the amount of salient information of the object and is therefore the most recognisable view. This view is object specific but it generally conforms to a $^3/_4$ view of the object. Humphreys and Riddoch (1984) also noted that their agnosic patients found it difficult to recognise objects from 'unusual' views but they were capable of recognising them in canonical or prototypical, views.

One problem with the Palmer et al. (1981) study lies in their non-arbitrary choice of objects as stimuli for the experiment. All of these objects have well-defined gravitational uprights and highly salient focal features. For example, the canonical view of an alarm clock (determined by rating study, the most imagined view, the view chosen to be photographed and the view most quickly recognised) is a 2–dimensional view which shows the face of the clock only, and this 2-dimensional, frontal view is also the canonical view of a house (see Figure 6, Chapter 2). For all other objects studied, the canonical view has preserved the depth information of the objects by depicting these objects so that the principal axes and the salient features are fully resolved. The fact that the canonical view is peculiar to each object suggests lack of efficiency in the visual system. It seems likely that the canonical view of an alarm clock maximises information about the face because that is its salient feature but there

must be instances where the visual system needs more global information about clocks e.g.. for discrimination between clocks and watches. Do they suggest that there are separate canonical views for different examples of objects within the same category and each of these canonical views are further divided according to the type of information needed for the task at hand? Other objects which are also recognised on the basis of more local or feature-based information are faces. These objects however, do not necessarily have a single canonical view but are represented as a set of characteristic views (Thomas et al, 1990). The studies of canonical views lack a comprehensive analysis of the different views one is likely to observe of any object and generally only a small selection of views are tested.

Is the canonical view merely the most encountered view of that object or the view that maximises the amount of visual information of that object? If the former then one would expect to find a facilitation for uprightness in objects that are more likely to be seen one single upright orientation of the principal axis. If the latter then more than one canonical view per object would emerge in each axis irrespective of uprightness. In other words, if a set objects were rotated in the three major axes X, Y and Z and their combinations, then each time the principal axis was fully exposed and the salient features were present a canonical view effect would be apparent.

In the Marr and Nishihara (1978) model the first step towards representing an object in object–centred coordinates is to assign a direction to the principal axis of the object and from there build a 3–D object model (see discussion in previous chapter). According to Marr, the process of representing an object in terms of a description based on the object's principal axis renders recognition invariant over different orientations of the object. However, the ease of assigning the direction of the principal axis depends on how much of the axis is available. When an object is rotated in depth it could be argued that an increase in the deviation between the orientation of the principal axis of the image and the standard orientation of the 3–D model would cause an increase in the difficulty of assigning the direction of the principal axis. For example, as an elongated object becomes more rotated in depth, the principal axis becomes more foreshortened. A description of the object's contours relative to the intrinsic axis of the object would therefore become more difficult to derive as the axis becomes more foreshortened. It could be argued therefore that this model predicts a linear effect in recognition times with orientations in depth away from a standard view or direction of the principal axis because an increase in recognition times would reflect the difficulty in deriving the information about the object's principal axis as the object becomes foreshortened.

Rock (1973) asserted that the phenomenal shape of an object changes with orientation in the picture plane. Why does this happen? The relative positions of the parts of the objects have not changed but their overall direction has. Like Marr, Rock also argued that recognition proceeds by assigning a direction to a disoriented object relative to a reference frame but he argued that the objects are described relative to the environmental upright and

deviations from the upright in the picture plane are more difficult to recognise because assigning the direction is more difficult. An object's orientation is therefore seen relative to the environment with vertical orientations recognised fastest. Rock therefore argued that changes of orientation in the picture plane have more of an effect on recognition than orientations in depth because the direction of the object relative to the environmental upright need not change with orientations in depth. Other studies have supported this view that objects are recognised relative to the environmental upright when rotated in the picture plane (Cooper and Shepard, 1973; Jolicoeur, 1985; Koriat and Norman, 1985). Humphreys (1984) demonstrated in a matching paradigm that changes in the orientation of the principal axis were more detrimental to matching than changes in the locations of the shapes' focal features (i.e.. top and bottom of the object). However, it has also been found that practice reduces the dependency on the environmental upright and objects are recognised equally fast in all orientations in the picture plane (Jolicoeur, 1985). In this case Rock argued that familiar stimuli carry their own intrinsic reference frames. Elongated objects have more salient intrinsic axes which makes the direction of the object more easy to detect.

3.1.1 Swivel 3D Package

An object-oriented drafting package was used to generate the stimuli for the experiments reported in this chapter and all of the experiments reported in this thesis. This package is called Swivel 3–D and is especially designed for use with an Apple Macintosh.

The package allows 3–dimensional objects to be designed by building on descriptions of the cross, top and side sections of the object. The number of facets on the surface of the object can be manipulated allowing more smoother object rendering. A specific advantage of this package is that once the object is drawn it can be shown in different orientations with respect to the viewer. Orientations can be manipulated in each of the axes (X, Y, and Z) independently in increments of 1 degree. A combination of orientations in the different axes can also be produced. The orientations of each of the objects drawn was therefore carefully controlled.

3.1.2 A Description of the Orientations Tested

A set of objects were drawn using the Swivel 3–D package which allowed the objects to be viewed from any angle. The objects were presented in increments of 30° in the three major axes of rotation; X, Y and Z and in their combinations; XY, XZ, YZ and XYZ. Figure 8 shows the major axes of rotation. All of the orientations are relative to the 0° position shown in the illustration below. This 0° position was the foreshortened view of the objects with the top of the object facing the viewer. Orientations in the Z axis were rotations around the object's axis and did not yield any extra information about the objects other that what was available at the 0° degree position. In the 0° degree position the principal axes of the objects are parallel

to and in the same direction as the line of sight, in other words the principal axis is completely foreshortened in this view. Other foreshortened orientations include the 180° orientations in the X and Y axes and are indicated in Figure 8. At 90° and 270° in both the X and Y axes and also all the combination axes (XY, XZ, YZ, and XYZ), the principal axis is viewed orthogonal to the line of sight and parallel to the picture plane. For rotations in the X, XY, and XZ axes the upright version of the objects was shown in the 270° orientation and the inverted version was shown in the 90° orientation. The orientations of the principal axis in the picture plane differs across axes in these views.

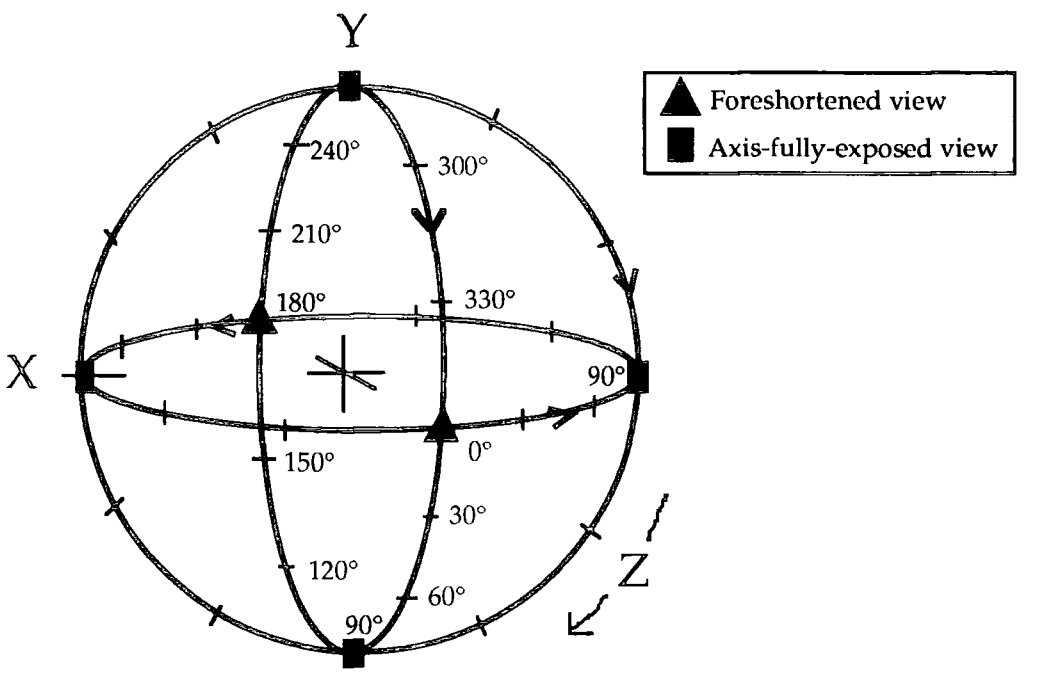

Figure 8: Illustration of the orientations used in all the experiments reported below. The objects were viewed relative to a 0° position which is indicated on the viewing sphere above. The top of the object was in the foreground in this position. The 90° and 180° positions in both the X and Y axes corresponded to the same view. The orientations are indicated where the objects were viewed with either their principal axes fully foreshortened or fully exposed.

Figure 9 below illustrates a selection of objects, taken from the initial three experiments reported in this chapter, which are shown rotated in the different axes of rotation. The objects illustrated are shown oriented in increments of 30° from left to right between the 0° orientation and the 180° orientation illustrated in Figure 8 above. The bottle is shown rotated in the X axis and the lamp is shown rotated in the Y axis. The other objects are shown rotated by a combination of axes; the glass is rotated by a combination the X and Y axes, the jug is rotated by in the XZ axes and the frying pan is rotated by all three axes, i.e.

the XYZ axes. The source of illumination was fixed with respect to gravity.

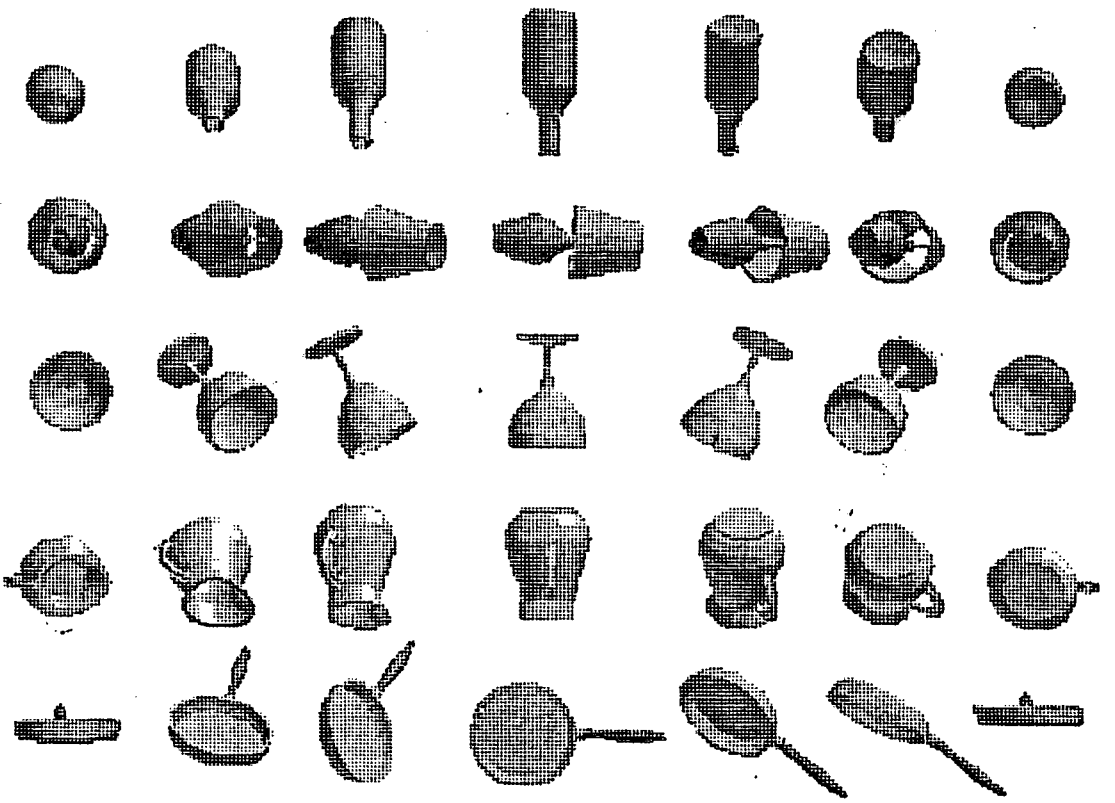

Figure 9: Illustration of a number of objects used throughout the thesis shown in orientations from 0° to 180° in a selection of axes of rotation.

3.2 Rating Study

A set of objects which were to be used as stimuli in the experiments reported in this chapter were rated for their typical orientation in the environment. A set of objects, taken from Snodgrass and Vanderwart (1981) were chosen based on their likelihood of being found in a single, upright orientation or in any orientation. This rating study was conducted in order to confirm that the choice of objects represented two separate sets, one a set of objects typically upright in the environment and the other a set of objects with non-typical orientations.

Subjects

Twenty–one members of the Department of Psychology, University of Durham participated in this study. Their ages ranged from 23 to 56 years.

Stimuli and Apparatus

A rating scale was developed in which a set of 15 objects were rated for 'uprightness' in the environment. These objects corresponded to the following; light bulb, bottle, chair, lamp, clothes peg, frying pan, screw, cricket bat, padlock, rolling pin, whistle, nail, glass, jug

and clock. Two extra objects (car and tennis racket) were given as examples on how to rate the objects. The scale ranged from 1 to 4. If the objects were likely to be found in any position in the environment then the object was rated as 1. On the other hand, if the object was typically found in one, upright orientation, the object was rated as 4. Ratings of 2 and 3 corresponded to 'not often found in one orientation' and 'mostly found in one orientation' respectively.

Design

The study was a rating study which was based on a Likert Scale of rating. A scale from 1 to 4 was used to measure objects on their typical uprightness in the environment.

Procedure

Subjects were asked to rate a set of 15 objects on typical uprightness in the environment. Two objects were included as example objects by the experimenter; A car was rated as 4 indicating that it was always found in one orientation and a tennis racket was rated as 1 indicating that it could be found in any orientation. Subjects were instructed to imagine each object and to decide whether the object was likely to be found in different orientations around the picture plane, or whether they were typically found in only one orientation, the upright orientation.

Results

Figure 10 below indicates the porportion of times the objects were rated on or above a score of 3 (mostly found in one orientation and always found in orientation). The objects are depicted from left to right as follows; light bulb, bottle, chair, lamp, clothes peg, frying pan, screw, cricket bat, padlock, rolling pin, whistle, nail, glass, jug, clock. It was found that seven of the objects were rated as being mostly found, or always found, in one orientation by more than 70% of the subjects. Conversely, the other eight objects were rated as having no typical orientation more than 50% of the time.

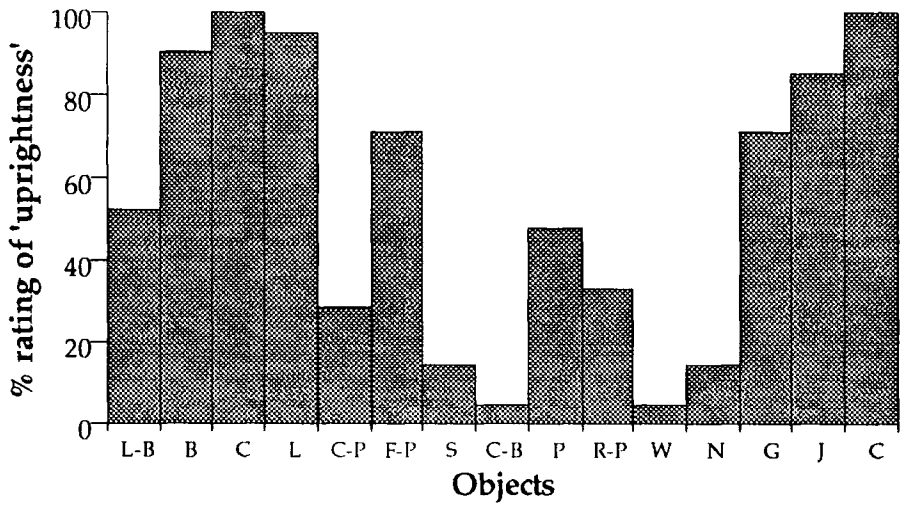

Figure 10: Percentage number of subjects who rated each object as being mostly or always found in an upright orientation.

Discussion

It was decided that objects rated higher than a score of '3' by more than 70% of the subjects was the criterion for determining that the object had a single, typical orientation. Therefore, bottle, chair, lamp, frying pan, glass, jug and clock are objects having a typical orientation whereas light bulb, clothes peg, screw, cricket bat, rolling pin, whistle, and nail are objects having no typical orientation. For the purposes of this thesis, only elongated objects were chosen as stimuli, therefore, the clock was not used as a stimulus.

3.3 Experiment 1

Experiment 1 was designed in order to determine the canonical view of a set of elongated objects which are typically found in a single, upright orientation. According to Palmer et al. (1981) canonical representation refers to a single, view-centred representation which maximises the amount of salient information about the object in question. Palmer's approach predicts a facilitation for the recognition of objects in a canonical view and also that deviations from the canonical view should affect recognition time with reaction times monotonically increasing with increasing deviations from the canonical view. This canonical view need not necessarily be the view that holds the maximum information about the principal axis but could instead be sensitive to the gravitational upright.

This experiment was run in order to test the notion of canonical or prototypical views as representations in memory. Six elongated objects with strong gravitational uprights were selected from among a larger set of objects that were rated by subjects for uprightness (see rating study above). These objects were judged by the subjects as having a typical orientation in the environment which corresponded to the upright orientation.

3.3.1 METHOD

Subjects

Nine Psychology postgraduate students from the University of Durham participated in this experiment. Five of these subjects were male and four female. The age range of all the subjects 21 to 49. All subjects had normal or corrected–to– normal vision.

Materials

A selection of 6 common objects was drawn on an Apple Macintosh IIx using the computer graphics package *Swivel 3D*. The objects drawn were as follows; frying pan, bottle, jug, glass, chair and lamp. A picture of each object was presented in a total of 78 views i.e. a 0° degree view and 11 rotations in increments of 30° around each of the X, Y, Z, XY, XZ, YZ and XYZ axes.

A stimulus was made up of an object view which was presented on a screen twice,

once with a label which matched the name of the object and once with a mismatched label taken from the names of the other objects in the experiment. Each stimulus was presented for 1 second with an inter-trial interval (ITI) of two seconds. The object was presented in the middle of the screen with a label shown above it. All stimuli were shown against a black background and the screen remained black between trials.

A hood was used to cover the Macintosh screen in order to control for different light conditions in the laboratory and also to ensure that every subject was equidistant from the screen (i.e. 57 cm). The visual angle between the object and the label subtended 5 degrees.

The onset of each stimulus triggered a timer in a BBC microcomputer through a photoreceptor which was attached to the bottom of the Macintosh screen. The photoreceptor was hidden from the subjects and did not interfere with the task. The offset of the timer was triggered by a response from the subject. A response was made by the subject depressing either the 'SAME' key or the 'DIFFERENT' key on a response box which the subject held in their hands. Left handed subjects were allowed to turn the box in order that the verification, or 'SAME', key was depressed by the dominant hand.

Design

The experiment was based on a 3–factor, repeated measures design. The main experimental factors were objects, axes of rotation and orientations. The experiment was based on a match/ mismatch design where subjects had to decide as quickly as possible whether a label shown with an object was the correct (or incorrect) label of the object shown. The reaction times to each trial were recorded although the analysis was conducted on the match trials only.

The objects factor contained six levels each corresponding to an object. There were 7 levels to the axes of rotation factor which included the three major axes X, Y, and Z and their combinations. Finally, there were eleven levels to the orientations factor, each level corresponding to orientations in 30° increments from 30° to 330° and the 0° orientation was shown once for each object to avoid learning across the foreshortened views.

To ensure that each subject became highly practised with the procedure, the nature of the drawings and the different orientations, each subject was initially presented with a practice block of trials which included all of the orientations with objects as a nested factor.

The order of the trials was randomised for each subject. The experiment was divided into nine experimental blocks and the experimental conditions were counter-balanced across blocks.

Procedure

The total number of trials was randomly divided into nine experimental blocks,

with 100 trials in the first eight blocks and 136 trials in block 9. Each block contained equal numbers of match and mismatch trials. The experiment was performed in 3 sessions, one per day and the subjects received three experimental blocks per session. Each session lasted approximately 20 minutes. Dummy trials were included at the beginning of each experimental block because a speed up effect over the first three or four slides may have occurred. The presentation of the blocks across subjects was counter balanced using the Latin square method.

Each subject was instructed to respond to a stimulus by either pressing the 'SAME' key if the label matched name of the object shown or the 'DIFFERENT' key on the response box if the label did not match the name of the object. The subjects were instructed that a response should be made as fast as possible to every trial without making too many errors. They were asked to attend to the screen at all times for the onset of each stimulus until the end of each block. The reaction times and errors were recorded and stored on the BBC microcomputer.

3.3.2 RESULTS

The errors made to the matched trials only were not more than 5.6%, with an average of 26 errors per subject in 468 match trials. There was no evidence found for a speed/accuracy trade off. Figure 11 shows the mean number of errors made to each orientation. The errors were not subjected to further analysis.

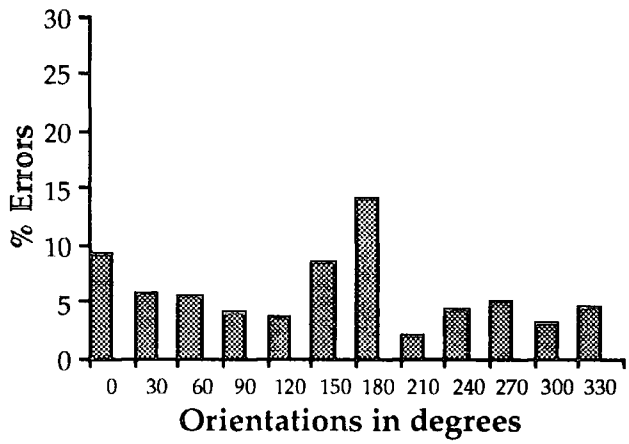

Figure 11: Percentage of errors made across objects shown in the different angles of orientation.

Figure 12 shows the mean reaction times to objects shown in the different orientations. The reaction times were collapsed over the objects and axes factors. A two-way, repeated measures ANOVA was conducted on the axes and orientations factors across all subjects.

A significant effect of orientation was found, $F(11,88)=7.635$, $p=0.0001$. A Newman-Keuls post-hoc analysis revealed that the 180° orientation was significantly different from

all other rotations at p≤0.01 level of significance and 0° orientation was different from both the 60° and the 120° orientations at p≤0.05 level of significance. Figure 12 below shows the mean reaction times across all of the objects in the different orientations.

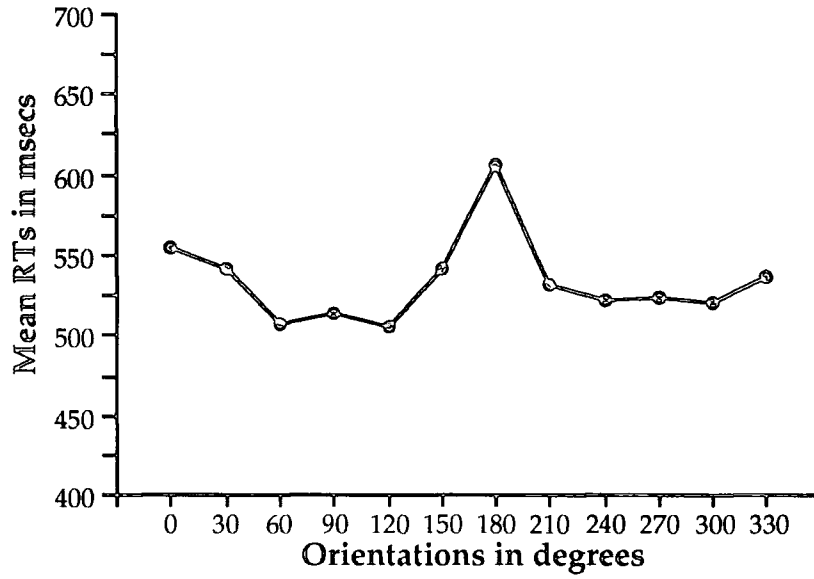

Figure 12; Subjects overall mean reaction times to the different orientations across all the objects viewed in Experiment 1.

A significant effect of axes was also found, F(6,48)=24.673, p=0.0001. Figure 13 shows the mean reaction times to objects shown in different orientations in each of the axes of rotation. A post hoc Newman-Keuls showed that at p<0.01 level of significance the Z axis was different to all other axes. No other differences were noted within these effects.

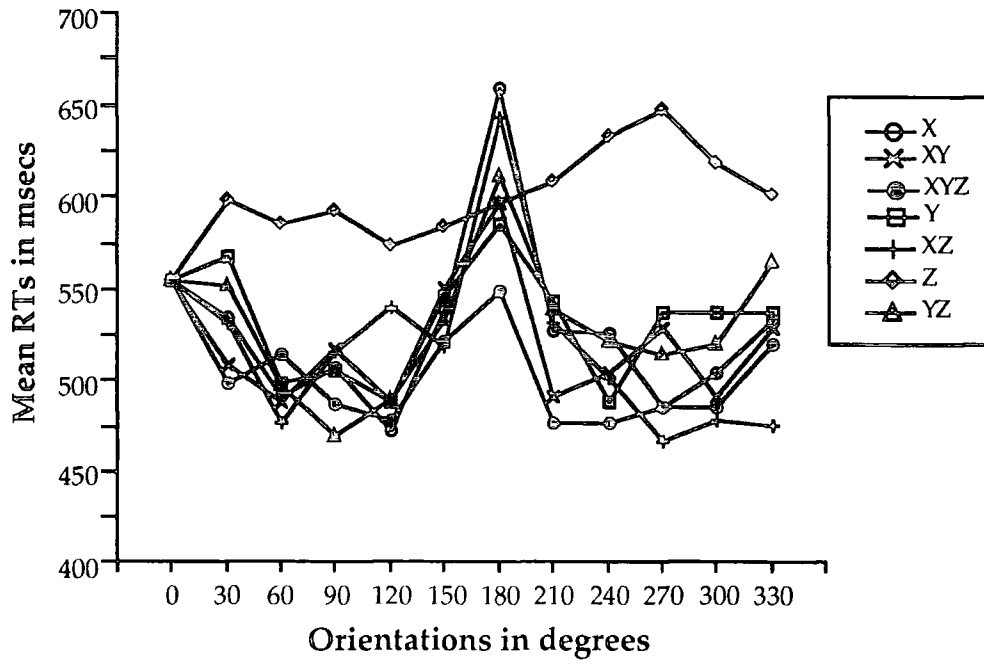

Figure 13; Subjects overall mean reaction times to the different orientations of the objects viewed in each of the axes of rotation in Experiment 1.

A significant interaction between axes and orientation was also found, F(66,528)=1.807, p=0.0002. This interaction was attributed to orientations in the Z axis.

The individual objects were each subjected to an ANOVA to determine the effect of rotation and to see if a canonical effect was apparent in each of the objects. The results of an analysis of variance across all subjects are listed below for each individual object (Figure 14).

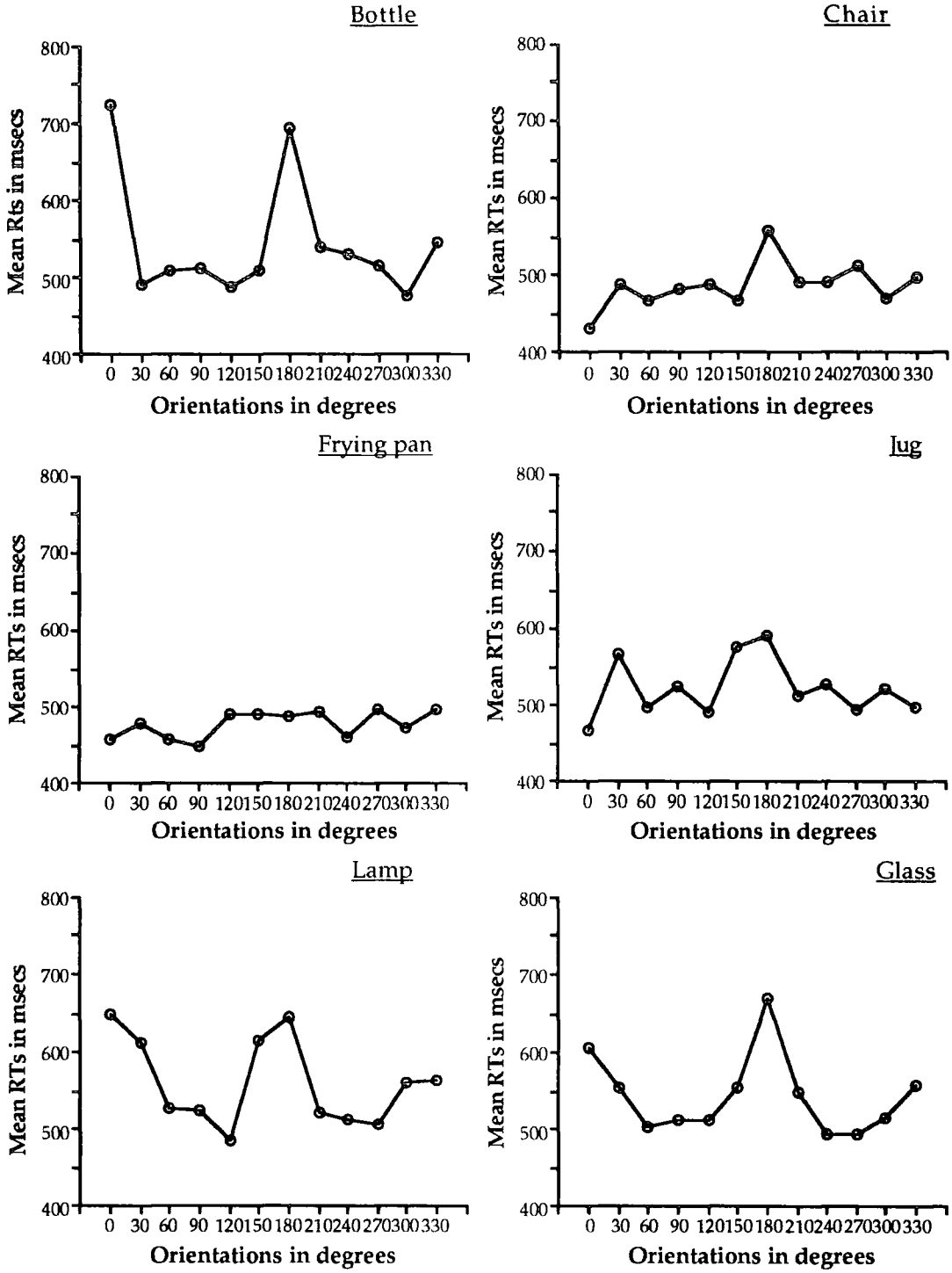

Figure 14: Individual plots of the mean reaction times to the different orientations of each object shown in Experiment 1.

For the bottle, a two-way ANOVA revealed significant effects of both axes, $F(6,48)=11.277$, $p=0.0000$ and orientations, $F(11,88)=6.464$, $p=0.0000$. For the chair, a two-way ANOVA showed a significant effect of orientation, $F(11,88)=2.848$, $p=0.0031$ but not of axes, $F(6,48)=1.451$, $p=0.2153$. An analysis of the data to the frying pan proved significant for axes, $F(6,48)=2.486$, $p=0.0356$ at $p<0.05$ level of significance and not significant for orientation. A two-way analysis of variance on the reaction times to the glass proved significant for orientations, $F(11,88)=2.823$, $p=0.0033$ and for axes, $F(6,48)=14.861$, $p=0.0001$. A significant interaction between the variables was also found, $F(66,528)=1.346$, $p=0.0428$ ($p<0.05$). This interaction was attributed to the differential effect of the orientations in the Z axis. For the jug, an analysis of variance proved significant for both orientation, $F(11,88)=3.989$, $p=0.0001$ and axes, $F(6,48)=4.946$, $p=0.0005$. A significant interaction between the two variables was also found, $F(66,528)=1.709$, $p=0.0008$. Finally, an analysis of the reaction times to the lamp revealed a significant effect of orientation, $F(11,88)=1.848$, $p=0.0576$ and of axes $F(6,48)=3.252$, $p=0.0092$. A significant interaction between the two variables was not found, $F(66,528)=0.922$, $p=0.6511$. For none of the objects was there evidence of a unique best view (see Figure 14).

3.3.3 DISCUSSION

The analysis of the results did not reveal a benefit for a single view of these objects or indeed in any single object. Nor was there any particular benefit for the upright view (270° in the X, XY and XZ axes). Indeed there was no differential effect between views that contained the maximum amount of information about the principal axis and those views with reduced information about the principal axis apart from the fully foreshortened views (0° and 180° views). It therefore seems that these results do not support Palmer's conclusions on single canonical views as representations of 3D familiar objects.

The results do suggest however, that the foreshortened view is the most difficult view to recognise. Once information about the principal axis emerges, then the object is readily recognised. However, reaction times to orientations 30° off the foreshortened view in Figure 11 suggest that these views are also difficult to recognise. It was decided that this observed trend justified, for subsequent orientation studies, planned orthogonal comparisons between the grouped reaction times to orientations 30° off the foreshortened view and reaction times to the axis-fully-exposed views ±30°.

One of the reasons why these results do not concord with those of the Palmer et al. study may lie in their choice of objects as stimuli for the experiment. Many of the objects used in their study have highly salient focal features. For example, the canonical view of an alarm clock and a house is the 2-D, frontal view. This suggests a different type of representation of objects where information about the features is important. For other objects, the canonical view preserved some depth information. The fact that the canonical view is

peculiar to each object may suggest lack of efficiency in the visual system.

Another difference between the present study and the Palmer et al. (1981) study was the number of orientations tested. This experiment tested a larger range of orientations than the Palmer study. It could be said that the latter study was not a comprehensive one of the views that facilitate speed of recognition.

These results also present problems for the 3–D, object-centred approaches. For these models to be built, the information about the principal axis is resolved from the $2^1/_2$-D sketch. As already discussed, this assertion should therefore predict that the reaction times to views where the principal axis is fully resolved would yield the fastest reaction times with monotonically increasing reaction times as the information about the principal axis is reduced. Such a linear effect, from the views with the principal axis fully exposed to the foreshortened views, was not observed for any of the objects and indeed there was no difference found between the reaction times other than the foreshortened views.

3.4 Experiment 2

It has been suggested from previous studies that practice and familiarity with the stimuli reduce the orientation effect and that the response times to different aspects become more uniform (Jolicoeur 1985, Tarr and Pinker 1989 and Bulthoff and Edelman 1991). In order to verify that the orientation effect found in the first experiment was not a product of practice and therefore over-learning of the orientations, the experiment was repeated using a larger number of subjects, each tested on a small number of views of each object.

The subjects in this experiment were trained on a different set of objects to the test objects and were presented with only four examples of the test objects in the experimental block. Practice effects with orientations of objects other than the test objects does not transfer across different sets of stimuli (Tarr and Pinker 1989). The following experiment was therefore set up in order to test whether subjects, who were not previously exposed to the stimuli used in the experiment, would show the same orientation effects as the subjects in the previous experiments.

3.4.1 METHOD

Subjects

Thirty six Psychology undergraduate students from the Department of Psychology, University of Durham participated in this experiment. Twelve of these subjects were male and 24 were female. Their ages ranged from 18 to 22 years. All subjects had normal or corrected–to–normal vision. None of these subjects had participated in the previous experiment.

Stimuli

See Experiment 1 for a description of the six objects used. A selection was chosen from among the views used in Experiment 1. Only rotations in the X and Y axes were used.

An extra twelve objects were also drawn and similarly rotated in the X and Y axes to be used in practice blocks. These objects were chosen arbitrarily from the set of objects in Snodgrass and Vanderwart (1981). The objects presented in the practice blocks were drum, cup, stool, whistle, brush, pot, bowl, table, clothes peg, light bulb, padlock and rolling pin. These objects were divided into two practice blocks which every subject was presented with.

The same apparatus used to display and record the responses in Experiment 1 was again used in this experiment.

Design

The experiment was based on a 3 factor, mixed design. The factors involved were axes of orientation, orientations and objects which was a nested factor under the orientations factor. The axis factor had only two levels; rotations in the X and Y axes. All twelve orientations were tested in each axis. Each subject was therefore presented with all orientations in both axes of rotation with the objects nested under these factors. Each object was presented four times to any one subject and the orientations of the objects shown were counterbalanced across all subjects so that all views of all objects were shown across the experiment. Subjects therefore saw all orientations within the experimental block.

The experiment was again based on a match/ mismatch design as in the previous experiment. Reaction times and errors were recorded as a measure of the recognition of objects in each orientation.

Procedure

Each subject was initially presented with two practice blocks of 60 trials each followed by an experimental block of 48 trials. The order of the trials across the practice blocks was randomised across subjects. All subjects received a self-timed break between blocks.

The procedure followed that given in Experiment 1: The subjects were instructed to respond as fast as possible to each trial without making too many errors. It was stressed that a response should be made to every trial. The reaction times and the 'SAME'/ 'DIFFERENT' responses were recorded and stored on the BBC microcomputer.

3.4.2 RESULTS

Responses made to the match trials only were analysed. The percentage errors made to the trials in this condition were 7.38% and indicated no evidence for a speed/accuracy trade off. Figure 15 below shows the number of errors made to each orientation. The errors

were not subjected to further analysis.

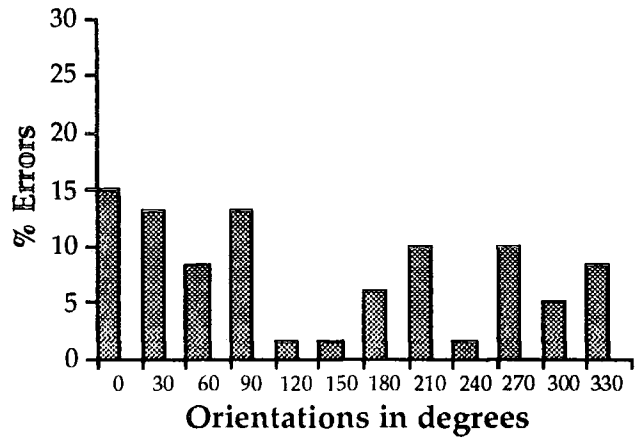

Figure 15: Percentage errors made across the different orientations of all objects.

Figure 16 below illustrates the mean reaction times to the different orientations collapsed over all other factors. A two factor repeated measures ANOVA was conducted on both the axes and orientation factors. This analysis yielded a significant effect of orientation, $F(11,319)=11.369$, $p=0.0001$. No other effects were found.

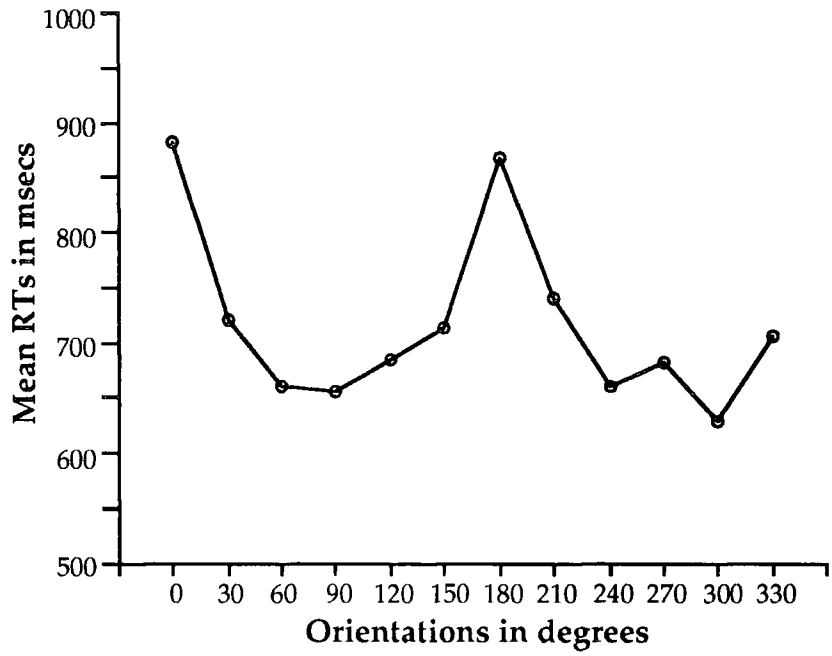

Figure 16; Subjects overall mean reaction times to the different orientations across all the objects shown in Experiment 2.

A post-hoc Newman-Keuls analysis on the orientation effect revealed that the mean reaction times to the orientations of 0° and 180° were significantly different than reaction times to all other orientations (except each other) at $p< 0.01$ level of significance. A significant difference was also found between the mean reaction times to the orientations of 300° and 210° at $p<0.05$ level of significance. No other differences were found between the

orientations.

The results of the previous experiment suggested that the time to recognise objects shown 30° from the foreshortened view was slower than views which show more information about the principal axis. Figure 16 shows the same pattern. A planned orthogonal comparison was conducted between the grouped reaction times to the views 30° off the foreshortened views and views which had more information about the principal axis namely the axis-fully-exposed views ±30° off this view. This analysis proved significant $F(1,261)=15.589$, $p=0.0001$, indicating that the reaction times to orientations off the foreshortened views were slower than those with more information about the principal axis.

Figure 17 below shows the mean reaction times to the orientations in each of the axes of rotation (X and Y). There was no difference found between the axes and no interaction between the orientations and the axes factors.

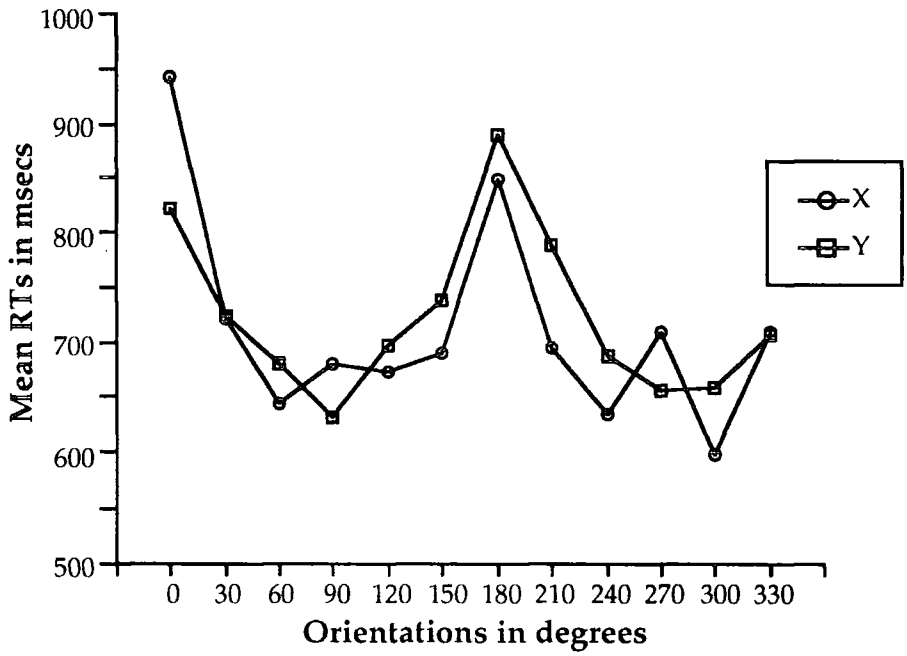

Figure 17; Subjects mean reaction times to the different orientations of the objects shown in the different axes of rotation in Experiment 2.

3.4.3 DISCUSSION

The results of subjects who were not trained on the stimuli beforehand were highly comparable to those of subjects who were highly trained on the stimuli. In both experiments the subjects took longer to recognise objects when shown with the principal axis foreshortened. A facilitation for an upright view of the objects was not observed for either group of subjects (270° in the X axis in the present experiment). The orientation pattern was highly similar for both experiments which suggests that elongated objects are faster to

recognise when the principal axis is available than when it is not. It is not necessary, however, to have the axis fully exposed for rapid recognition but there is not enough information available when the object is shown 30° from the foreshortened view.

These results suggest that the principal axis is important in the representation of rigid objects. The difference in the time taken to recognise objects in the foreshortened position and 30° from this position may reflect the time taken to normalise the object to match the nearest stored representation. Because there was no indication of a facilitation of recognition to objects shown with the axis fully exposed relative to views 30° off then this may suggest that these views (or views close to this) are represented. In other words there is more than one single representation per object.

3.5 Experiment 3

Experiment 3 tested whether the results found in the previous experiments generalised to a different category of objects. As was already mentioned, the objects used in the previous experiments all had strong gravitational uprights (determined by a rating study). There could be an argument therefore that the results reflect a bias towards the amount of information available in the most familiar view. In other words, the most familiar view that is, the upright view of the set of objects used in the previous experiments has maximum information (or near maximum information) available about the principal axis. It could be argued that any views that share the same amount of information about the principal axis as the most familiar view will be easier to recognise than other views. Indeed the similarity of results between the previous experiments add strength to this argument. According to this argument, objects that have no particular upright (or a bias of having a familiar orientation) may not exhibit the same orientation effects.

A set of objects were chosen that are typically viewed and recognised in a number of different aspects including the foreshortened view. These objects were selected due to the results of a rating study reported above where 21 subjects rated a set of objects on uprightness. The objects chosen for this experiment were rated low for uprightness.

The principal axes of these objects was not a necessary feature of the familiar view. If, as was suggested by previous work (Jolicoeur, 1985; Tarr and Pinker, 1989), the familiarity of the view determines the ease of recognition, then a flat function relating view to recognition times was predicted in this set of objects because familiarity of a particular view was not a bias. All of the objects used were elongated as in the previous experiments.

3.5.1 METHOD

Subjects

Seven Psychology students from the University of Durham participated in this experiment. Three of these subjects were male and four female. The age range of the subjects was 24 to 48. All subjects had normal or corrected to normal vision.

Stimuli

A selection of five common objects was drawn on the Macintosh IIx computer. See Experiment 1 for a description of how the objects were drawn. The objects drawn for this experiment were as follows; clothes peg, screw, rolling pin, light bulb, and whistle and were particularly chosen because all had non-defined uprights (see Rating study). A stimulus included a picture of an object-view and a label which either matched or mismatched the name of the object. All of the orientations in all axes were used as in Experiment 1.

The apparatus used in the two previous experiments was also employed to display the objects and record the reaction times.

Design

The experiment was based on a 3 factor, repeated measures design with objects, axes of rotation and orientations as factors. The experiment was also based on a match/ mismatch design, using pictures of objects and their corresponding labels as in the previous experiments.

The objects factor contained five levels with each level corresponding to an object. There were seven levels to the axes of rotation factor which included the three major axes X, Y, and Z and their combinations. Finally, there were 12 levels to the orientations factor, each level corresponding to orientations in 30° increments. However, as in Experiment 1, the 0° orientation was shown once for each object in the match trials in order to avoid learning, due to repetition, of the foreshortened views. See Figures 8 and 9 for an illustration of the orientations used. The rolling pin was oriented in four different axes of rotation. This was because the rolling pin is bilaterally symmetrical and rotations in the total number of axes would have resulted in a repetition of the views which may have resulted in a recognition bias for that object due to the number of repeated orientations. For this reason, the rolling pin was viewed in rotations around the X, Y, XY and XYZ axes of rotations.

For the purpose of this experiment it was considered necessary that the subjects be well practised. The subjects therefore received a practice block of 100 trials containing an example of each orientation with objects and match/ mismatch trials as nested factors.

Procedure

The general procedure followed that outlined in Experiment 1. The total number of trials was randomly divided into 7 experimental blocks, with 100 trials in block numbers 2 to

6 and 97 trials in both block numbers 1 and 7. Each block contained equal numbers of match and mismatch trials. Dummy trials were included at the beginning of each experimental block since pilot studies had shown a speed up effect over the first three or four slides. The presentation of the blocks across subjects was counter balanced using the Latin square method.

Each subject was initially presented with a practice block of 100 trials taken randomly from the experimental blocks. The seven experimental blocks separated into two groups of 3 blocks and one of 4 blocks with a self timed break between the groups.

3.5.2 RESULTS

Figure 18 shows the mean number of errors made to each of the orientations. There was no evidence of a speed/accuracy trade-off and the errors were not subjected to further analysis.

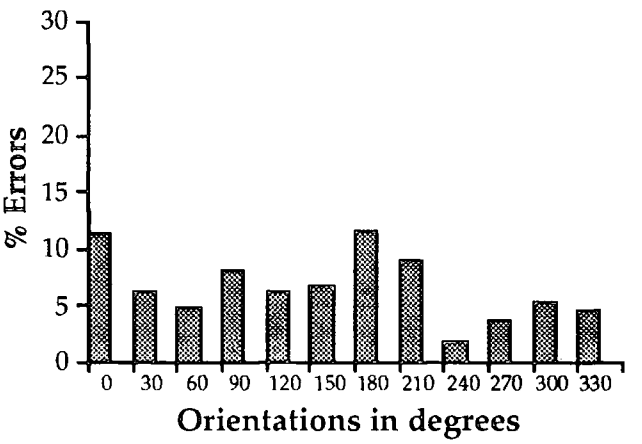

Figure 18: Percentage errors made to each orientations of all objects tested.

The reaction times to the different objects was collapsed over the other two factors and a two factor repeated measures ANOVA was conducted. This analysis proved highly significant for axes, $F(6,36)=34.859$, $p=0.0001$ and for orientation, $F(11,66)=11.861$, $p=0.0001$. A Newman-Keuls post hoc analysis on the axes effect revealed that the Z axis was significantly different from all other axes at $p<0.01$ level of significance. The same analysis on the orientation effect revealed that both the 180° and the 0° rotations were significantly different from all other orientations but not from each other at $p<0.01$ level of significance. There were no other differences found within either effect. Figures 19 and 20 below show the mean reaction times to the different views over all objects and axes and in each of the different axes of rotation respectively.

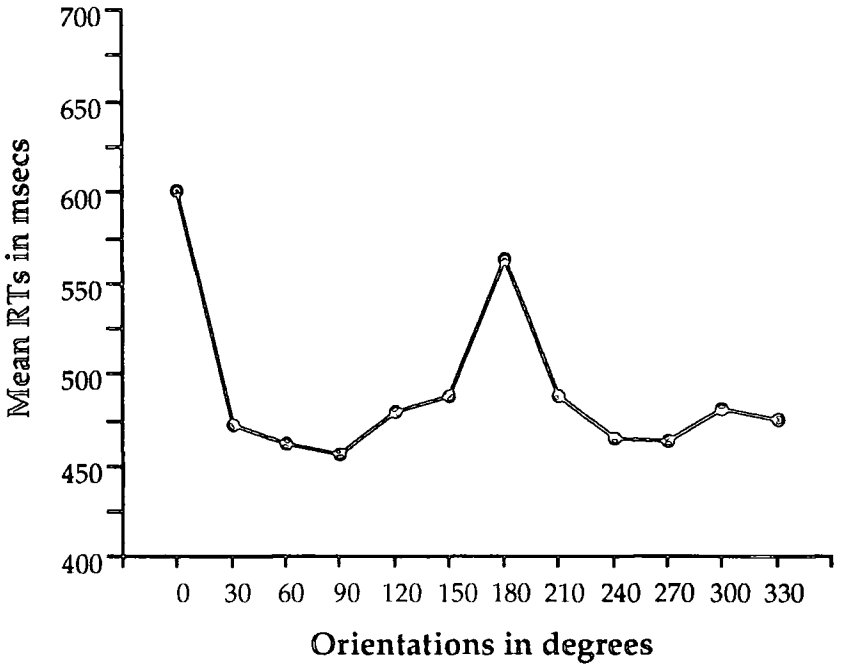

Figure 19; Subject's overall mean reaction times to the different orientations of the objects shown in Experiment 3.

A highly significant interaction between the two main variables was also found, $F(66,396)=2.325$, $p=0.0001$ which was attributed to the differential orientation effects in the Z axis relative to the other axes. There was no orientation effect observed in this axis.

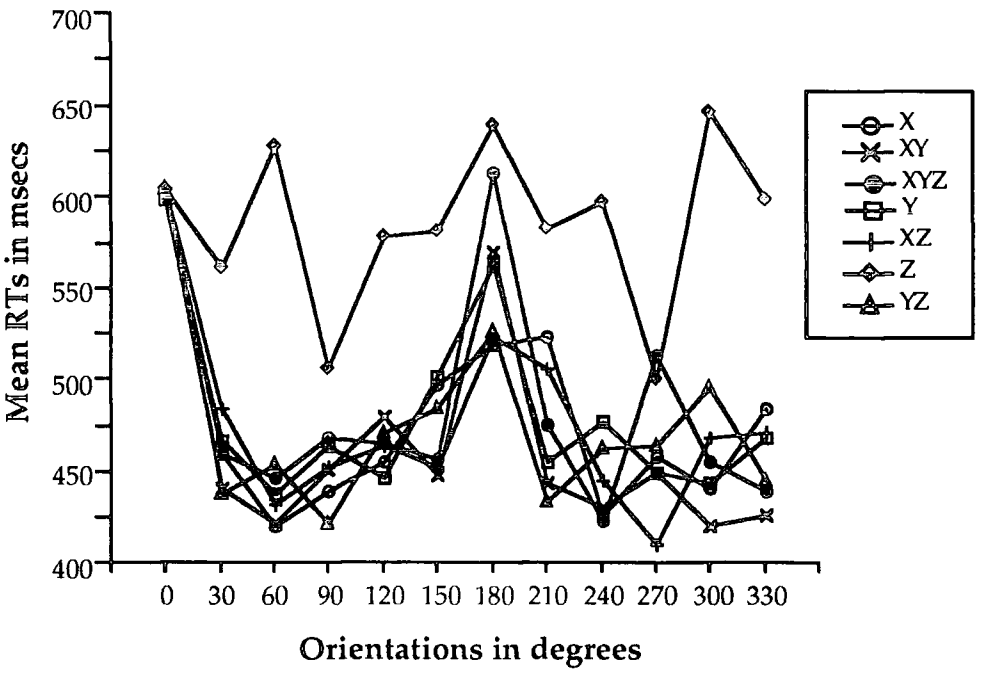

Figure 20; Subjects mean reaction times to the different orientations of the objects shown in each of the axes of rotation.

A planned orthogonal comparison was conducted between the reaction times to the

orientations 30° from the foreshortened view and views that contained more information about the principal axis i.e. the axis-fully-exposed views ±30°. This analysis proved significant, $F(1, 54) = 10.265$, $p = 0.0023$.

A separate analysis on each of the objects is given below. Figure 21 shows the mean reaction times to each of the objects across all of the orientations and collapsed over the axis factor.

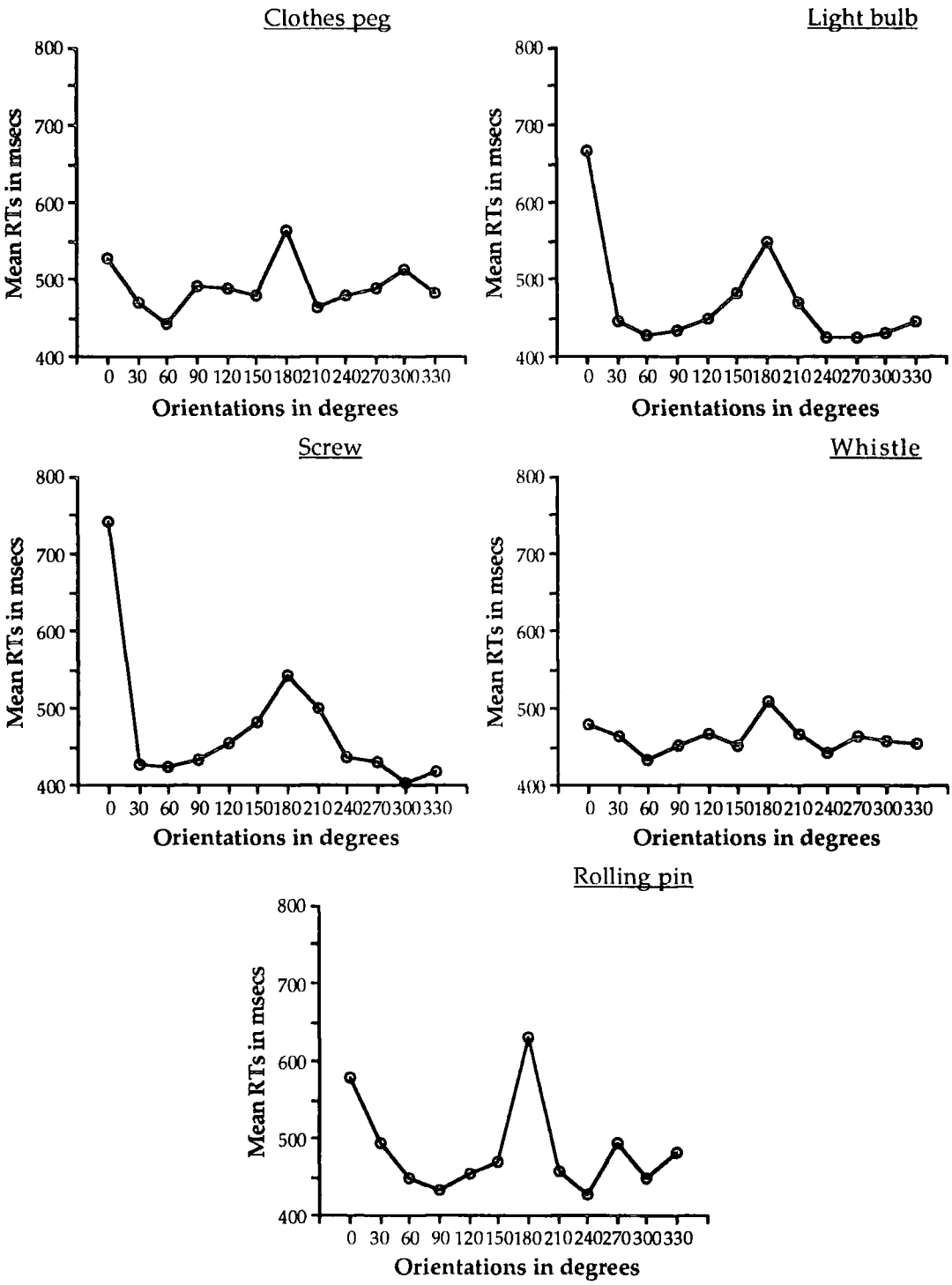

Figure 21: Mean reaction times to the different orientations of each object shown in Experiment 3.

An analysis of variance on the reaction times to the clothes peg found a significant effect of axes, F(6,36)=3.612, p=0.0066 but no effect was found for orientation. A significant interaction between the two variables was also found, F(66,396)=1.662, p=0.0018. An analysis of variance on the raw data for the light bulb proved significant for axes, F(6,36)=5.093, p=0.0007 and for orientation, F(11,66)=7.330, p=0.0001. A significant interaction between the two variables was also found, F(66,396)=1.982, p=0.0001. For the purposes of this experiment the rolling pin was only oriented in 4 axes; X, Y, XY and XYZ (see design section above). An analysis of variance proved significant for orientation, F(11,66)=5.678, p=0.0001 but not for axes. An analysis of variance on the reaction times to the screw proved significant for axes, F(6,36)=25.910, p=0.0001 and for orientation, F(11,66)=4.898, p=0.0001. A significant interaction was also found, F(66,396)=1.821, p=0.0003. Finally, an analysis of variance on the reaction times to the whistle showed no significant main effects. No other significant effects were found for any of the objects. There was no evidence for a single 'best' or more recognisable view for any of the objects.

3.5.3 DISCUSSION

The results of this experiment were highly consistent with the results of the previous experiments. They indicate that recognition times did not favour a single view of the objects but rather a number of views were recognised equally as fast. Again, an increase in the reaction times was caused by views with a foreshortened axis. An orthogonal comparison revealed that the recognition times to views that are 30° away from the foreshortened views were significantly different from views that contained more information about the principal axis. This result concords with the result found in the second experiment. It suggests that recognition time is dependent on the view of the object and that having little information about the principal axis causes an increase in recognition time. On the other hand, there was no observed facilitation for the views with the principal axis fully exposed which may suggest that these views are not isolated to represent the object in memory but that several views are stored, each stored view containing a high degree of information about the principal axis. It would seem plausible that the visual system store objects in this way because in order to avoid identity mistakes then information about the object would need to be maximised such that each representation is sufficiently unique. It could be suggested that the delay in reaction times to views 30° from the foreshortened view reflects the time taken to transform or normalise the object to match its nearest stored view.

3.6 Experiment 4

The results from the previous experiments revealed a significant effect for orientations in depth. It was argued that these effects may be due to the fact that the visual system may not store representations of objects where the information about the object in the

image is reduced with orientation and therefore when an image of this sort is imputted for recognition it is normalised to match the nearest stored view of the object.

If this is so then rotations of objects in the picture plane (i.e. plane perpendicular to the line of sight) should have no differential effect on the recognition times. This is because there is no information loss of the elongated axis with rotations in the picture plane. Indeed there was no difference found in the previous experiments between the upright and inverted orientation of the objects (i.e. 270° and 90° in the X, XY and XZ axes). Rock (1973) argued that there is a phenomenal change in the image of an object with rotation in the picture plane which one does not get with rotations in depth. He claimed therefore that objects rotated in the picture plane are more difficult to recognise than objects rotated in depth. Jolicoeur (1985) found a linear effect in recognition times with rotations in the picture plane away from the upright. He found that with practice this effect diminished suggesting that when the familiarity of the view increased, recognition time decreased. This experiment helps to address this discrepancy between the effects of rotations in depth and in the picture plane.

3.6.1 METHOD

Subjects

Nine undergraduate and post-graduate students of the Department of Psychology, University of Durham participated in the experiment. The age range of the subjects was 18 to 33. There were four female and five male subjects. All subjects had normal or corrected-to-normal vision.

Stimuli

A set of six objects were drawn using the Swivel 3D package for the Macintosh. These objects consisted of a bottle, glass, lamp, clothes peg, light bulb and screw. The first three of these objects had well-defined environmental uprights and the latter three had no typical environmental orientation (see rating study). The package allowed the images to be rotated in the picture plane in increments of 30°. The picture plane refers to the plane which is perpendicular to the line of sight. Twelve images of each object were therefore constructed. All of the objects were rotated from the same, corresponding starting or 0° position. Figure 22 below illustrates the 0° position for all of the objects and the relative orientations around the picture plane. The equivalent 0° position for all of the objects was the inverted position. Rotation in the picture plane corresponded to the rotation of the principal or elongated axis of the objects.

Once the images were constructed labels were placed over each object image using a package called Enhance. These labels corresponded to either the correct name of the object shown or the incorrect name of the object. A stimulus therefore consisted of an image of an object in a particular orientation with a label shown directly above it. All objects were shown

against a white background. An inter-trial-interval consisted of a blank white screen which remained for 1 second until the onset of the next stimulus.

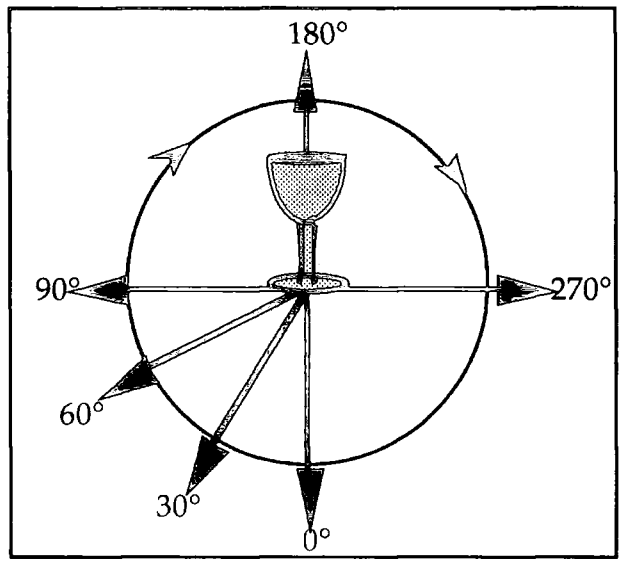

Figure 22 : Illustration of the orientations in the picture plane of the objects used in the experiment. The thick, straight lines represent the direction of the object's elongated axes (for example the glass would be shown upside-down in the 0° position). The circle indicates the direction of the orientation increments. The objects were therefore shown in increments of 30° in a clockwise direction from equivalent 0° reference orientations.

Apparatus

A Macintosh IIx computer was used to display the stimuli and record the data. The stimuli were displayed and the data recorded using a custom-built programme written for the Macintosh. Each stimulus was displayed on the screen until the subject had responded or for a maximum of 5 seconds. A response consisted of depressing the appropriate key on a response box which was attached to the Macintosh. Reaction times were measured from the onset of the stimulus to the subjects response.

Design

The experiment was based on a two-way repeated measures design with objects and orientations as factors. The objects had six levels corresponding to the six different objects; bottle, glass, lamp, clothes peg, light bulb and screw. The orientations factor had twelve levels each corresponding to a particular orientation of the image shown; 0°, 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300° and 330°. Figure 22 above illustrates the orientations.

The experiment was of a match/ mismatch design were the subject had to decide as fast as possible whether the label shown with an object was the correct or incorrect name of the object shown. There were equal numbers of match and mismatch trials across the

experiment and the names of the other objects served as mismatch labels for each object. The objects and labels were counter-balanced across the experiment.

The experiment was repeated three times in order to test for the effects of practice on the reaction times to the different orientations. Each run of the experiment was referred to as an experimental block.

Procedure

Each subject was initially presented with a short practice block in order that they were familiar with the nature of the drawings of the objects and the task. The practice block contained 12 trials, one for each orientation, with nested objects and trial types. The practice block was short in order to test for effects of practice within the experiment proper.

The experimental block of 144 trials immediately followed the practice block. This experimental block was repeated 3 times in total and subjects were allowed short breaks between each block.

The subjects were instructed to decide whether the label shown over the drawing of the object was the name of that object. They were instructed to respond as fast as possible without making too many errors. A response consisted of depressing a 'yes' button if the label was the name of the object shown or a 'no' button on a response button box. Reaction times were recorded by the Macintosh IIx.

3.6.2 RESULTS

The mismatch trials were not subjected to analysis. The errors across all nine subjects in the match trials of the experiment totalled 2.4%. Figure 23 below shows the mean percentage errors made to each orientation. As there was no evidence of a speed/accuracy trade-off the errors were not subjected to further analysis.

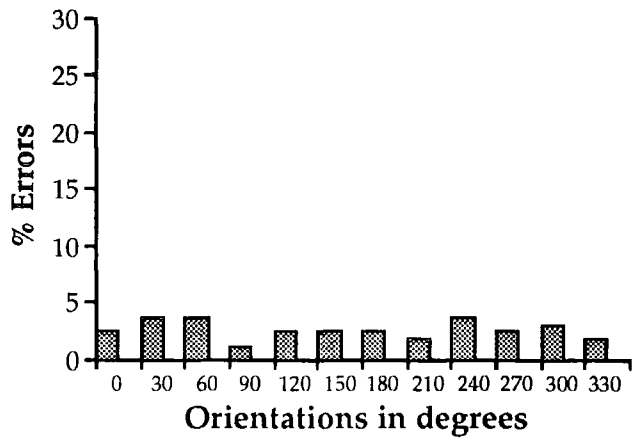

Figure 23: Percentage errors made to the different orientations of the objects in the picture plane.

The reaction times across all subjects in each block were subjected to a two-way, repeated measures analysis of variance with objects and orientations as factors. Figure 24 below shows the mean reaction times to the different orientations in each of the experimental blocks. The results to each of the experimental blocks is outlined below.

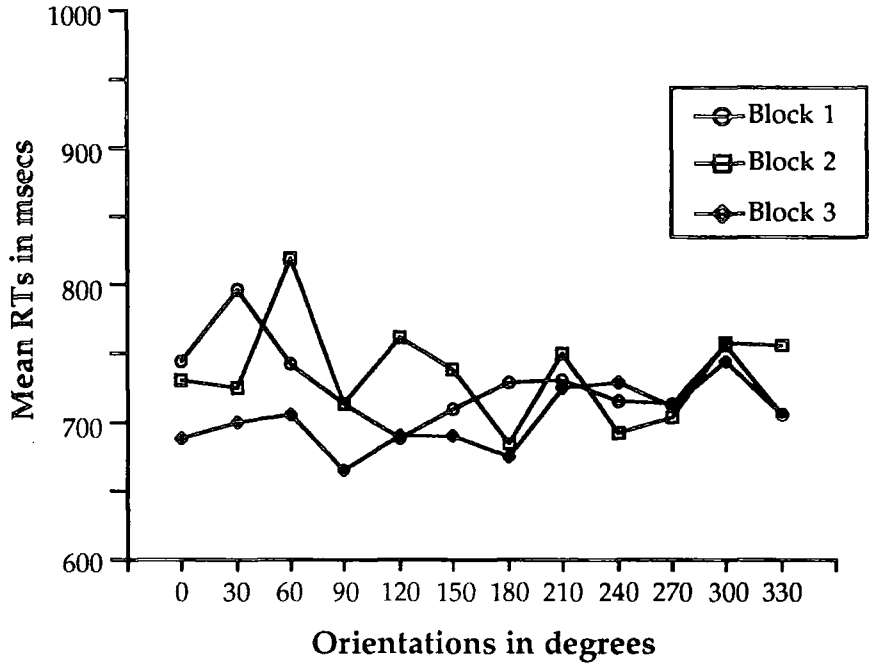

Figure 24; Subjects mean reaction times to the orientations of the objects in each of the experimental blocks in Experiment 4.

Block 1;

There were no significant effects found for any of the main factors: For the objects factor F(5,40)=0.601, p=0.6997 and the orientations factor F(11,88)=0.697, p=0.7379. There was no interaction found between the two factors, F(55,440)=0.676, p=0.9629.

Block 2;

A significant effect of objects was found, F(5,40)=2.502, p=.0461. There were no other significant effects found for the orientation factor F(11,88)=1.262, p=0.2601 and no interaction F(55,440)=1.005, p=0.4686. The main effect of objects was subjected to a post-hoc Newman-Keuls analysis which revealed that the Bottle was recognised significantly faster than the Glass at $p \leq 0.05$ level of significance.

Block 3;

There were no significant effects found for any of the factors in the third block: for the objects factor F(5,40)=1.519 p=0.2057 and the orientations factor F(11,88)=0.933, p=0.5127. There was no interaction found between the two factors F(55,440)=0.835, p=0.7935.

In general, there were no differences found between the different orientations in each

of the blocks. Figure 25 below shows the overall mean reaction times to the objects in each of the orientations. There was no evidence for a differential speed of recognition over the orientations.

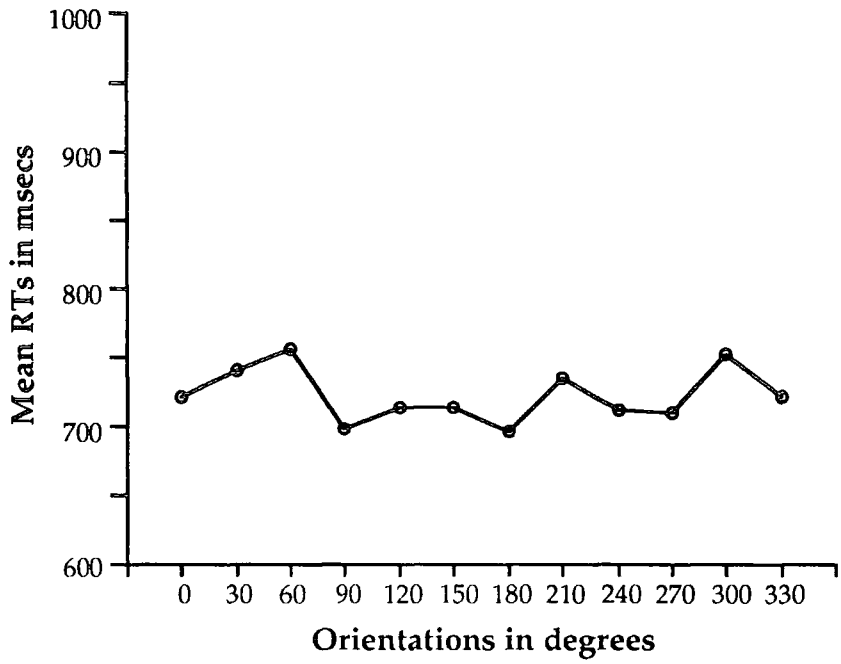

Figure 25: Subjects overall mean reaction times to the orientations in the picture plane of the objects shown in Experiment 4.

A one-way Anova across all reaction times, with block order as a factor, was not significant, $F(2,16)=0.567$, $p=0.5784$. Figure 24 above indicates that there is little observable difference between the blocks across the orientations.

Figure 26 below shows the mean reaction time to each of the objects over all three blocks. It can be observed that three objects with a previously highly rated upright did not show decreased recognition times to that view (see bottle, glass and lamp) and that reaction times to different orientations of the objects without a typical orientation in the environment (clothes peg, light bulb and screw) were not different to objects with a typical upright.

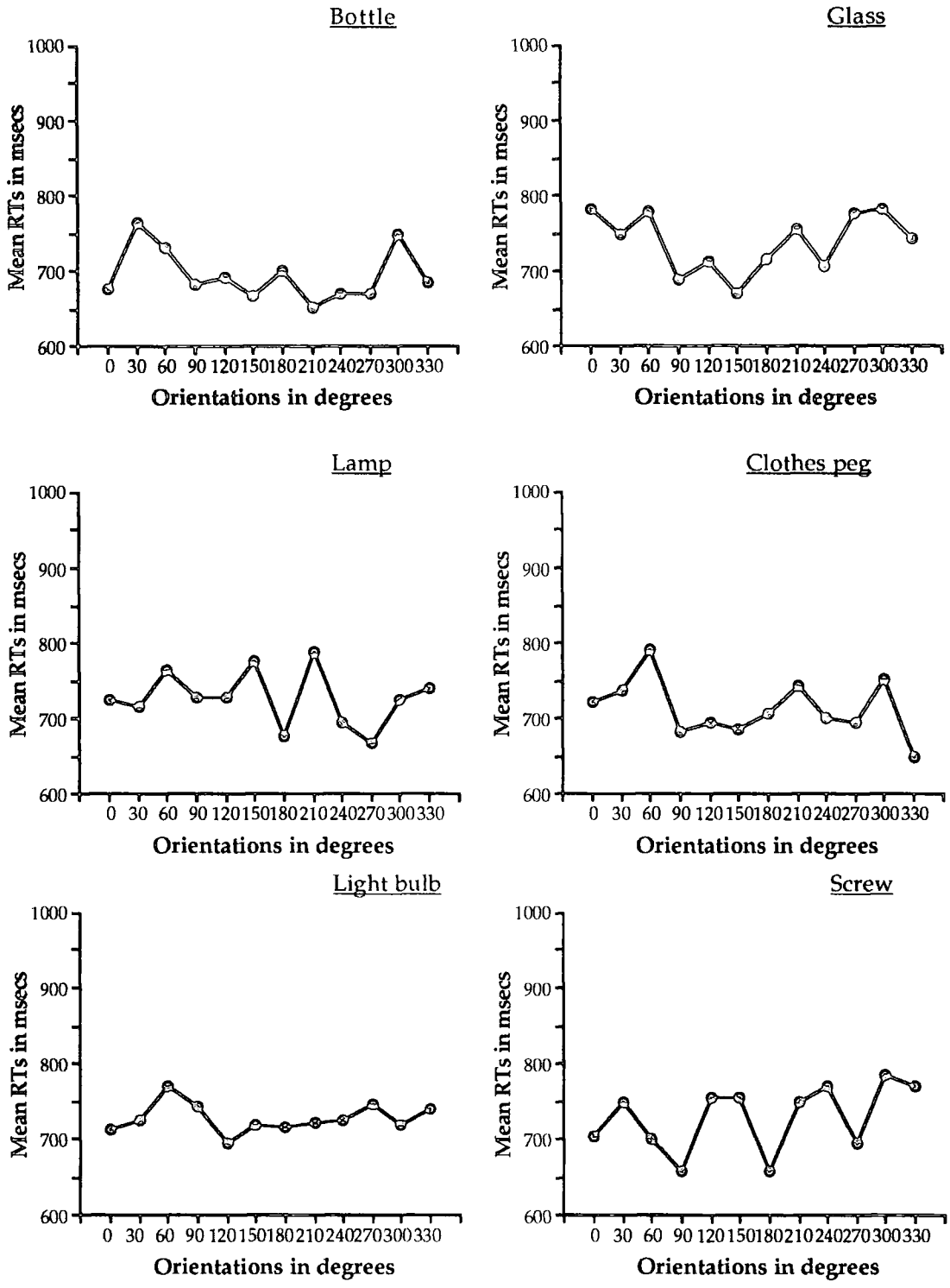

Figure 26; Mean reaction times to the different orientations of each of the objects shown in Experiment 4.

3.6.3 DISCUSSION

There was no significant effect of orientation found in any of the experimental blocks. There were no differences found between the orientations in the first block and

practice had no effect on the recognition times of objects in different orientations across the experiment. This result is different from Jolicoeur's (1985) findings and may reflect on the types of objects used in the experiments. The objects used in this experiment had the constraint of having strong principal axes. There was no such constraint on the objects used in Jolicoeur's experiment. The nature of the stimuli was also different: While Jolicoeur mostly used line drawings of his objects, this experiment used shaded drawings generated from a 3–D object oriented drafting package. This difference in information may reflect the ease of recognising objects in different orientation in the picture plane. Representing objects only by their contour may result in an orientation effect due to the lack of information available.

Another possible interpretation of these results is that the nature of the representation in memory is such that it can compensate for orientations in the picture plane quite readily, provided the stimuli are highly familiar. It could be suggested that most of the orientations tested are familiar orientations and that these views are represented in memory. Small deviations from these views would not result in a significant increase in reaction times because normalisation to the nearest stored view may occur very rapidly.

In the second experimental block a significant difference was found between the recognition time to the bottle and the glass, the glass being slower to recognise than the bottle. As this effect was quite weak, and as there was no interaction found between the objects and the orientations in any of the other blocks then this result was not considered important to the overall results of the experiment.

3.7 General Discussion

In general, the results of the four experiments reported above have shown highly consistent results. The initial intention was to isolate a single canonical view which represents objects. If objects are represented in terms of a canonical view then a facilitation for this view was expected (Palmer et al., 1985). However, none of the experiments revealed a facilitation for any single view. This was true for objects with strong gravitational uprights and objects with arbitrary uprights. Practice with the objects did not cause the loss of a canonical view in Experiment 1 (due to the fact that other views were over-learned) because subjects in Experiment 2, who were not trained on the test objects, yielded the same orientation effects. A 3/4 view facilitation was also not found for any single object, as Palmer et al. had predicted.

On the contrary, there was a strong suggestion that several views of objects are stored and that unusual views of those objects are mentally transformed and matched to the nearest stored view. The results from the initial three experiments indicated a differential effect between recognising objects shown from the foreshortened view, 30° from the foreshortened view and other views that contained more information about the principal axis (i.e. the axis-

fully-exposed views ±30°). This differential effect observed may reflect the time taken to mentally transform the objects to align to the nearest stored representation. There was no particular benefit found to views that showed the maximum amount of information i.e. the views with the elongated axis fully exposed, over views 30° from those views. It could be suggested that because recognition is equal across the views which maximise the information about the principal axis and views 30° away, that these views are represented. This suggestion would account for the result that orientation is invariant across these views. The conclusion that objects are represented as a number of views is consistent with more recent models of object recognition, particularly that of Tarr and Pinker (1989).

The differential effect on recognition found between the different views shown in the experiments remained constant with practice. The results of Experiment 4 indicated that orientation effects are not vulnerable to practice as was suggested by previous studies (Jolicoeur, 1985 and Tarr and Pinker, 1989). That orientation effects were found to disappear with practice in both the Jolicoeur and Tarr and Pinker studies can be attributed to the fact that different orientations of these stimuli were unfamiliar but practice rendered many more orientations familiar and therefore more easily recognisable. As already discussed, Jolicoeur (1985) studied the effects on naming times of different orientations of natural objects. These objects are typically seen in an upright position (e.g. dog and elephant) which makes that orientation more familiar and therefore more recognisable than others. On the other hand, the objects used in Experiment 4 may be seen from a variety of orientations due to handling of the objects or even different positions of the observer (even though half of the objects had well defined uprights). More views of these objects may therefore be familiar. The lack of an orientation effect in the picture plane was thought to arise because familiar views are more wide spread across orientations in the picture plane. If the stored views were sufficiently close to each other and spread out over the entire range of orientations, then this would result in no differential effect for the recognition of objects rotated in the picture plane. In other words, recognition would readily generalises to novel views of the objects between two stored views.

The second major finding from these experiments is that the principal axis is important in the representation of rigid elongated objects. This would seem consistent with the Marr and Nishihara (1978) approach to object recognition. However, according to their approach, an orientation effect should be observed with response times increasing monotonically as information about the principal axis is reduced. This increase in response time reflects the time taken to extract information from the $2^1/_2$-D sketch in order to resolve the principal axis and build a 3-D model. This was not found in the studies reported in this chapter. There was a difference in response times to objects shown in the foreshortened orientation, ±30° from that view and other views which contained more information about the principal axis with no particular benefit for the axis-fully-exposed views. It seems, therefore, that views that have no information about the principal axis, or a minimum

amount of information, are considered redundant for the purposes of representation and only views which contain a larger amount of information about the principal axis are represented. These representations would seem to be useful and result in the minimum number of identity errors, not only for recognition purposes but also for discrimination purposes. Novel or unusual views are therefore transformed or normalised to match the nearest stored view of the object.

However there are no other consistencies between Marr's notions of representation and those suggested by the results reported here. The results suggest that an object-centred description is not built and that, in fact, representations of objects are collections of view-centred representations (Tarr and Pinker, 1989). This conclusion is supported by the fact that recognition is affected by some orientations and not others. Another reason for suggesting that representations are view-centred is that, according to the object-centred idea, practice should eliminate the need to mentally transform the objects and that the object-centred representation would be accessed directly. This was not found. In fact the non-practised subjects had the same orientation effects as the well practised subjects. Also, with practice it was found that the overall recognition speed increased but the orientation effect remained unchanged.

Why are these views represented? Perhaps the choice of stored representations are determined by the familiarity of the views as Tarr and Pinker found. It seems that the visual system chooses those views that have optimal information about the object and knows what views would be redundant as representations. Representations for different objects that are elongated and rigid seem to be highly consistent.

The issue of the nature of the views stored as representations is discussed in Chapter 5. The next chapter includes evidence for the nature of the informaion in the image that is used in object recognition.

Chapter Four

The Information used for Recognition

4.1 General Introduction

The results from the experiments described in the previous chapter revealed that some views of objects are recognised more quickly than others, and that foreshortened views and views 30° off the fully foreshortened view are less readily recognisable views. Recognition was found to be affected by rotations in depth (Experiments 1, 2 and 3) rather than rotations in the picture plane (Experiment 4). In general it was found that recognition was faster to views which maximised the amount of information about the object. It was concluded that recognition is view-dependent rather than object-dependent. The important questions to ask therefore are what information in the inputted image is important in order that a match between a stored representation and the input can occur and how is this information used to match novel views to a stored representation of the object?

In general, models of recognition have proposed that one of three types of information is accessed from an object's image and is transformed to match a stored representation of that object. An image of an object can therefore be transformed to match a representation based on either an image-like template, a select number of features of the image or on the basis of some perceptual reference frame. This chapter includes a discussion of the previous evidence for each of these proposals and an experimental investigation into the nature of the information used to match an image to a stored representation.

Many of the investigations into pattern matching and recognition have concluded that the entire image is transformed or normalised in order to match to a stored representation (Shepard and Metzler, 1971; Kubovy and podgorny, 1981; Koriat and Norman, 1984; Bartram, 1976 and Tarr and Pinker, 1989). Koriat and Norman (1984) found that subjects persisted in mentally rotating letters to the upright and not to the orientation of the preceding stimulus when subjects had to decide whether a rotated letter string constituted a real word. This effect remained across different intervals between the sequentially presented stimuli. They concluded that the image itself was transformed rather than some other abstract information such as the object's reference frame. Similarly, Bartram (1974 and 1976) found that subjects' naming latencies of objects increased when the view of the preceding stimulus was different from the view that the subject had to respond to. Naming latencies were fastest when the same views of an object were shown in sequence and slowest when different objects having the same name were shown. With practice however, subjects could name objects that were preceded by either the same view of the object or a different view

equally fast. He also found that objects with high frequency names showed no difference in naming latencies between conditions where the object was preceded by the same view or a different view. These results suggested to Bartram that the objects were represented in terms of picture codes and that when there is a change in view-point between two picture codes, then the representation of one of the stimuli is normalised to match the other (see Bartram, 1976).

Jolicoeur and Kosslyn (1983) provided an alternative account of the image transformation account. They argued that two pictures of the same object, whether different views or the same views share more features than pictures of different objects and the more overlap between the features, the faster the naming latencies. This argument can explain the findings reported by Bartram (1974, 1976) and other picture matching studies (Kubovy and Podgorny, 1981).

However, models which assume that recognition occurs by aligning an image-like template with a stored representation need to specify how this alignment process occurs. Recognition models which mentally rotate novel views to match the nearest stored view would require some sort of recognition of the object before the visual system can decide that the view observed is a rotated version of a stored representation (Tarr and Pinker, 1989). This does not seem feasible. An alternative approach to this process was proposed by Ullman (1989). He argued that inputted views are compared to a large set of possible stored views of objects in parallel on the basis of a limited number of features. Only three landmark features of an input would be needed in order to align it with the stored representations (Ullman, 1986). A small fraction of the shape information in the input image is therefore sufficient to isolate a set of possible stored views with which to align the input. These landmark features are referred to as the 'alignment key' which cues the object's orientation in the input image independently of its identity (Ullman, 1985). Such landmark features can include distinctive features or the principal axis of the object (see also Humphreys, 1984; Warrington and James, 1986). Similarly, Cutzu and Edelman (1992) postulated that inputted images are aligned with a stored view of the object by correlating the summed Euclidean distances between the features in the objects image to those in the nearest stored views. To such alignment models, information about the features or edges are important for determining the amount of alignment needed to match to a stored view and this process can proceed without any recourse to top-down information.

Other theorists have also postulated that the information in the edges or contours of images is important in building a representation which can be matched with the object representation in memory (Marr and Hildreth, 1980; Biederman and Ju, 1988). Biederman and Ju (1988) found that subjects could recognise or verify objects (match a name to an object) equally fast when shown either as a full colour photograph or as a line drawing showing only the objects major components. They concluded that representations based on the information from the edges of the objects mediate recognition, in contrast to surface information. Marr

(1982) also argued that there is sufficient information in the occluding contour with which to build a representation of the image in order to mediate recognition.

An alternative approach to the alignment of an image-like template or extracted features to a stored representation is a description of an object based on a perceptual reference frame (see Quinlan, 1991). In general, two types of reference frames have been postulated that mediate the recognition of objects. A representation of an object can either be described based on the object's intrinsic reference frame such as the object's principal axis (Marr and Nishihara, 1978) or described relative to an extrinsic reference frame such as the gravitational upright. Marr and Nishihara (1978) argued that a description of an object in co-ordinates based on the objects principal axis (axis of elongation or symmetry) is necessary for recognition. However, other studies have found support for the notion that objects are described relative to their orientation in the environment. For example, such studies have found that disoriented letters (Cooper and Shepard, 1973; Koriat and Norman, 1984, 1985; Robertson et al. (1987) and Jolicoeur, 1990) and pictures of objects (Rock, 1973; Jolicoeur, 1985) are mentally rotated to the upright before being recognised.

Jolicoeur and Kosslyn (1983) and Hinton and Parsons (1981) found that subjects could use representations based on either extrinsic or intrinsic reference frames. Similarly, Humphreys (1983) found that the nature of the perceptual reference frame used in describing objects depended on the shape of the object. For example, shapes with a salient principal axis such as a symmetrical axis were more likely to be described relative to an intrinsic frame of reference based on the principal axis rather than an extrinsic reference frame. However, when shapes had an ambiguous principal axis (such as a square for example) then they were recognised by aligning an axis to an extrinsic reference frame such as the environmental upright. Palmer (1989) however, found no evidence that an object's intrinsic reference frame is built relative to its principal axis i.e. the axis of symmetry or elongation.

If representations are view specific and not object-centred as was concluded from the results of the previous experiments, then textural cues to the orientation of the object such as the shading patterns on the surface of the object or even prior indications of the expected orientation of the object in the environment should not have any differential effect on the recognition times of the objects in the different orientations. Also, if images or features are aligned to match a stored representation as opposed to a perceptual reference frame, then prior knowledge of the orientation of the object should not affect recognition times to different views of the object. An experimental investigation into the image variables that affect recognition is described in the next sections.

4.2 Experiment 5

The results from the previous experiments suggested that the recognition of

elongated objects is dependent on the view of the object seen. The objects shown in the previous experiments were shown under controlled light conditions with the light source unchanging across the views shown and across the different experiments. The results would seem to reflect changes in the shape transformations and that each shape is mapped onto a representation in memory. If object representations store shape information only, then information about light intensity and source should not affect recognition. Recognition should therefore be invariant over changes in light intensity or light source.

The notion that shape information is the only information required for recognition and that recognition proceeds by matching the projected image onto the nearest stored view-centred representation of the object was tested in this experiment. Recognition times to two different versions of the same object i.e. a shaded version and a non-shaded or silhouetted version were measured. It was predicted that there should be no differential effect on the recognition times of the objects in the different orientations because information about the shape is available to both stimulus types and both sets of drawings would trigger the same representations in memory.

The results from the experiments reported in the previous chapter suggested that practice does not change the orientation effects observed in experiments where the subjects are well practised. This result does not tally with other studies on the effect of practice on the orientation effect. Tarr and Pinker (1989) found that practice diminishes the orientation effect observed in the initial blocks. Jolicoeur (1985) also found that practice on recognising different orientations of natural objects caused an initial orientation effect to disappear. The similarity of orientation effects reported between practised and non-practised subjects in Experiments 1 and 2 may indicate that representations to objects are already well defined and that practice cannot induce new representations. In other words the number of representations per object are limited (Tarr and Pinker, 1989). An alternative explanation may be that the initial orientation effect reported by Jolicoeur (1985) was peculiar to line drawings which are novel, unfamiliar versions of objects that are unlikely to be found in the natural environment and therefore may need to be relearned when shown in different orientations. A second prediction stated that practice would not affect the recognition times to different orientations of 2-D, silhouetted objects when the shape of the stimulus is already familiar (e.g. a common object).

4.2.1 METHOD

Subjects

Eight members of the department of Psychology, University of Durham participated in this experiment. Three of the subjects were female and five male. Their ages ranged from 22 to 30 years. All subjects had normal or corrected to normal vision. These subjects had not participated in any of the previous experiments.

Stimuli

4 common objects selected from among the objects used in Experiments I and III were used. These objects were a glass, a lamp, a light bulb and a rolling pin and were chosen because each object exhibited significant orientation effects in the previous experiments. All of the objects had strong elongated axes and minimum surface features.

The objects were drawn in the same way as those described in Experiment 1. Copies were made of these original drawings and were presented as silhouettes. The silhouetted versions of the stimuli were achieved by assigning the colour black to each of the 3–dimensional views within the Swivel 3-D package. Each object drawing was presented at five orientations relative to the original foreshortened position of 0° in increments of 30° in each of the X, Y and XY axes only. The stimuli were produced by adding a label to each object view in the same way as in the previous experiments. All objects were presented with an appropriate and inappropriate label as in the previous experiments. The stimuli were photographed from the Macintosh screen and formed into slides and back-projected onto a screen which was placed 114 cm away from the subject. Each image measured no more than 10 cm in either direction of the picture plane. The largest image on the screen did not subtend 5° of visual angle. The objects were shown against a white background.

The presentation of each slide was controlled by a three field projection tachistoscope which was operated by a BBC microcomputer. The microcomputer was programmed to trigger a shutter on the slide projector via the tachistoscope. Once a slide was presented it remained on the screen until the subject responded. This response triggered the offset of the slide. An interval of 1 second followed each trial to the onset of the next trial. A "SAME" or "DIFFERENT" key press on a response box and the reaction times were recorded by the BBC.

Design

The experiment contained two parts and was based on a match/mismatch design as in the previous experiments. The first part of the experiment was based on a four-factor, repeated measures design with drawing type, objects, axis of rotation and orientations as factors. The drawing type factor had two levels, silhouetted drawing and shaded drawings. Four objects were used in the experiment. The axes of rotation factor contained three levels each corresponding to the X, Y and XY axes. Six different orientations were used; 0°, 30°, 60°, 90°, 120° and 150° views. See Figure 8 in Chapter 3 for an illustration of the views used.

The experiment was divided up into four different experimental blocks for the purposes of testing for an effect of practice on the orientation effect. The factors outlined above were counter-balanced across all blocks. An object orientation was not repeated across match and mismatch trials within a block. For the match trials in blocks 1, 2, 3 and 4 their incorrect versions were in order of blocks 3, 4, 1 and 2 respectively. Consequently, each stimulus

was seen by half of the subjects with the correct version seen before the incorrect version and in the opposite order by the other subjects. It was thought necessary to counter balance this order in case learning should occur across the same stimulus views. Besides this constraint, the order of the slides was randomised within each block across all subjects.

A preceding practice block consisted of four different objects than those used in the experimental blocks. These objects were jug, bottle, clothes peg and frying pan. This practice block was based on a four-factor repeated measures design with axes and orientations as factors and objects and drawing type as nested factors. Equal numbers of match and mismatch trials were given. Different objects were used in the practice block to avoid learning effects.

Procedure

The instructions given to this set of subjects were the same as those given in the previous experiments. The practice block preceded the four experimental blocks. These four blocks were presented in one session to each subject with a self-timed break between each block. A warm-up trial of 4 slides taken randomly from the other blocks preceded each block given to each subject because previous pilot studies suggested a speed-up effect over the first few trials. These dummy trials were not subjected to any subsequent analysis. The presentation order of the blocks was counterbalanced across subjects according to the Latin Square method.

4.2.2 RESULTS

The total number of errors in the match trials in the experiment was 8.19%. Figure 27 below shows the mean errors made to each orientation in each stimulus condition. The error count for the 3-dimensional drawings in the match trials was 7.07% and the error count for the silhouettes was 9.32%. A Wilcoxon Signed-rank test revealed that significantly more errors were made to the silhouetted drawings than to the 3-dimensional drawings; z (corrected for ties) ≤ -3.044, $p=0.0012$ across the whole experiment.

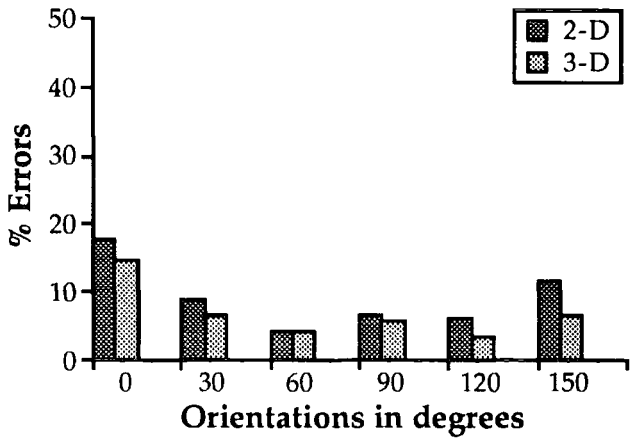

Figure 27: Percentage errors made to the different orientations of the objects shown in silhouetted (2-D) or shaded (3-D) form.

Figure 28 below shows the mean reaction times to the different drawing types in each orientation. A four factor repeated measures ANOVA was conducted across the reaction times of all subjects to the matched trials. Significant main effects of objects ($F(3,45)=9.884$, $p=0.0001$), axes ($F(2,30)=15.504$, $p=0.0001$) and orientation ($F(5,75)=27.521$, $p=0.0001$) were found. There was no effect found for the drawing type.

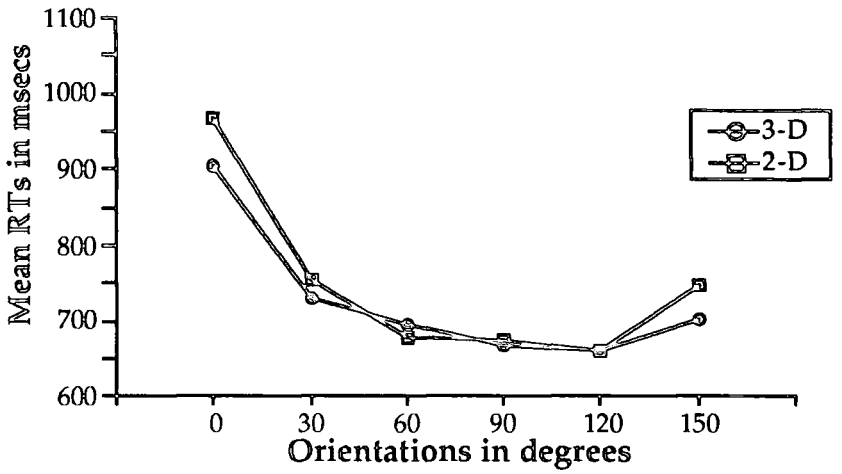

Figure 28: Mean Reaction times to orientations of objects presented in both silhouetted (2-D) and shaded (3-D) drawings.

Each main effect was analysed using a Newman-Keuls post hoc analysis. For the objects effect it was found that reaction times to the Rolling Pin were significantly faster to reaction times to all of the other objects at $p<0.01$ level of significance. A Newman-Keuls analysis on the axis effect revealed that reaction times to the Y axis were significantly slower than reaction times to the other axes at $p<0.01$ level of significance (see Figure 29 below). Finally, a post hoc analysis on the orientation effect revealed that reaction times to 0° orientations were significantly slower than reaction times to other orientations at $p<0.01$ level of significance. The 30° orientation was found to be slower than both the 90° and the 120° orientations at $p<0.05$ level of significance. No other differences were found within the main effects.

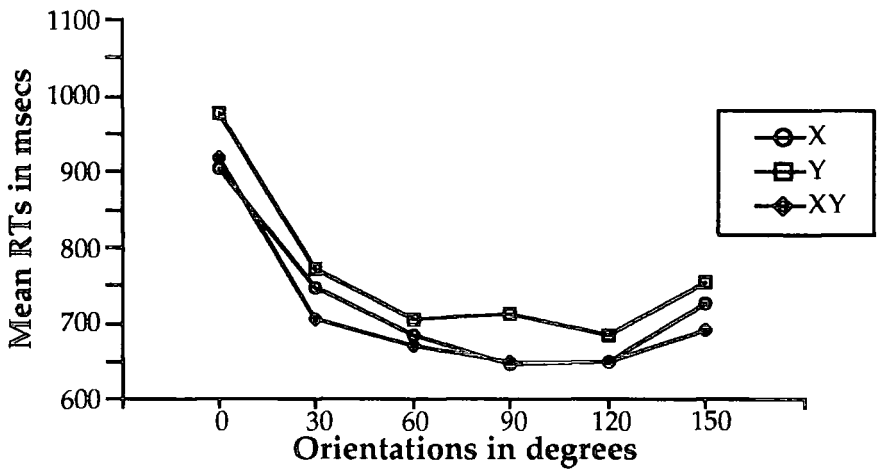

Figure 29; Mean reaction times to the different orientations of the objects rotated in each of the major axes.

A significant interaction was found between objects and rotation, F(15,225)=1.738, p=0.042. This interaction was attributed to the differential orientation effect to the Rolling Pin to all other objects. Figure 30 below shows the individual objects' mean reaction times to each of the stimulus versions across the different orientations.

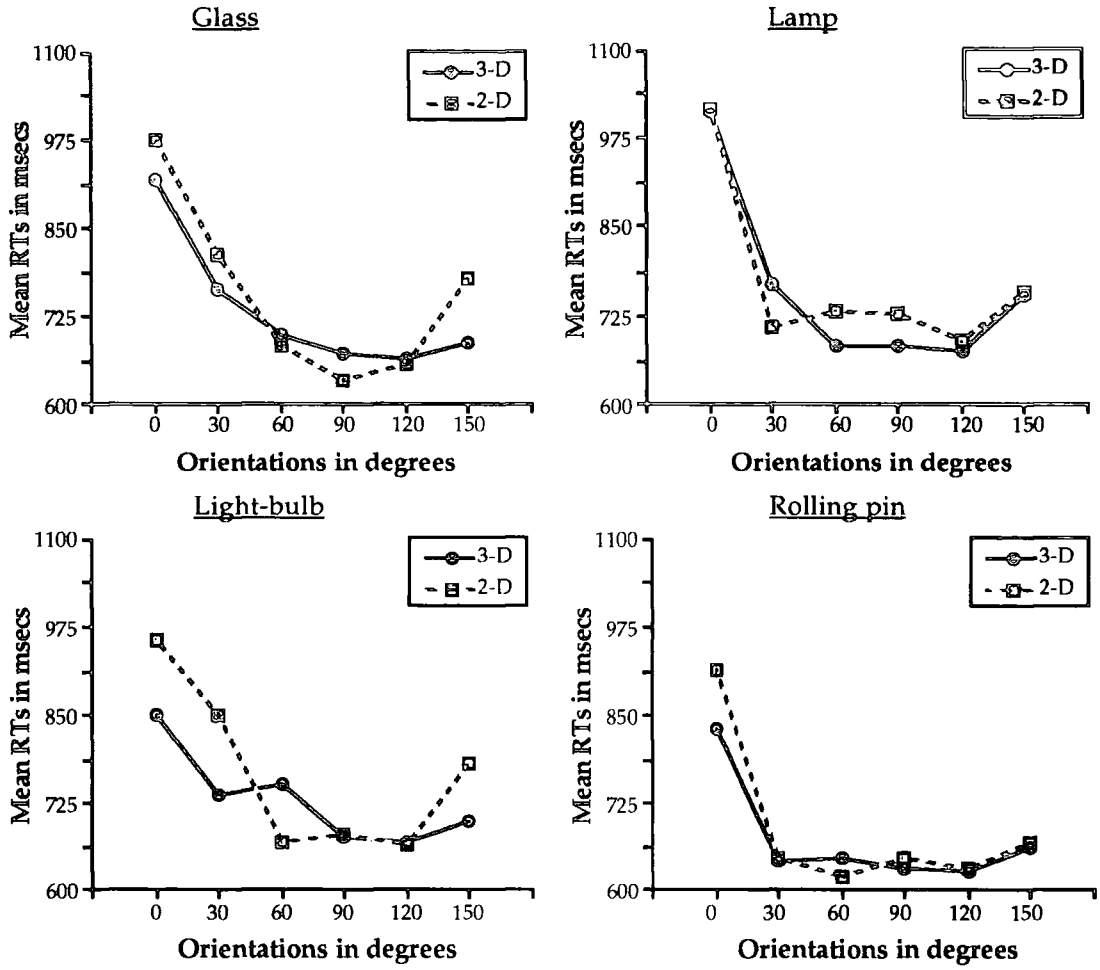

Figure 30: Individual objects mean reaction times to orientations of 3D and 2D images.

Orthogonal comparisons between the mean reaction times to orientations 30° from the foreshortened view and the axis-fully-exposed views ± 30° revealed a significant effect for both the shaded drawings F(1, 60) = 31.891, p=0.0001 and for the silhouetted drawings, F(1, 60) = 18.329, p=0.0001.

A second analysis on the data was conducted in order to test whether practice had an effect of making the orientation effect disappear. Figure 31 below shows mean reaction times to orientations collapsed across objects and axes for each block. The blocks shown correspond to the presentation order of the blocks that the subjects received.

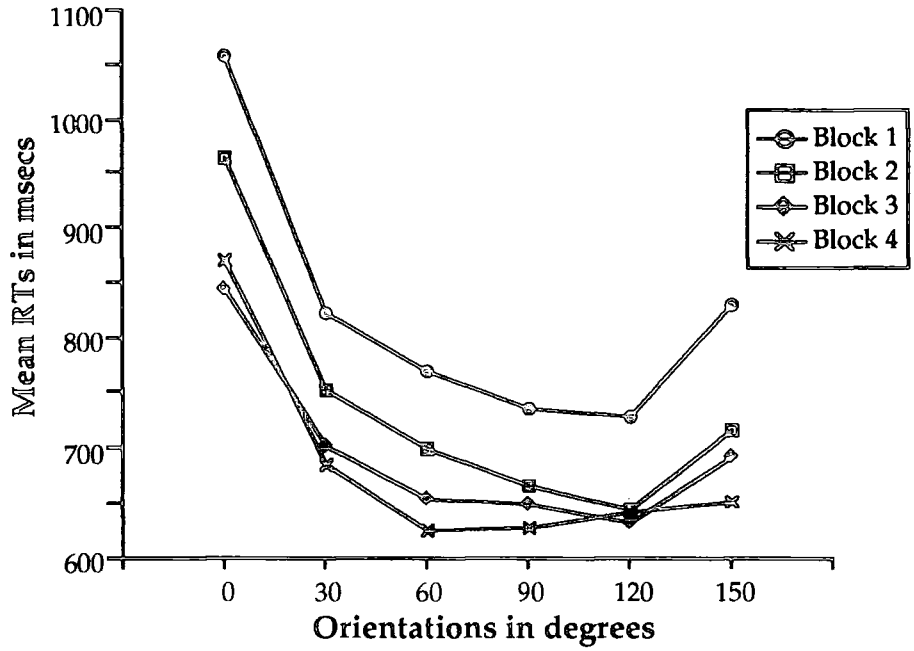

Figure 31: Mean Reaction times to orientations of objects collapsed over drawing type, objects and axes of orientation and shown for each block. The block titles given above correspond to the presentation order that the subjects received.

A four factor repeated measures ANOVA was conducted on reaction times to the match trials only with block order, drawing type, axes of rotation and orientations as factors. A significant main effect for block order was found, F=24.410, p=0.0001 and Figure 32 below illustrates the increase in overall speed across the different blocks.

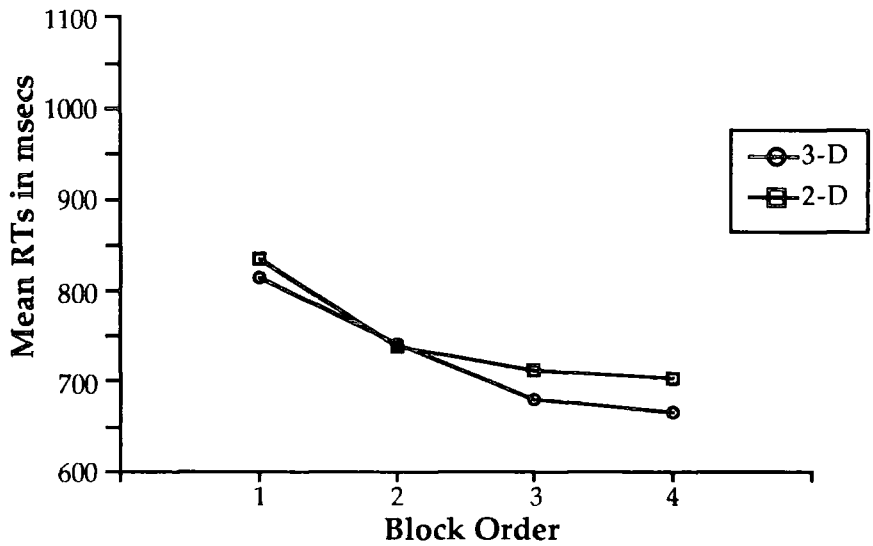

Figure 32: Mean reaction times given to each of the blocks for each stimulus version of the object.

A Newman-Keuls analysis on the block order effect showed that the reaction times to the stimuli given in the 1st blocks were significantly slower than the reaction times in all other blocks at $p \leq 0.01$ level of significance. Also the reaction times to the 2nd blocks were

significantly slower than those in the 4th block at $p \leq 0.01$ level of significance and the 3rd block at $p \leq 0.05$ level of significance. There was no difference found between the 3rd and 4th blocks given. There was no interaction found between the block order and the orientations or the block order and the drawing type. Other effects noticed in this analysis are the same as those given above in the initial analysis of the data.

4.2.3 DISCUSSION

The fact that no difference was found between the drawing types of the objects confirms the notion that representations of rigid, elongated objects do not need information about the surface or shading in order to recognise the object. The results suggest that information about the edges of an objects image are important in building a representation of the object to match to a stored view in memory.

The results also show a good deal of consistency with the results from the previous experiments in that there is a differential effect of recognition of objects shown 30° from the foreshortened view and objects shown with the axis fully exposed ±30°. This difference is present in both drawing types. It was also observed that the orientation effect did not disappear with practice which confirms the previous argument made that representations to these objects are already determined and the number of representations per object is limited. This result is not in accordance with Jolicoeur's (1985) study on the effect of practice on the recognition of different views of objects. This may be because Jolicoeur used line drawings of objects which are unfamiliar versions of the objects and therefore these version-specific stimuli had to be re-learned (see Tarr and Pinker, 1989). Shaded versions of objects may resemble real objects more than line drawings. The overall speed of recognition across the blocks did decrease but this could be attributed to the familiarity with the task. Also, both types of drawings were affected to the same degree with practice. The increase in speed over the experiment may be due to some motor effect for example. The results suggests that although the correct representation is initially accessed by a novel image of an object, familiarity with the task facilitates the speed with which this is done.

It was suggested that the interaction between objects and orientations which was attributed to the rolling pin may be due to the fact that the rolling pin is bilaterally symmetrical and therefore may have two principal axes in order to build a representation. Also, the aspect ratio is more pronounced in the rolling pin compared to other objects and aspect ratio may be an important metric variable in the representations of shapes. It is beyond the scope of the present thesis to test the effects of different metric variables on recognition but nevertheless the effects of such factors in the representations of shapes must be considered.

In sum, the results of Experiment 5 suggest that shading is a minor cue to recognition and indeed that the presence of shading is not necessary to recognise these objects. Lowe

(1985) asserted that shading and other depth cues are included after the recognition of the object. It may be the case therefore that depth cues are another route to object recognition but are not necessary in the recognition of rigid, elongated objects but that the shape of the object is necessary. This finding in itself is consistent with Marr's notion of object representation in that he claimed that there is enough information in the occluding contour of the object in order to resolve information about the principal axis of the object. However, Marr proposed that the principal axis is resolved from the information in the occluding contour. The following experiment addressed the question of whether an object-centred description which is invariant over view-points is derived from objects' axes of elongation.

4.3 Experiment 6

The notion of whether a reference frame intrinsic to the object is transformed, or whether a template-like image of the object is transformed into alignment with the objects stored representation was addressed in the following experiment. It was predicted that if the transformation of the object's reference frame hypothesis holds, then recognition times would be more uniform across the different views if the objects reference frame was given prior to the picture of the object. Conversely, if the image of the object is transformed into alignment with the nearest stored view of the object then priming the subject with the orientation of the object should have no differential effect between the recognition times to the primed views and the orientations that were not primed.

Based on previous studies, it was decided that the primes would align with the orientations of the elongated axes of the objects (Marr and Nishihara, 1978; Humphreys, 1983). Previous studies have shown priming effects of the orientations of stimuli but only when the identity of the test stimulus is also given with the orientation prime. Humphreys and Quinlan (1988) found priming effects of disoriented shapes when the frame-based descriptions of the primes and targets were similar. Subjects performed a task judging whether a 2-D shape had three or four sides. They were faster when primed with the identity and the orientation of the test stimulus than when primed with either the identity or the orientation alone. Cooper and Shepard (1973) also found that priming the orientation and identity of a letter decreased the time to decide whether a letter was shown in its normal version or backwards. Both studies found no effect of priming when the identity of the stimulus was inappropriate or omitted. However, Hinton and Parsons (1981) found that when subjects were given the orientation and the handedness of the subsequent stimulus that reaction times were faster than when no advance information was given.

There has been no work published to date on priming the orientations of familiar objects that have been rotated in depth where the task for the subject was to recognise the disorientated objects. It was therefore decided to run an orientation experiment to test the effect of primed orientations on the recognition of an orientated object. According to the

multiple-view, 2D representation approach we expected to find no benefit for primed orientations over unprimed orientations.

The following experiment also provided a test of Marr's recognition model that objects are represented by 3-D object-centred descriptions derived from the objects intrinsic reference frame. According to Marr, the main goal of the recognition system is to derive a 3-D object representation from the retinal image. This occurs by resolving the main or principal axis of the object from the 21/2-D or viewer-centred representation in order to build a 3-D object-centred representation. A failure to impose the correct object-centred reference frame to an object leads to an incorrect object-centred representation and consequently mistaken identity. It was assumed that if the correct object-centred reference frame was given prior to an object stimulus, that recognition of the foreshortened views (or views where the principal axis is difficult to derive) would be facilitated such that all views would be equally recognisable. This would be because the direction of the principal axis would already be specified, therefore the object-centred reference frame would be resolved and the time to represent the object in 3-D object-centred coordinates would be constant for all views.

4.3.1 METHOD

Subjects

Nine subjects from the Department of Psychology participated in this experiment. All of the subjects had normal or corrected-to-normal vision. The age range of the subjects was between 22 and 30.

Stimuli and Apparatus

A set of four objects was drawn using the Swivel 3D package. These objects were a lamp, a glass, a bottle and a light bulb. Each object was shown in 6 different orientations (0°, 30°, 60°, 90°, 120° and 150°) in three different axes (X, Y, and XY). The length of the elongated axis was the same for all objects. The package was also used to draw the priming stimulus (an arrow). This arrow was shown in all of the orientations in all axes mentioned above, with the arrow head corresponding to the position of the top of the proceeding object. Figure 33 below illustrates a primed trial. Each stimulus object was paired with either its correct or incorrect name.

The stimuli were displayed using a 2 field tachistoscope and were back-projected onto a screen. The subject was seated 114cm from the screen and the image on the screen was at most 10 cm in diameter. The visual angle therefore subtended 5 degrees for the views of the objects that maximised the elongated axes and less for the foreshortened views.

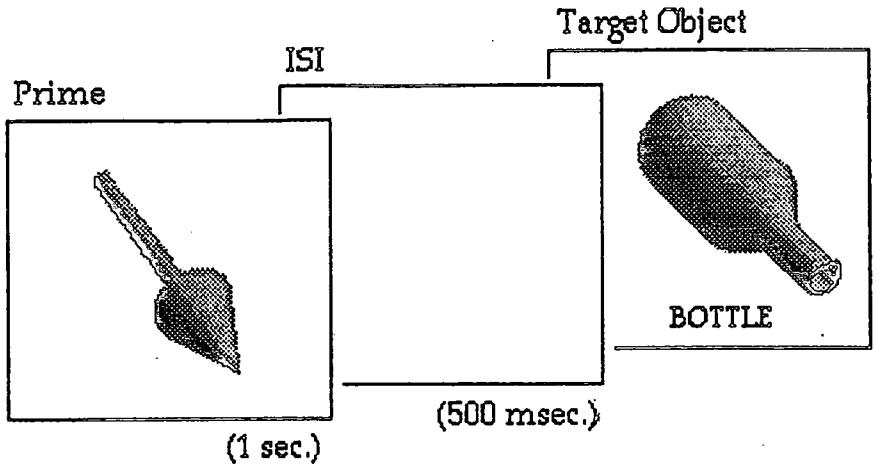

Figure 33: An illustration of a primed trial. The arrow was absent for unprimed trials.

The onset of each trial was triggered by a BBC micro-computer which also recorded the reaction times and the responses made by the subjects. Half of the trials were preceded by a priming stimulus which indicated the orientation or direction of the following object-stimulus and also the length of the objects principal axis (see Figure X above). The other stimuli were not primed and were therefore preceded by a blank slide. The timer was triggered from the onset of the second (or object) stimulus. The stimulus onset asynchrony (SOA) was 1.5 seconds: the first stimulus remained on the screen for 1 second which was followed by an inter-stimulus-interval (ISI) of 500 milliseconds and then the object slide remaining on the screen for 5 seconds unless the subject responded within that time. The subject's response therefore triggered the offset of the second stimulus and the onset of the next trial.

Design

The experiment was based on a four factor, repeated measures design with priming conditions, objects, axes and orientations as factors. It was also based on a match/ mismatch paradigm. All of the stimuli in the match condition were seen in both priming conditions that is, the orientation was primed and also unprimed. Half of the mismatch trials were primed. The orientation of the prime was always aligned to the orientation of the following object across the match and mismatch trials. A blank slide preceded the unprimed trials. The other three factors were counterbalanced across these conditions. The number of match trials totalled 216. The order of presentation of the trials was counter-balanced across subjects.

Each subject was initially presented with a practice block of 24 trials. The objects used in the practice block were different to those in the experimental block. The practice trials were balanced for priming conditions, match/mismatch trials and orientations. The axis type and objects were nested factors within the priming and orientation factors.

Procedure

The subject was instructed to decide as quickly as possible whether the label presented with the object was the correct name of that object. They were instructed to press the appropriate 'match' key as soon as they decided that the label was the name of the object and the 'mismatch' key if it was not the name of the object. They were told that in half of the trials the orientation of the object would be presented before the object and to respond to the object stimulus.

The subjects were initially presented with the practice block. The experimental block of 216 trials followed the practice block. A BBC micro-computer recorded the responses and the reaction times to each trial.

4.3.2 RESULTS

The total percentage of errors made across all subjects was 8.94%. There was no evidence of a speed/accuracy trade off in the experiment. The total number of errors made to the different orientations in both the primed and unprimed conditions are shown in Figure 34 below. There were proportionally more mistakes made to the 0° orientations in both conditions.

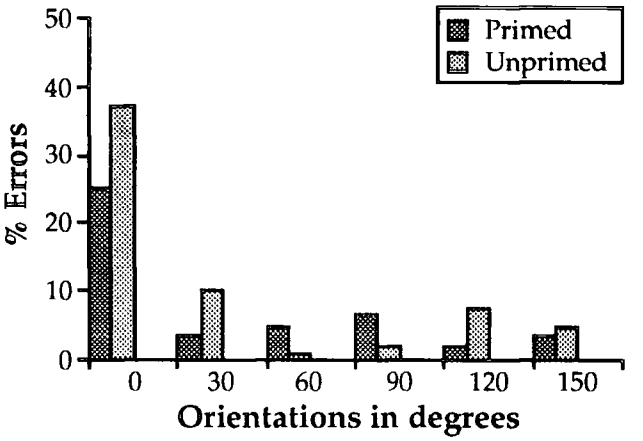

Figure 34: Percentage errors made to objects across primed and unprimed orientations.

The mean reaction times to the match condition for both the primed and unprimed orientations across all objects and axes of rotation are shown in Figure 35 below.

A four factor analysis of variance was conducted on the reaction times across all subjects. This yielded a significant main effect of priming, $F(1, 14)=9.970$, $p=0.0070$. The mean reaction times to the unprimed condition were significantly slower than those in the primed condition. A significant main effect was also found for the orientations factor, $F(5, 70)=32.566$, $p=0.0001$. A post-hoc Newman Keuls analysis revealed that the reactions times to 0° orientations (foreshortened) were significantly longer than reaction times to all other orientations at $p<0.01$ level of significance. There were no other differences found within this effect. The effects of axes of rotation and objects were not significant.

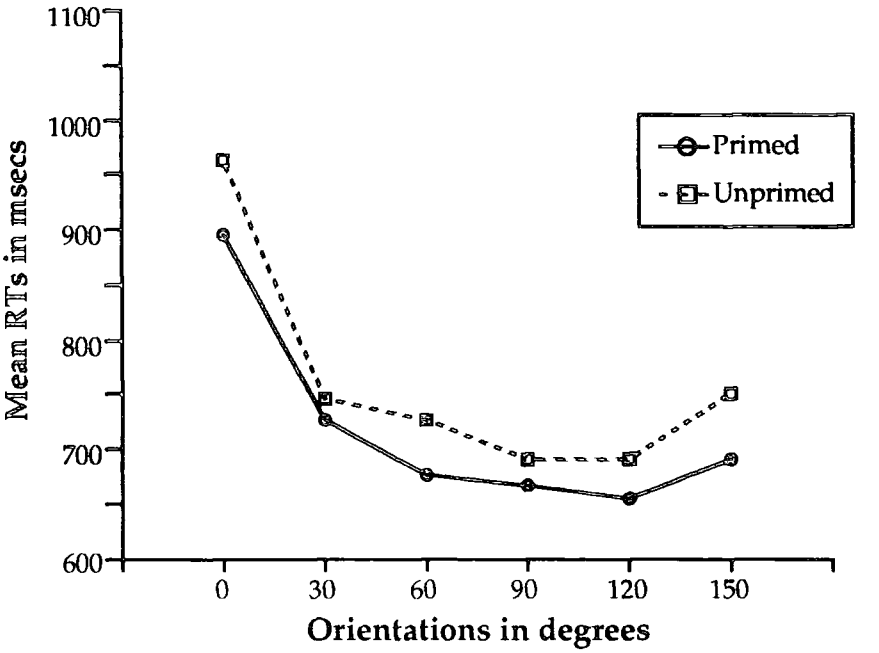

Figure 35: Subjects overall mean reaction times to the different orientations

There was no interaction found between the priming conditions and the orientations, $F(5,70)=0.952$, $p=0.4533$. A significant interaction was found between the priming conditions and the axes of rotation. This interaction is shown below in Figure 36.

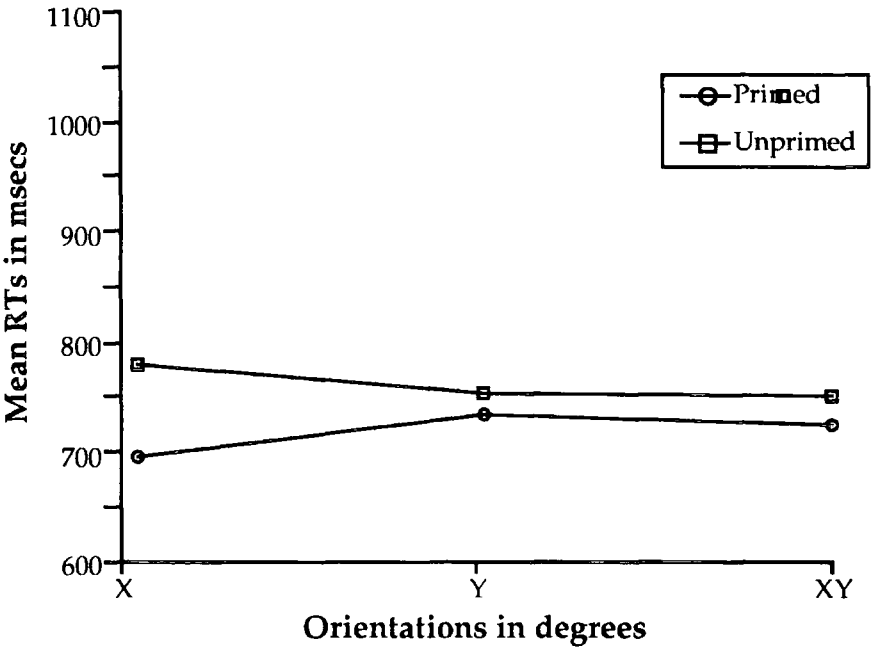

Figure 36: Mean reaction times to the axes of rotation in both the primed and unprimed conditions.

It was noticed that the reaction times to orientations in the X axis were slower in the unprimed condition and faster in the primed condition than the mean reaction times in the other axes of rotation. A post-hoc Newman-Keuls analysis showed a significant difference

between the primed and unprimed conditions of reaction times to the X axis at $p<0.01$ level of significance. The reaction times in the X axis in the primed condition were significantly slower than the unprimed Y axis and the unprimed XY axis at $p<0.05$ level of significance. The primed XY axis were significantly different from the unprimed X axis at $p<0.05$ level of significance.

A significant interaction was also found between the objects and the orientations $F(15,210)=3.789$, $p=0.0001$. Figure 37 below shows the mean reaction times to each of the objects across the different priming conditions. (It could be suggested that the interaction is caused by the difference between the objects at 0° orientation and possibly at 30° also. A post-hoc pairwise comparison proved impossible to do due to the large data set.)

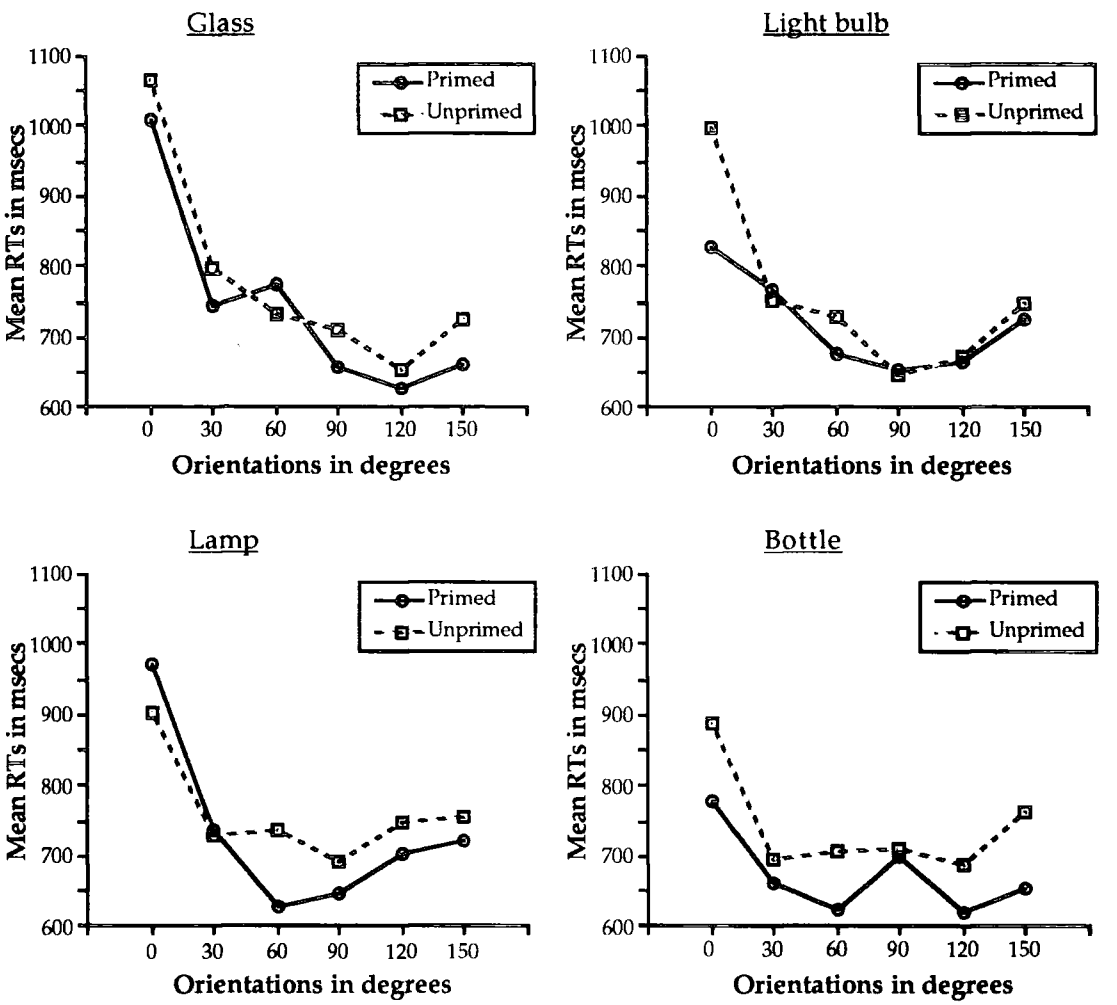

Figure 37: Mean reaction times to each object shown in different orientations across all other conditions.

A three-way significant interaction was found between the priming, objects and orientation conditions, $F(15,210)=1.716$, $p=0.0497$. As this interaction was barely significant it was suggested that the effect was probably due to the difference between the objects in the different orientations as was suggested by the previous result shown. An interaction was also

found between the different axes and the orientations, $F(10, 140)=3.818$, $p=0.0001$.

A post-hoc Newman Keuls analysis revealed that the 0° orientation of the X, Y, and XY axis was significantly different from all other orientations except each other at $p<0.01$ level of significance. The 30° orientation of the X axis was significantly different from the X axis 120° and the Y axis 90° orientations at $p<0.01$ level of significance. At $p<0.05$ level of significance X axis 30° was different from X axis 90°, XY axis 60°, 30°, 90° and 120° and Y axis 60° and 120°. Orientations of 30° in the Y axis were significantly different from X axis 120° and Y axis 90° at $p<0.05$ level of significance.

A significant three-way interaction was found between the priming condition, axes of rotation and angles of orientation, $F(10, 140)=2.442$, $p=0.0102$. It could be suggested that this interaction was based on the orientation effect in the X axis in the unprimed condition only, as a previous interaction between the axes and priming conditions was also significant.

The objects, axes and orientations factors also produced a significant three-way interaction, $F(30,420)= 1.504$, $p=0.0451$. Again this interaction may be as a result of the mean reaction times to orientations in the X axis.

Also, a significant four-way interaction between all the factors (priming, objects, axis of rotation and orientations) proved significant, $F(30, 420)=1.876$, $p=0.0040$. The mean reaction times to orientations in the X axis in the unprimed condition only may contribute to this significant interaction.

4.3.3 DISCUSSION

The results seem to suggest that when the direction of the object is given prior to recognition, this does affect the overall recognition speed but does not affect the relative recognition speed across the different orientations. The overall speed up may be due to the subject anticipating the onset of the next stimulus because of the preceding prime. The fact that the prime itself did not change the function relating recognition time to orientation suggests that an object-centred reference frame is not transformed but that a template-like image of the object is transformed to match the nearest stored view. A facilitation effect for the primed condition was expected if the representations of the objects in memory were based on resolving the principal axis of the object in order to describe the objects coordinates relative to this axis. An increase in reaction times across different views may reflect the difficulty in resolving this axis. However, if the appropriate reference frame was primed prior to the object onset, then the time taken to build a description of the object in coordinates relative to this axis should have been uniform across the different orientations. As this was not the case, then it could be argued that representations are not built around object–centred coordinates based on the object's principal axis.

However, there may be other reasons why there was no difference found between the orientations in the primed and unprimed conditions. Cooper and Shepard (1973) found that priming the orientation of the proceeding stimulus had no effect on rotation times unless the identity of the stimulus was also primed. This suggests that transformations are specific to the object itself. Humphreys and Quinlan (1988) also supported the finding that orientations cannot be primed unless the identity of the object is also given. Bartram (1974) found that

> "unless the priming information specifies the critical perceptual features, there is no gain in performance".

However, a model of object recognition that includes transforming the object by a process of, say, mental rotation to match to the nearest stored view would have to assume that this process occurs without prior knowledge of the objects identity.

The length of the ISI was not manipulated which may also contribute to the cause of the null orientation effect across priming conditions. Allport et al (1985) found that SOAs of above 200 milliseconds caused a decrease in recognition times relative to short SOAs when subjects had to match two views of the same object. This differential effect in matching times caused by the different SOAs in the Allport et al. study may suggest that a differential effect could have been found with different ISIs in the previous experiment.

It could be suggested that the overall speed-up may have been due to the subjects anticipating the onset of the following stimulus. As there was no effect of the prime on the recognition times across the different orientations, then it can only be assumed that the prime itself had no effect on the actual recognition of the following stimulus. In light of this result, the unprimed stimulus should not have been preceeded by a blank stimulus.

The differential reaction times to unprimed orientations of the objects in the X axis only is probably not a typical result because this differential effect was not observed in any of the previous experiments in this thesis.

4.4 General discussion

The results of Experiment 5 support the findings of Biederman and Ju (1988) that surface-based information is not necessary to derive a representation of an image in order to match it to a stored representation. The findings seem to suggest that information about the edges of the objects is important to recognition as was proposed by Marr and Hildreth (1980). The results also support the notion that representations in memory are a collection of 2-D views of the objects to which the inputted images are aligned and the degree of the match between the view of the image and the stored representations was reflected in the reaction times.

It seemed that the surface information which indicated the orientation of the object

in space was not exploited by the visual system. The projected image of the object relative to the viewer, irrespective of its orientation in the environment, seemed to be the only information required. It could therefore be asserted that cues to the object's reference frame are not important. It is the match between the input and the representation that is important. It is the number and kind of 2-D shape of representations in memory that is important, not the orientation of the image per se. The view of the object is chosen not for its orientation but for other reasons such as the amount of information that is available in that particular view of the object or the familiarity of the view for example (Jolicoeur, 1985; Larsen, 1985).

The findings from Experiment 6 support the view that object representations are not built upon the object's intrinsic reference frame. If that were the case then the recognition of objects in different views would be affected by advance knowledge of the orientation of the reference frame such that there would be no difference in the time taken to recognise the object across the different views. As this was not found it was asserted that representations are not transformed according to their reference frames, but that a 2-D image of the object is transformed. Other studies which have tested the effect of priming the orientation of the test stimulus have also found null effects unless the identity of the stimulus was also given in advance (Shepard and Cooper, 1973 and Humphrey and Quinlan, 1988).

Although the results of Experiments 5 and 6 indicate that images are matched to their corresponding stored representations by aligning the image with the stored view and not by extracting information about the orientation of the object in 3-D space, the results do not indicate why specific views are stored as representations over other views. The following chapter addresses this issue.

Chapter Five

The Recognition of Unfamiliar Objects

5.1 Experiment 7

In the previous chapters a number of experiments were reported which studied the effects of orientation in different axes of rotation on the recognition of objects and it was found that recognition times were slower to views that were fully foreshortened (0°) and views that were 30° off this foreshortened view than other views which contained more information about the principal axis. What does this mean? It was concluded from the results of the previous experiments that the increase in recognition times was due to the time needed to transform or normalise the object in order to align it with its nearest stored view. However, an alternative answer to this question could be proposed. It may be that the objects are simply unfamiliar in these orientations. This would seem unlikely given that there was no observable difference found in the orientation effect between subjects that were highly practised at the views of the objects and subjects that had not seen the objects prior to the experiment (see Figures 12 and 16, Chapter 3). Also, the overall effect of practice within Experiment 5 did not change the orientation effect but merely caused an overall speed-up in the reaction times (see Figure 31, Chapter 4). The experiment discussed in this chapter addresses these issues and investigates the role of familiarity on the representation of objects in memory.

It could be argued that the results observed in the previous experiments may be peculiar to the set of stimuli used. The objects used were familiar objects (Snodgrass and Vanderwart, 1980) and therefore, by virtue of the fact that they are highly over-learned stimuli they may have become immune to the effects of some orientations (see Jolicoeur, 1985 and Tarr and Pinker, 1989). In other words, the effects observed may be due to over-learning of the orientations rather than having representations of those views. This problem was addressed in Experiment 2 where the subjects were trained on a different set of objects other than the experimental objects. The same orientation effects were observed in Experiment 2 as in other experiments where the subjects were highly practised on the experimental stimuli. Although this suggests that the orientation effects observed with familiar objects are not dependent on previous practice with the different views it does not explain whether the orientation effects are attributed to the representation of familiar views. The results may reflect some property of the nature of representation of highly familiar objects.

Many studies on the recognition of disoriented objects have asserted that representations are built around the most familiar views of the objects and are therefore

view-centred rather than object-centred (Rock, 1973; Rock and DiVita, 1989; Jolicoeur, 1985; Larsen, 1985; Tarr and Pinker, 1989; Edelman et al, 1989; Edelman and Bülthoff,1990). Rock (1973) proposed that at least under some conditions, objects are not represented by 3–dimensional object models. He found that subjects who were previously trained to recognise 3–dimensional wire-frame objects found it difficult to recognise these objects when rotated in depth by 90° (Rock et al, 1981). In a further study, Rock and DiVita (1987) found that subjects who were trained to recognise 3-dimensional wire-frame objects in a particular position had difficulties in recognising the same objects when shown in a different position in the display. This difficulty was most evident when the change in position also changed the retinal projection of the objects. They concluded that the subjects perceived these objects relative to an egocentric reference frame. In other words a viewer-centred description was employed for recognition purposes, not an object centred description. Recognition of novel objects was therefore constrained by the most familiar view and position of the objects.

A view-centred model of object representation predicts that recognition times are dependent on the disparity between the input view and the stored view. Moreover, no single, preferred view should exist for objects that are likely to be seen in any orientation. Familiarity of the views should therefore determine the views that are stored as representations of the object. Indeed previous studies have reported that the difference in time taken to recognise common objects, stick-like figures or wire-frame objects in different orientations was reduced with practice (Jolicoeur, 1985; Tarr and Pinker, 1989; Rock et al, 1981; Edelman, Bülthoff and Weinshall, 1989).

Edelman et al. (1989) looked at the effect of familiarity on recognition times by training subjects on different sets of restricted views of novel, 3D wire-frame objects. They found that initial recognition times were dependent on the views of the objects seen in the training session but practice had the effect that the recognition times became more uniform across objects and that the view most quickly recognised (which was related to the trained view) disappeared with practice. The variation of recognition times over the different views was found not to depend on the stimulus complexity. These results would seem to suggest that the advantage of some views over others is linked to the familiarity of those views. The more familiar a view is then the more salient that view will be in memory. This conclusion is also supported by the work on the recognition of stimuli that are highly learned in one orientation such as faces (Yin, 1969 and Diamond and Carey, 1986), dog breeds (Diamond and Carey, 1986), alphanumeric characters (Jolicoeur and Landau, 1984; Cooper and Shepard, 1973) and common objects (Jolicoeur, 1985). Although many of these studies reported with–in category effects, in general, these data seem to clearly support the notion that familiarity determines the views of objects that are stored.

The findings on orientation studies by and large imply that familiar views or orientations of objects are favoured for recognition purposes. Yin (1969) reported that people find it very difficult to recognise faces that have been inverted. It was subsequently revealed

that these findings were not specific to faces and did not reveal a peculiarity of face representation but that an increase in the familiarity of the views rendered recognition more sensitive to orientation. Diamond and Carey, (1986) replicated Yin's findings with inverted faces but also found that dog trainers were less able to recognise dog breeds when inverted. They concluded that recognition is sensitive to orientation when perceivers are experts at representing objects in certain orientations. The representation of dogs for dog breeders is different than the representation of dogs for people who are not dog experts.

The orientations examined in most of the studies reported above were in the picture plane. Information about the objects themselves remains constant with rotations in the picture plane. It could be suggested from the studies be that as long as the same information is available, in any orientation, that recognition times will be equal. For example, if views of objects are stored that are spaced equally around possible orientations in the picture plane, then there should be no difference in recognition times observed across these views. The experiment reported in this chapter addresses the issue of familiarity by looking at the effects of orientation in depth of nonsense, 3D objects.

From the results of the previous experiments it was concluded that objects are represented by a number of stored views and that novel views are transformed to match the nearest stored view. It was argued that these stored views represent the views that contain the maximum amount of information about the object. Therefore it could be predicted that for novel objects, representations will be built around the views which contain the maximum amount of information about the object and not necessarily the most familiar views. However, familiarity should have an initial effect in that the most familiar view should be the most recognisable view of the object, irrespective of the type of view and the amount of information it contains. However, as more views of the object become available, the visual system can store representations that hold the maximum amount of information about the object. Therefore, if subjects are trained on a set of elongated, novel objects shown in specific views then an initial effect of the most familiar view should be observed. With practice on other views of the objects however, the visual system should build representations that contain the most information about the axis of elongation. In other words the results of recognising novel, elongated objects should show the same effects as those found for familiar, elongated objects.

For these reasons it was decided to test the effects of familiar views on the recognition of novel objects seen in a variety of different views. The experiment neatly separated familiarity from informativeness by training the subjects on two different views of four novel objects one of which was highly informative and the other more foreshortened and therefore less informative. These views were therefore as familiar as each other. The views selected corresponded to the 90° and the 30° views in the previous experiments (see Figure 8, chapter 3). Figure 38 below indicates the different views that were presented to the subjects in both the training and the test phases of the experiment.

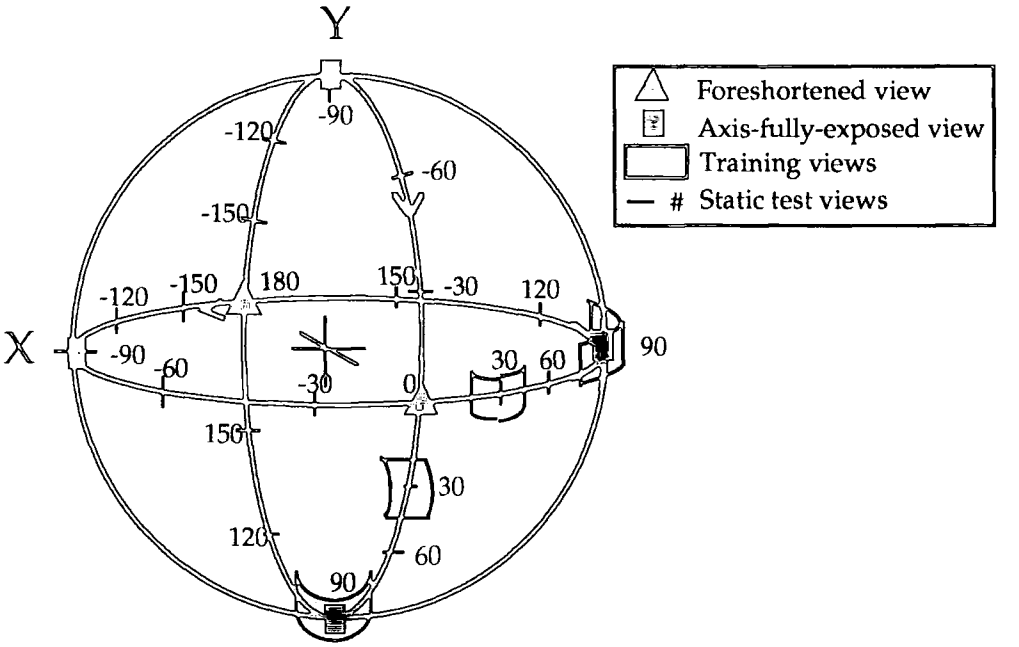

Figure 38: Illustration of the training and test views of the objects used in Experiment 7. The training views consisted of the objects oscillating ±10° around the following views; 30° and 90° in the X axis and 30° and 90° in the Y axis. The test views were static views of the objects seen in 4 different quadrants; the positive and negative quadrants of both the X axis and Y axes.

This experiment was designed to test the effect of familiarity on the nature of the representation of a 3-dimensional object in memory and to test the generalisability of recognition from familiar to novel views. From the results of previous experiments it was noted that there was a difference in recognition times to objects viewed with their elongated axes fully resolved and 30° either side in comparison with objects viewed 30° from the foreshortened view. It was argued that the visual system represents a number of views per object and that recognition proceeds by transforming each new view of an object to the nearest stored view. However, there is a strong suggestion that information about the principal axis, particularly in elongated objects, is important for recognition and therefore the stored views of the objects hold maximum information about the object. The results of the previous experiments yielded a highly consistent finding that elongated objects are recognised faster when information about the principal axis is available. It could be suggested that the visual system is organised in such a way that views of elongated objects which include the maximum information about the axis automatically represent the object in memory and that novel views are normalised (transformed) to match the nearest stored view. A further aim of this present experiment was therefore to test whether it is characteristic of the visual system to store views of elongated objects with the maximum information about the object when all views of the object are equally familiar.

In the experiment the subject was initially trained to match a specific label to each of 4 different, novel objects. The labels used were four lettered, nonsense words. The objects

were initially shown in motion between the orientations of 80° and 100° in either the X or Y axis and also between 20° and 40° in either axis. This motion sequence was achieved using the More package which presented the different orientations of the objects to move between two different orientations on the screen (see Figure 38 above for an illustration of the views). A 3-dimensional appearance of the objects was created due to the kinetic-depth effect. It was thought that the subjects were more likely to create 3-dimensional models of the objects if the objects were perceived as 3-D rather than 2-D. The subjects were subsequently tested on static views of the objects in two different experimental stages. They were initially tested on static orientations seen in the training session. This stage also acted as a learning stage and the errors and reaction times were recorded in this block. In the experiment proper, the subjects were tested on novel views of the objects in three different conditions of orientations: 1) the SAME condition, where the orientations of the objects shown included the orientation of the objects that the subjects were trained on, 2) the -SAME condition, where the test orientations included the negative orientations of the practice views, for example, if the object was shown in 90° in the practice block it would be shown in the -90° (or 270°) orientation in this condition, 3) the ORTHOGONAL condition, where the objects were viewed in the same orientations as the practice trials but in the opposite axis and 4) the -ORTHOGONAL condition, where the views included the negative orientations of the practice views in the opposite axis. The task was a match/mismatch one where the subject had to decide whether the label shown over the object shown was the correct name of that object. The task used was the same as that used in the previous orientation experiments.

According to the 2–D, multiple view model of representation (Tarr and Pinker 1989, Edelman and Bülthoff 1990), there should be an initial decrease in recognition rate and accuracy the further the test view is from the trained views. During the course of the experiment other views of the objects become more familiar, therefore recognition times were expected to become more uniform. However if, as was argued from the results found in the previous experiments, it is characteristic of the brain to store representations of elongated objects which include the maximum information about the principal axis of objects then training subjects on the 30° views of the objects should not be faster than views which include more information about the principal axis once all views are familiar.

5.1.1 METHOD

Subjects

Sixteen members of the University of Durham which included students and members of staff participated in this experiment without pay. The age range of the subjects was 21 to 32. All subjects had normal or corrected-to-normal vision. Six of the subjects were female and ten were male.

Materials and Stimuli

A set of four 3-dimensional stimuli was drawn using a 3D object-oriented drafting package called Swivel 3D. These stimuli were presented on the screen with a white background and the stimuli were displayed in shades of grey. This package allowed the objects to be viewed in any orientation around the X and Y axes.

The objects are shown in Figure 39 below. Two of the objects drawn had symmetrical axes parallel to the elongated axes and the other two objects were asymmetrical. Each object was paired with an arbitrary label which the subjects learned to associate with the object in the training phase. The labels used were as follows; Ress, Wike, Kolb and Chup. Each object stimulus was paired with an appropriate label in the training session. In the experimental stages each object stimulus was paired with either the appropriate label or an inappropriate label which was one of the other object names. The labels of the objects were placed above the object drawing and were seen in that position for all trials.

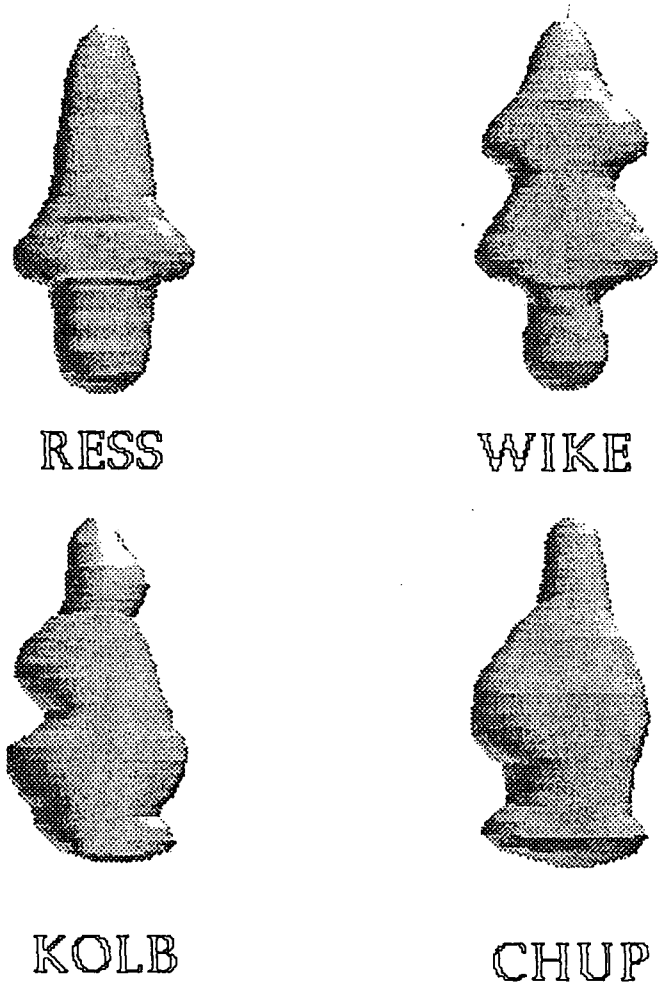

Figure 39; Novel objects used as stimuli in Experiment 7. Two of the objects were designed with elongated axis which was also the axis of symmetry and two objects were asymmetrical along the elongated axis. The length of all objects' elongated axes was equal.

Apparatus

A Macintosh IIx micro-computer was used to display the stimuli using a display and recording package designed for the Macintosh (see Chapter 3 for a description) The subject responded by pressing a key on the keyboard as soon as a decision was made. The Macintosh recorded the reaction times and the response type made to each trial.

Design

The experiment consisted of three different sessions; the training session, testing of the views shown in the training session and the main experimental blocks.

Training session

The subjects were randomly assigned to one of four training session groups. All subjects were trained on the 30° view in both the X and Y axes and the 90° views in both axes (see Figure 38 above) with the narrow end of the object pointing towards the subject. The objects were nested under these factors. The objects were assigned to each factor according to a Latin based design which yielded four different groups of training sessions. There were four subjects assigned to each of these groups so group 1, for example, received objects 1 and 2 shown in orientations 90° ±10° in the X axis and objects 3 and 4 in orientations 30°±10° in the Y axis. The orientations, axis of orientations and objects themselves were counter-balanced across all four groups of subjects.

Post-training test

Static views of the objects shown in the training session were presented in a random sequence in the post-training test. The objects were shown in the orientations that were given in the training session, for example, if object 1 was shown in the 30°±10° view in the Y axis then the subject was tested on static views of the object within that orientation range. No new views of the objects were therefore seen in this session. Twelve views of each object were shown. Copies were made of these stimuli and the objects were paired with inappropriate labels. Based on a match/mismatch design, the subject had to decide as fast as possible whether the label shown with the object was the name given to that object in the training session. There was a total of 96 trials in this session. This session also served as a learning phase in that the views of the objects were repeated and therefore increased their familiarity.

The main experiment

The experiment was based on a four way factorial design, with objects as a nested factor. The four main factors included the conditions of orientations, the previous training views, the axes of rotation and the orientations.

There were four levels to the conditions factor: SAME (orientations were 0° to 150° i.e. including orientations shown in the training session); - SAME(orientations were -0° to -150° i.e. including the negative of those in the training session); ORTHOGONAL

(orientations were 0° to 150° in the opposite axis) and finally - ORTHOGONAL (orientations were -0° to -150° in the opposite axis). The trained views factor contained two different levels; objects that were originally trained on the 30° view and objects that were trained on the 90° view. The objects used were a nested factor under this condition and were counter-balanced across the two training view levels. The axes of rotation factor were the X and Y axes respectively. These axes corresponded to the same axes of rotation as were used in the previous experiments (see Figure 38 above). Finally the orientations factor included six different orientations; 0°, 30°, 60°, 90°, 120° and 150°.

The experimental procedure was based match / mismatch design experiment where the subject had to decide whether the name presented over the object was either the correct or incorrect name of that object.

An experimental block contained 192 trails, 96 were match trials and 96 mismatch trials. The conditions, objects, axes of rotation, previous trained views and orientations were counter-balanced across the experiment. The experimental block was repeated three times in order to test the effects of practice on the recognition of the novel views across the experiment.

Procedure

The subjects were initially presented with the training session in which the set of test objects were shown oscillating ±10° around the training view with the correct label shown above the object. Subjects were randomly assigned to one of four experimental groups according to the design constraints of the experiment. This training session was self-timed and the subject ended each object learning session as soon as they decided that the object and its associated name were familiar.

The post-training session immediately followed the training session. In this session subjects were tested on the recognition of static views of the objects shown within the range of orientations in the training session. The static views corresponded to 20°, 30° and 40° views of the objects that were shown oscillating around 30° in the training session and 80°, 90° and 100° views of objects that were shown oscillating around the 90° view. The subjects received no feedback during this or any of the subsequent sessions. Subjects were instructed to respond as fast as possible to each trial without making too many errors. The reaction times and errors were recorded for this session.

The final experimental session consisted of testing the recognition of views not previously seen in the training sessions. The total number of trials in the experimental block was 192. This experimental block was repeated 3 times in the experiment. The task for the subject was to match the label shown over the object with the object and they were instructed to respond as fast as possible without making too many errors to each trial. A response consisted of depressing a 'match' key on a response box if the subject decided that the label

was the appropriate name of the object shown or the 'mismatch' key if the label was not the name of the object. The reaction times and errors were recorded for all subjects.

5.1.2 RESULTS

The results for the test of the training session (post-training block) and the results of the experimental blocks were analysed separately.

Post-training block results;

The mean number of errors made per subject in the training sessions was 6 which gave a proportionate number of 6.25% across both match and mismatch trials. The mean reaction times per subject was 1297 milliseconds in the training session.

A one-way, repeated measures ANOVA was conducted on the reaction times to the different orientations of the objects in the test condition and was not significant, $F(5,640)=0.872$, $p=0.4993$.

Experimental block results;

The mean percentage of errors made to the match trials only per subject were as follows; 11.84% in block 1, 9.83% in block 2 and 7.75% in block 3. A Friedman ANOVA was conducted across the error counts and proved significant, Chi_r Squared= 10.906, $p=0.0043$. Figure 40 below illustrates the percentage of errors made to each of the conditions of orientation. There were more errors made to the 0° than other orientations across all experimental conditions.

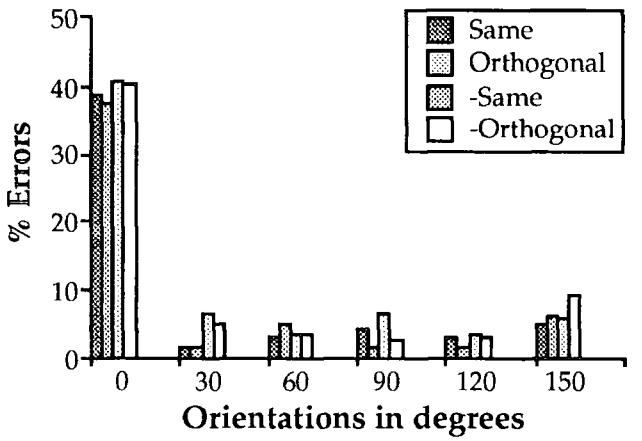

Figure 40: Percentage of errors made to each orientation in each condition of rotation.

Initial effect of training on recognition

In order to assess the initial effect of the trained views on the recognition of the objects shown in other views, reaction times to the different views of the objects in the Same condition, block 1 only were analysed. This small set of data was analysed for two reasons; (a) the trained view only occurs in one of the experimental conditions i.e. the Same condition

and (b) learning of the other views over the experimental blocks may affect the recognition times in such a way that a facilitation for the trained view may have been obscured.

Figure 41 below shows the mean reaction times across the match trial for objects that were initially trained on 30° or 90° viewing positions. A three factor ANOVA was conducted on the reaction times across subjects with trained view, axis of rotation and orientations as factors. The trained view factor contained two levels, the 30° and 90° trained views. There were two axes of rotation, X and Y axes. Finally for the purposes of this analysis, the fully foreshortened 0° orientation was removed and the analysis was performed using data to the 30°, 60°, 90°, 120° and 150° orientations. There was no significant effect found for trained view; $F(1,15)=0.707$, $p=0.4135$, axis of rotation; $F(1,15)=0.709$, $p=0.4130$ or orientations; $F(4,60)=0.993$, $p=0.4184$. There was no interaction found between the trained view and the orientations, $F(4,60)=1.006$, $p=0.4116$.

From Figure 41 it can be seen that the mean reaction times to the 30° orientations were in fact faster than reaction times to other orientations of the objects initially trained in that view (although not significantly faster). Reaction times to the 90° view for objects that were initially trained on that view did not prove to be any faster that reaction times to other views.

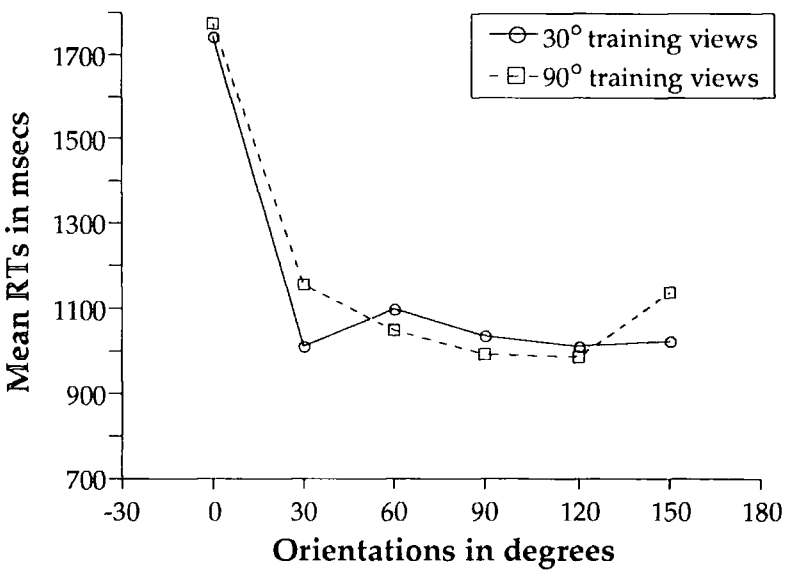

Figure 41: The effect of the trained view on initial recognition times to the objects shown in the Same condition in Block 1.

The main experiment

The reaction times to the match trials in each of the experimental blocks were subjected to a 4 way, repeated measures ANOVA with condition, trained view, axis of rotation and orientations as factors[1]. The conditions factor contained 4 different levels; Same axis and

[1] Due to the large data set it proved impossible to analyse all of the data together. The data was therefore analysed separately per block.

orientation (Same), Same axis but negative orientation (Same-), Orthogonal axis (Orthogonal) and Orthogonal axis negative orientation (Orthogonal-). The trained view factor contained two levels; the 30° view and the 90° view. The axes of rotation were the X and Y axes. The orientations included 0°, 30°, 60°, 90°, 120° and 150°.

The analysis of variance across the data in each block showed no effect of condition in block 1, $F(3,45)=1.733$, $p=0.1738$, a main effect of condition in block 2 $F(3,45)=3.167$, $p=0.0333$ and no effect in block 3, $F(3,45)=1.105$, $p=0.3569$. Figure 42 below shows the mean reaction times to each condition across all blocks. This figure also indicates that the overall reaction times to the trials in the first block were slower than the reaction times in the other two blocks. A Newman-Keuls post-hoc analysis on the condition effect observed in block 2 indicated that the reaction times to the orthogonal condition were significantly slower than the reaction times to both the Same and -Same conditions at $p \leq 0.05$ level of significance.

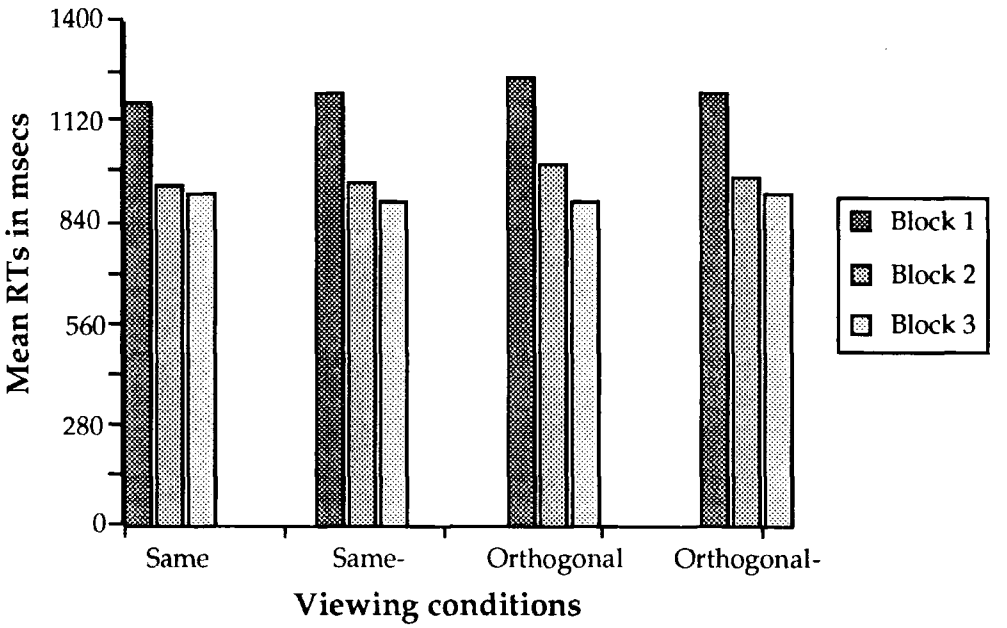

Figure 42: Mean reaction times across all subjects to the different conditions of orientation in each of the experimental blocks.

There was no main effect of trained view found in any of the blocks; $F(1,15)=0.469$, $p=0.5038$ in block 1, $F(1,15)=0.000$, $p=0.9866$ in block 2 and $F(1,15)=0.001$, $p=0.9724$ in block 3. It was predicted that the trained views would have a differential effect on the reaction times to the different orientations in each condition therefore the interaction between the conditions, trained views and orientations was looked at in each block. There was no interaction found between the three factors in block 1; $F(15,225)=0.992$, $p=0.4643$, no interaction in block 2; $F(15,225)=0.687$, $p=0.7964$ and no interaction in block 3; $F(15,225)=0.453$, $p=0.9609$. Figure 43 below shows the mean reaction times to orientations in each condition when the subjects were previously trained on the 30° view of the objects. Figure 44 below shows the mean reaction times to the different orientations within each condition when the subjects were trained on the 90° view of the objects. There is little observable difference

between the two figures and because there was no interaction found between the factors it can be assumed that the trained views had no differential effect on the recognition of the objects across the experiment.

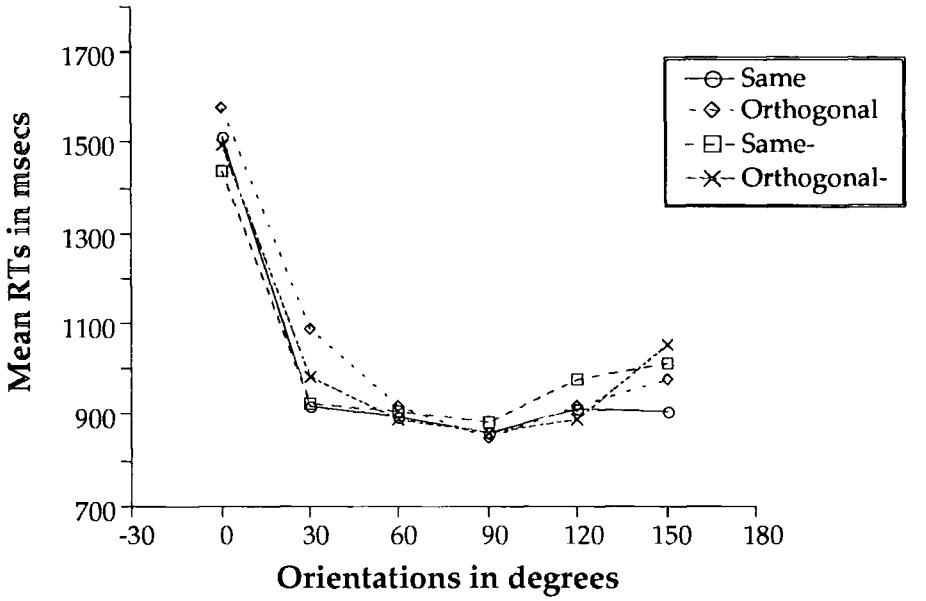

Figure 43: Mean reaction times to the different orientations in each of the conditions across all objects that were shown in a 30° ± 10° view in the training block.

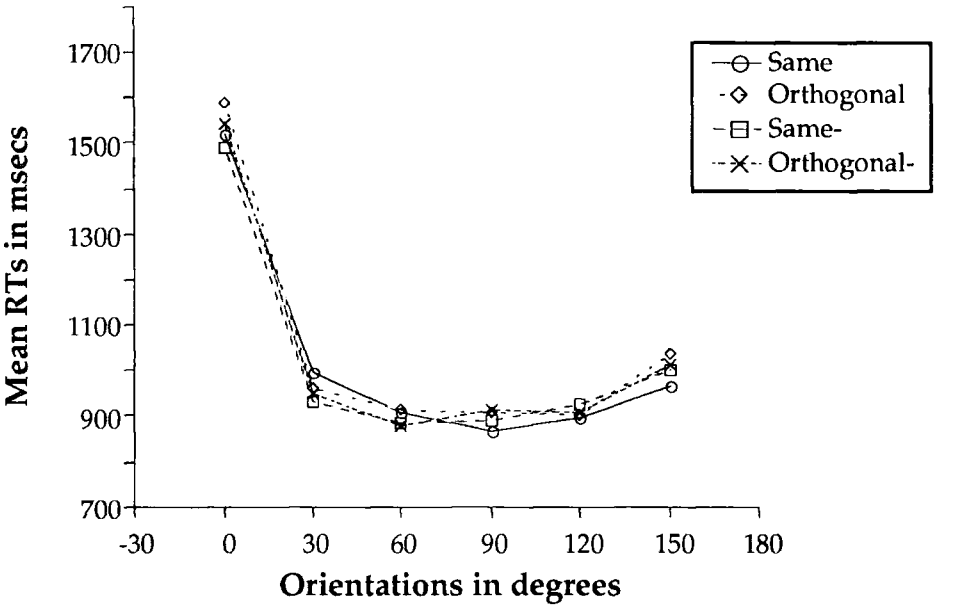

Figure 44; Mean reaction times to the different orientations in each of the conditions across all objects that were shown in a 90° ± 10° view in the training block.

A significant main effect of orientation was found in all blocks; $F(5,75)=54.227$, $p=0.0001$ in block 1, $F(5,75)=44.494$, $p=0.0001$ in block 2 and $F(5,75)=36.586$, $p=0.0001$ in block 3. Figure 45 below shows the mean reaction times across orientations in each block. A post-hoc

Newman-Keuls analysis on the main effect of orientations in each block were conducted. In block 1 there was a significant difference between the reaction times to 0° orientation and any other orientation at $p \leq 0.01$ level of significance and reaction times to 150° orientations were significantly slower than reaction times to 90° at $p \leq 0.05$ level of significance. A significant difference between 0° orientation and any other was again found in block 2 at $p \leq 0.01$ level of significance. The reaction times to the 0° orientation in block 3 were significantly slower than reaction times to any other orientation at $p \leq 0.01$ level of significance. There were no other differences found within the orientation effect.

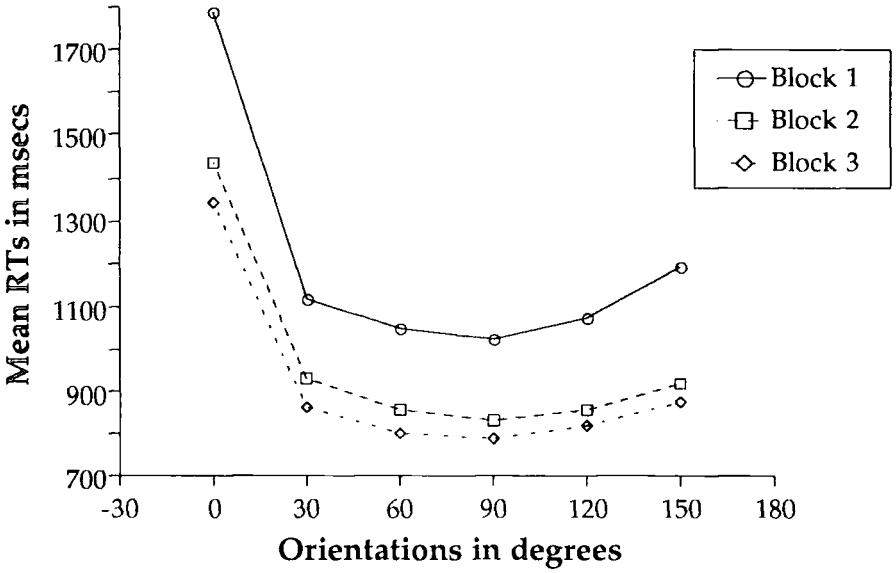

Figure 45: Mean reaction times across all objects shown in different orientations in each block.

An interaction between the axes and the training views was found in block 1 only, $F(1,15)=7.068$, $p=0.0179$. There was no interaction found in block 2, $F(1,15)=4.004$, $p=0.0638$ or in block 3, $F(1,15)=1.010$, $p=0.3308$.

An interaction between the conditions and orientations proved significant in block 2, $F(15,225)=2.213$, $p=0.0068$. This interaction was not found in block 1, $F(15,225)=0.811$, $p=0.6651$ or in block 3, $F(15,225)=0.922$, $p=0.5400$.

There was a main effect of axes found in block 3, $F(1,15)=4.549$, $p=0.0499$. This main effect was not found in any of the other blocks; $F(1,15)=0.432$, $p=0.5211$ in block 1 and $F(1,15)=0.4$, $p=0.5368$ in block 2. In block 3 a significant interaction between axes and orientations was also found, $F(5,75)=7.083$, $p=0.0001$. This interaction was not found in the other blocks; $F(5,75)=0.804$, $p=0.5501$ in block 1 and $F(5,75)=0.332$, $p=0.8922$ in block 2.

The data for each block was re-analysed with the reaction times to the 0° orientation removed. In block 1 the interaction between the trained views and the axes of

rotation was again found, F(1,15)=6.697, p=0.0206. The orientation effect also proved significant, F(4,60)=5.422, p=0.0009. A post-hoc Newman Keuls analysis revealed that 150° was significantly slower than all other orientations at p<0.01 level of significance. No other differences were found. However, a planned post-hoc orthogonal comparison between the grouped 30° and 150° and the grouped 60°, 90° and 120° proved significant, F(1, 60)=16.955, p=0.0001. There were no other effects observed in the re-analysis of block 1.

A re-analysis of the data in block 2 again yielded a significant interaction between the conditions and orientations, F(12,180)=2.563, p=0.0037 although the effect for conditions disappeared, F(3,45)=1.819, p=0.1572. A significant main effect for orientation was also found, F(4,60)=8.790, p=0.0001. A Newman-Keuls post-hoc analysis on the orientation effect revealed that the 30° was slower than all other rotations except 150° at p<0.01 level of significance and also that 150° was slower than all other orientations except 30° at p<0.01 level of significance. There were no other differences found.

Finally, a re-analysis of the data in block 3 removed the interaction between the axes of rotation and the orientations, F(4,60)=0.602, p=0.6626. A main effect of orientation was found, F(4,60)=6.294, p=0.0003. A post-hoc Newman-Keuls analysis revealed that 150° was significantly slower than 90° and 60° at p<0.01 level of significance and slower than 120° at p<0.05 level of significance. It was also found that 30° was slower than 90° at p<0.01 level of significance and also slower than 60° at p<0.05 level of significance. However, a planned post-hoc orthogonal comparison between the grouped 30° and 150° orientations and the grouped 60°, 90° and 120° orientations proved significant, F(1,60)=22.382, p=0.0001.

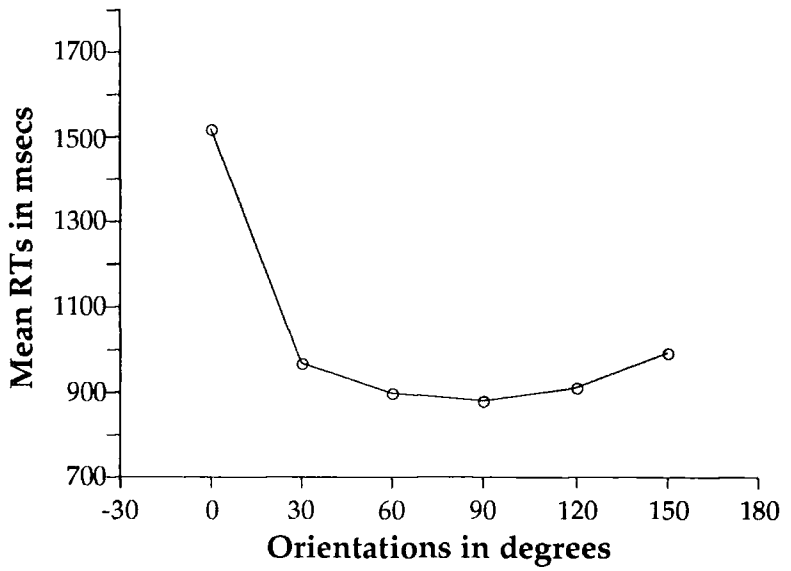

Figure 46: The overall mean reaction times to all objects shown in different orientations across all blocks in the experiment.

The overall recognition times to the objects shown in the different views is shown in

Figure 46 above. As indicated from the analysis on each separate block, the time taken to recognise the objects when viewed 30° from the fully foreshortened view (0° or 180°) is longer than the time taken to recognise views which are closer to the 90° or axis-fully-exposed view.

5.1.3 DISCUSSION

In sum, the effect of familiarity on the recognition of novel views of objects was initially observed but recognition times became more uniform as other views became familiar during the course of the experiment. The views that the subjects were trained on seemed to be recognised more quickly than other views although this advantage was not statistically significant. As other views of the objects became familiar during the course of the experiment proper, recognition times to some of the novel views became more uniform. Subjects still found it difficult to recognise the objects when shown in foreshortened views. This result is hardly surprising given that there is very little information available about the objects in these views. The recognition of the objects shown 30° from the foreshortened views were significantly slower than the recognition of other views around the axis-fully-exposed views in each of the experimental blocks. This result replicates the findings from the previous experiments which tested recognition times to different views of familiar objects.

This result may indicate that there is a characteristic way in which the visual system organises information about objects that is somehow dependent on the information that is available about the object's principal axis. In other words, it seems that the visual system can extract information about the object in order to represent it in its optimal or canonical views which include the maximum information about the principal axis. Palmer et al. (1981) argued that the visual system stores views which maximise the amount of salient information about the object. The results of this experiment suggest when all views are equally familiar, a number of views are stored as representations of the object. This can be asserted because according to Palmer et al. recognition times are fastest to the canonical view of the object. Recognition times are equal to a number of views (60°, 90° and 120°) across different quadrants and different axes, therefore there is not a single canonical view but rather multiple, canonical views. These views have a collective canonical aspect in that they are views which maximise the salient information about the objects.

The multiple–stored views model was proposed to alleviate the problems encountered by the single canonical view model (computational demands) or the template model (memory demands) (Ullman, 1989; Tarr and Pinker, 1989; Edelman et al,1989, 1990; Bülthoff and Edelman, 1992). A trade-off between the amount of memory invested in storing object representations and the amount of time or computation required to normalise a novel view is therefore implicit in such a model. However, if the visual memory stored all of the views that were equally familiar, this would be a return to the template model and the problems of memory capacity are again encountered. It would seem advantageous for the

visual system to store views that contained more salient information about the object rather than views that stored a minimum amount of information. If all such views were stored, recognition would be in chaos. For example, how could we recognise a wine bottle (viewed end-on) from a circle. Unless we could resort to other information such as stereopsis or environmental depth cues then the bottle would never be recognised in this position. It is therefore unnecessary to store views that contain minimum amounts of information about the objects.

The results do not support the object-centred 3-dimensional models proposed by Marr or Biederman (Marr, 1982; Biederman, 1987). This can be asserted for three reasons, 1) the familiarity of the view of an object affected the time taken to recognise that object in different views, 2) recognition did not generalise over different views despite the fact that all subjects reported to have seen 3-D versions of the objects in the training session, and 2) the final pattern of results did not indicate that recognition was invariant over orientation.

This ability of the visual system to generalise across to novel views of the object seems to be quite versatile. There was little difference found between the recognition of novel views that were in the opposite axis of rotation, in the negative orientation and views that were closer to the trained view i.e. in the same condition. There was a difference found between the different conditions in the second block but this effect was not robust across the experiment.

Another interesting finding from this experiment was the pattern of recognition times to the different views in the SAME condition in the initial part of the experiment (see Figure 41 above). Recognition of objects previously seen in the 30° position in the training block were fastest indicating a preference for the most familiar view. However, as can be observed from Figure 41, recognition times to the 150° view were almost equally fast as the 30° view and faster than other views. This 150° view is the mirror image of the 30° (previously trained) view. This result may indicate that recognition of some views can generalise quite easily to mirror-image views. However, this is not conclusive because the views were extracted from among a larger set of views which may have had an affect on the recognition of these initially familiar views. Nevertheless, other studies have reported that discrimination between mirror-image stimuli is difficult suggesting that these images are treated as equivalent by the visual system (see Corballis, 1988 for a review).

5.2 Conclusion

The experimental work reported in this chapter set out to investigate the effect of familiarity on the representation of objects. It was found that subject's recognition times were initially affected by the most familiar or previously trained views of a set of novel objects.

However, as other views of the objects became more familiar during the course of the experiment, the effect of the trained views disappeared and recognition times became more uniform around the views of objects that were less foreshortened. Views that were more foreshortened (i.e. 30° off the fully foreshortened view) were found to be less readily recognisable than views that contained more information about the elongated axes of the objects. This result was found to be independent of the familiar view.

It was therefore concluded that the recognition times to different views of objects are affected by the most familiar views. However, as more views become familiar, this effect is lost and the fastest responses occur to the most informative views. This suggests that the visual system characteristically stores a number of views as representations of the objects. These views collectively represent the canonical aspect of the object in that they maximise the amount of salient information about the object.

Chapter Six

Introduction to Visual Search

In visual search tasks, the subject is required to locate a target from amongst a set of nontargets. This task involves attending to the visual scene, directing the eye to move to a target and deciding whether the object is the target or not. The question usually asked in visual search tasks is what variables influence search efficiency. The number of features that define a target or a distractor and the size of the visual display are traditional manipulations in visual search tasks. Research into visual search has highlighted a number of different variables that can effect search efficiency.

In one of the earliest studies on visual search, Schneider and Shiffrin (1977) found that search efficiency depends on learning. They proposed that information is processed automatically when a learned sequence of responses stored in long term memory is triggered to certain inputs into the visual system. This process does not then require direct attention. Automatic processing was shown to develop following consistent mapping of stimuli to responses over trials. However, when a sequence was not previously learned then information was processed in a controlled fashion in that the subject controlled the search process and attention was applied to each element in a display in a serial fashion. They therefore concluded that the difference between automatic detection of a target and a serial search for a target was due to the amount of practice or learning involved in the search task.

Visual search tasks typically involve two types of attentive processing; preattentive and focal attention (Neisser, 1963). Treisman (1986) argues that focal attention is needed in order to identify a target unless the target differs from the distractors by some simple feature such as colour, orientation or movement. In this case the target pops out of the scene. However, when the elements in the display become more complex then focal attention is required to locate the target. She termed these different search processes parallel and serial search. In parallel search the whole of the visual scene is monitored. In serial search, each element in the visual scene is monitored in sequence until the target is found. Treisman likened the serial search strategy to Posner's (1980) model of attention in which he compares attentional shifts to a spotlight mechanism where each item can be monitored in a serial fashion.

More recent theories have suggested that a search model which incorporates both serial and parallel search strategies is too restricted. The model proposed by Treisman (Treisman and Gelade, 1980; Treisman and Gormican, 1988) implies that the nature of the features in a visual scene determine whether the subject can monitor the entire scene and find

the target without monitoring its location or searching from one element to the other until the target is found. Many researchers find this account of the attentive processes too simplistic (Wolfe et al., 1989; Duncan and Humphreys, 1989; He and Nakayama, 1992). They propose that for some visual displays, search strategies are neither strictly parallel nor strictly serial. The different models of visual search are outlined below.

6.1 Visual Search Theories

In an attempt to model the processes which influence the nature of what is selected by the visual system for further processing at least three contemporary theories stand out. Treisman's feature integration theory is one of the most influential search theories to date. This theory centres around simple features which are processed preattentively by the early visual processes and therefore do not require selective attention. Attention is required however for items that are defined by a number of different features. This account, although influential, has a number of drawbacks which were recognised by people such as Wolfe et al. and Duncan. They, in turn, proposed their own theories on the selection processes. Wolfe et al. (1989) proposed that search is guided by the outputs of the early visual processing. Duncan and Humphreys (1989), on the other hand, proposed that search efficiency is dependent on what items are included in a visual display, particularly to the degree of similarity between what is being searched for and the other items in the display and also between the non–targets in the display. A more detailed account of these theories is given below.

6.1.1 Feature Integration Theory

In 1980, Treisman proposed her feature integration theory which stated that early visual processing involves the extraction of simple features in parallel and that the integration of these features requires focal attention on the location of the item. The features automatically extracted from a visual scene are assembled into meaningful wholes or objects through selective attention. In a prototypical experimental design in visual search, Treisman and Gelade (1980) displayed an array of visual stimuli to subjects who had to respond to the presence or absence of a particular stimulus. They found that there was a qualitative difference in the search patterns for targets that were defined by a single feature relative to the surrounding distractors and targets that were defined by a conjunction of features. For example, the time to search for a blue target relative to red distractors is independent to the size of the visual display with search slopes close to zero. Search slopes refer to the function between search times and the number of items in a display. Treisman claimed that if the target differs from the distractors in some simple property, the target is said to "pop out" of a visual display. This suggested to Treisman that these simple properties are detected preattentively in that their detection does not require focal attention. On the other hand, the time to search for a target which is defined by a conjunction of features such as colour and orientation from among a set of distractors which are made up of different conjunctions of the

features (e.g. a red horizontal line from among a set of blue horizontal or red vertical lines) increases linearly with display size. The search slopes for target present trials are about half that for target absent trails (because on average targets are found after half of the distractors have been examined) suggesting that searching for a conjunction of features is serial and self terminating. She also found that when attention was not focussed on a target that was defined by a conjunction of features, subjects would respond to illusory conjunctions between the features of the target and features from surrounding distractors (Treisman and Schmidt, 1982). However, it was also found that previous knowledge affected the number of instances of illusory conjunctions. For example, when subjects were told to expect shapes that corresponded to meaningful stimuli such as a blue lake or an orange carrot as opposed to another group of subjects who were told to expect arbitrary shapes such as a blue ellipse or an orange triangle, the latter group were much more likely to respond to illusory conjunctions than the former group. These results indicate that prior knowledge can influence attentional processes in conjoining features of objects. The important point therefore about illusory conjunctions is that they seem to occur prior to any semantic categorisation of the objects.

In a further series of experiments, Treisman set about cataloguing the list of simple properties that could be detected preattentively and thereby inferring which features are coded automatically in early vision. She tested the effect of exchanging the target and the distractors and found results consistent with what she calls search asymmetries. In other words she found an imbalance between the search functions to items that were targets in one example and distractors in another. For example, Treisman and Gormican (1988) contrasted pairs of stimuli in each of 12 different dimensions where each member of a pair played the role of a target in one test and a distractor in the other. Many of the pairs gave rise to search asymmetries in that one member of a pair was detected through parallel processing whereas the other was detected through an item by item serial search. They found that when two stimuli were paired in search tasks, the more extreme value of a particular feature would be favoured as the target whereas the lesser value would be more difficult to locate as a target. For example, in one such experiment Treisman and Gormican found that longer lines were found more quickly when displayed among a set of shorter lines than vice versa. They also concluded that when a particular feature is absent from a stimulus that such a target would be difficult to find from among a set of distractors which contained this feature. For example, a tilted line '/' was readily detected from among different numbers of vertical lines, 'I', whereas a vertical line among tilted lines was detected after a serial search strategy. Treisman and Gormican suggested that 'tilt' is a feature that pops out of a display due to its presence and that 'vertical' is coded as the absence of a feature and therefore does not pop out. Their experiments resulted in a list of feature dimensions which may function as primitive that are processed in early vision. These dimensions included colours, line curvature, line tilt and quantitative values such as line length and contrast levels (provided the differences between the values were sufficiently large. Treisman argued that her findings suggest that some properties are encoded as standard or default dimensions and therefore do not elicit responses from early, preattentive visual processing but deviations, particularly positive

deviations elicit strong responses and are consequently detected in parallel. This notion successfully predicts the findings from her experiments in that deviations from a default value such as vertical line (default value) such as curved or tilted lines (deviations from the reference 'vertical') are processed preattentively and therefore pop out of a visual display.

Treisman and Patterson (1984) also found evidence for features that emerged from certain combinations of parts of shapes. Features such as closure emerged from a combination of parts into triangles. In this case, when subjects were given a brief display of the component lines of triangles and a feature which specified closure (e.g. a circle), illusory conjunctions between the lines and the closure feature were created which resulted in subjects reporting illusory triangles. Alternatively, subjects reported seeing illusory $ signs in a display of triangles and S shapes. They claim that this is evidence that triangles are not perceived holistically but are defined by a number of different features including lines, angles and closure which are all independent or free floating at the preattentive stage of processing and can therefore form illusory conjunctions with other features in a display. From her studies on illusory conjunctions Treisman added that not only do elementary features mediate parallel search, form easy texture segregation but they can also wrongly combine to form conjunctions with other features in a display and that these three characteristics of simple features correlate highly with each other. Simple features can emerge from a combination of a set of other simple features. For example, closure emerges when a set of lines and angles combine to form a triangle.

Treisman has amalgamated evidence from a number of different studies and formulated what she termed as her feature integration model (Treisman, 1986; 1988). In this model, information from early visual processing consists of a set of feature maps which store information about different features in a modular fashion. The information from these feature maps is integrated by the later visual processes into a temporary object file. Focused attention on the particular location of an object is required in order to integrate the relevant features that apply to a particular object therefore all of the features that are currently present in the selected location are linked. This temporary object representation is constantly updated according to the changes that occur to the object in the real world. Finally recognition occurs by mapping the integrated information found in the temporary object file onto a stored description of the object in memory allowing access the name and subsequent appropriate behaviour. Figure 47 below illustrates the model proposed by Treisman.

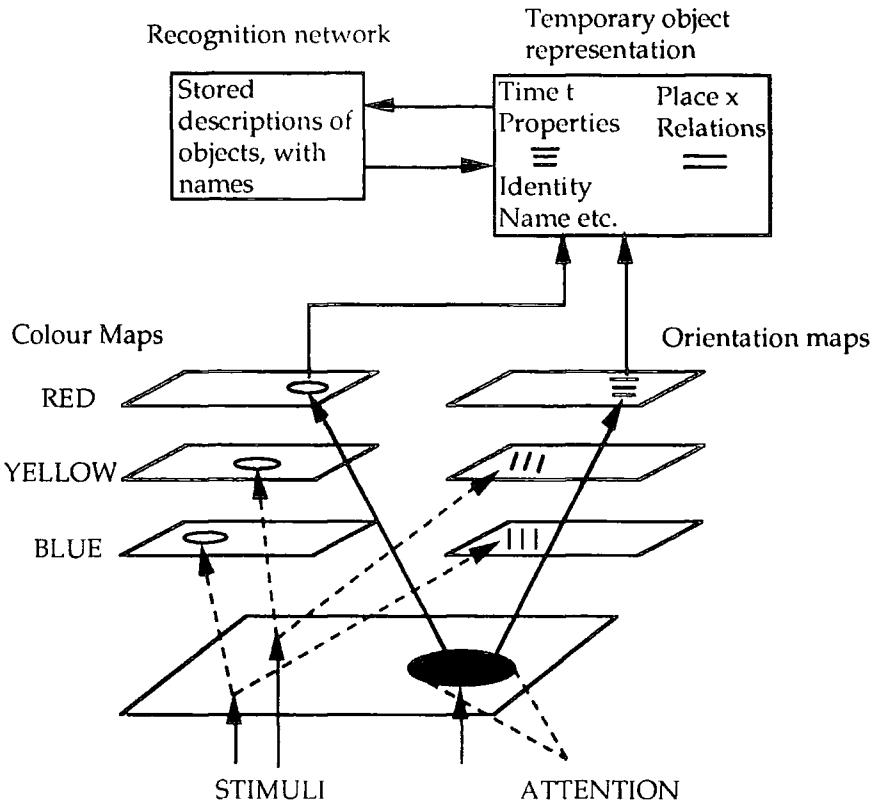

Figure 47: Treisman's model of the Feature Integration Theory (after Treisman et al., 1990).

Treisman's feature integration model, one of the most influential models of visual search, successfully accounts for the many findings from studies on visual search and has subsequently been adopted or amended according to new findings in the area. Her model also receives support from physiological findings. For example, the properties of V1 may subscribe to the properties of the early feature map in Treisman's model. Cells in V1 were found to be selectively tuned to orientation, size, colour, contrast etc. which are features which Treisman and her coworkers found to pop out of a visual scene when they differed from all other items in a display by a that single feature (Hubel and Wiesel, 1977). Marr (1982) argued that information is processed in a modular fashion by early visual processes and it is only in later vision that the outputs from the modules become integrated. The areas beyond V1 specialise in abstracting particular properties from V1 until an integrated representation of an object is created.

The important point about Treisman's feature integration theory is that features are processed in parallel and if a target differs from the distractors in a visual display along a single feature dimension then this target will 'pop out' of the display. Otherwise, targets that are composed of a particular conjunction of features then the target will be found from among a set of distractors only after a serial, self terminating search through each of the items in the display. Search for a target is either extremely efficient reflected by parallel

search processing, or inefficient, reflected by serial processing according to how the target differs from the background distractors. Treisman's model therefore suggests that parallel and serial processes are independent. Although Treisman (1982) did demonstrate that serial search need not be constrained to an item by item analysis but could be conducted across homogeneous subgroups of items nevertheless the independence of the parallel and serial processes remained a feature of her model. However, Treisman (1988) argued that preattentive and focused attentive processing varies along a continuum from broad attention across the entire display to focused attention on one particular item, but there was no specific link between the outputs of the preattentive level and the later attentive level. The following section includes recent amendments to Treisman's feature integration model.

6.1.2 The Guidance Theory

From their studies, Wolfe et al. (1989 and 1990) argued that the output from the parallel feature map can be used to guide a serial search scan of the visual display to find the target. Wolfe, Cave and Franzel (1989) found that subjects were a lot faster at finding targets that were defined by conjunctions of features than was predicted by the Treisman model. They found that the slopes of the search times to the display size were shallow for targets that were defined by conjunctions of colour and form, colour and orientation and colour and size. They also found that searching for triple conjunctions of features was a lot easier than searching for double conjunctions. These findings could not be accounted for by Treisman's proposal that conjunctions of features are searched for in a serial, self terminating fashion. Wolfe et al. therefore suggested a modification of this model. They proposed that parallel processes can guide attention towards likely targets whilst ignoring unlikely targets. This could proceed by either inhibiting the nontarget locations or by increasing the saliency of the candidate targets. They referred to their model as a guided search model. To illustrate their hypothesis, they gave the example that if the target is say a red 'X' from among a set of green 'X's and red 'O's, then a parallel colour map can divide the display into red and green items such that search for the target can be conducted from among the red items without searching through the green items which are unlikely target contenders. Their ideas are somewhat analogous to Watt's model (1988). Watt proposed that a hierarchy of filters work in parallel over the output of the last filter until a representation of the visual field is created. These processes work in parallel across each spatial scale until the target is located at the finest spatial scale and identified by matching it to a representation in memory. In other words, different stages in early visual processing filter out the candidates for the target.

In a subsequent study Wolfe et al. (1992) found that there are limitations on the parallel guidance of search. They found that searching for a conjunction of features that were from the same feature dimension was significantly less efficient than searching for a conjunction of features across different feature dimensions. These results suggested a constraint on the structure of the parallel stage of processing. The distinction between searches within

feature dimensions and between feature dimensions led Wolfe et al. to postulate that single instances of a feature are not processed independently in a modular fashion but are instead processed together which renders search more inefficient than if instances of two different features were processed independently. Wolfe et al. therefore argued that search can be guided by a single feature type and not by many instances of the same feature type. Similarly, Watt (1988) argued that parallel processes operate on coarse grain information and that finer grain information is only specified by attending to the location of the item.

This amendment to Treisman's original model would seem appropriate in the light of recent findings from studies on visual search. It also makes intuitive sense. The parallel processing of shape information with other information such as colour would seem to be necessary for the integration of surface details with object shape.

6.1.3 The Similarity Theory

Duncan and Humphreys (1989) found that search efficiency was determined by the degree of similarity between the target and distractor items in a display and also the similarity within the distractors themselves. They claim that search efficiency decreases continuously with an increase in the similarity between the target and nontarget and with decreasing similarity between the non-targets. Treisman's model does not account for this finding. Instead of agreeing with Treisman that there are two distinct selection processes, Duncan and Humphreys propose that a continuum of search efficiency results from the degree of similarity across target and distractors and within the distractors themselves.

In the Duncan and Humphreys (1989) study, a number of experiments were run in order to test the effect of similarity across the target and distractors. All of the experiments measured the reaction times to locate the target in varying sizes of visual display. In the first two experiments subjects had to locate either an 'L' or a tilted 'T' from among a set of either upright or 90° rotated 'T's. These distractors were either homogeneous (all of the same orientation) or heterogeneous (of mixed orientations). The letter size of the targets and distractors was varied as a further factor. They found that the different combinations of the experimental factors all produced flat search slopes suggestive of parallel search. When smaller letters were used, the search slopes became slightly steeper. These results were independent of the exposure times of the search display with both long and short exposure times producing the same effects. Duncan and Humphreys noted that although the targets could differ from the distractors by a feature (e.g. a 'T' among tilted 'T's) or by a conjunction of features (e.g. an 'L' among 'T's) the overall results were the same. There were slight variations in the slope between the feature difference and conjunction search however and Duncan and Humphreys tested the effect of similarity further by asking subjects to locate an 'L' from among a set of 'L's rotated 90° clockwise or anti-clockwise. They found that although search functions for an 'L' amongst homogeneous distractors were almost flat, search functions for an 'L' amongst heterogeneous distractors was very difficult. This variable effect of

similarity on search times across display sizes was investigated further in an experiment where subjects had to locate an 'L' from among a set of 'T's tilted either 0°, 90°, 180° or 270°. These particular distractors were chosen because in a pilot study 'T's tilted by 0° or 90° were faster to identify than 'T's tilted by 180° or 270°. They found that search for an 'L' among 'T's shows an increasing effect of display size even when the distractors were homogeneous. Duncan and Humphreys concluded that the results from their experiments support the notion that the degree of similarity between the target and the distractors and the degree of heterogeneity between the distractors themselves interacts to create a continuum of search efficiency rather than support for a dichotomy between parallel and serial search. When the target is sufficiently different from the distractors, then the similarity between the distractors will have no effect on the search slopes, with search times remaining independent of display size. Search slopes increase as the similarity between the target and the distractors increases and as the homogeneity between the distractors increases. Figure 48 below illustrates this interaction. Duncan and Humphreys argue that Treisman's feature integration model neglects the important variable of the degree of similarity between distractors.

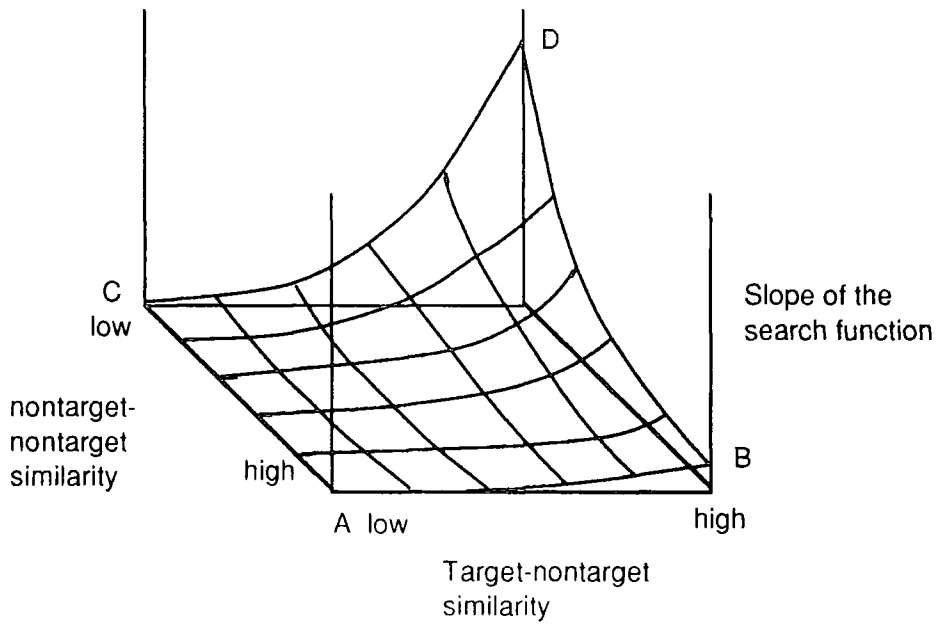

Figure 48: A search time surface illustrating the effects of increased similarity between the target and nontargets and the decreasing similarity between the nontargets on search efficiency (after Duncan and Humphreys, 1989).

Their model makes specific predictions on the nature of the selection processes and Duncan and Humphreys have proposed a theory of why selection processes are sensitive to the similarity of the items in a visual display. A brief outline of their theory will be discussed: In searching for an item in a display, the correct item must be selected for access to

the visual short term memory (VSTM), which in turn allows the selected item to be the focus of current behaviour. However, the surrounding items in the display are also competing for access. The visual system must select the correct item irrespective of all other items in the display. As access to VSTM is limited, its resource can be divided across the items in the display. The more of this resource that is assigned to an item in the visual display the more likely it will be selected and processed further by the visual system. However, although this provides a plausible model of the bottle-neck system to VSTM it is just a description and does not provide an explanation of how VSTM's resources are distributed among the items in a display. Duncan and Humphreys propose that each item is assigned a weight relative to the surrounding items in the display which reflects the strength with which it competes for selection. Weights are assigned according to the degree of similarity across the items but also according to the information which is available about the nature of the target. An item is selected for access to VSTM according to the relative strength of the weights between the items in the display with larger weights being more likely to be selected than smaller weights.

This similarity theory differs from Treisman's feature integration theory for two reasons. First, Duncan and Humphreys propose that parallel and serial search strategies are two extremes of the same continuum of search efficiency. Treisman would argue that these processes are autonomous and peculiar to feature and conjunctive search. Second, Duncan and Humphreys argue that the relations between the distractors can also effect search efficiency, whereas Treisman does not consider the relations among items in a visual display other than the relations between the target and the distractors.

Duncan and Humphreys model of the search processes successfully accounts for a lot of the different results found in the literature especially the instances where neither a strictly parallel or strictly serial search function was found (see Duncan, 1983; Nakayama and Silverman, 1986 and Wolfe et al, 1990, Dehaene, 1989). However, there are some findings in the search literature which the Duncan and Humphreys model does not account for (see Dehaene, 1989). One such finding is the notion of search asymmetries mentioned above. Treisman has shown that when the target and distractor are interchanged in different studies, different search slopes are evident across the different trials (Treisman and Gormican, 1988). According to the similarity theory, the degree of similarity across the target and distractors should remain the same regardless of which item is the distractor or which the target. Treisman accounts for this asymmetry by proposing that the detection of a target that is defined by the absence of a feature relative to the distractors is more difficult because the activity of the target in the feature map is reduced relative to the activity of the distractors therefore an item by item analysis of the display is required to find the target.

Another finding from Treisman's work is that when subjects cannot attend to the location of the item due to the short exposures of the visual display that illusory conjunctions are often reported between the features of the target and the flanking distractors despite the

lack of similarity across the items. If similarity across the items in a visual display is indicative of the efficiency of the search processes, why should features be perceived as being independent when attention is not directed to the location of the target. In other words, the Duncan and Humphreys model does not incorporate the important findings which support the idea that features are processed in parallel in a modular fashion. Instead their model suggests that the parallel stage of processing is a single multidimensional space with search becoming more efficient with greater distance between targets and distractors along this multidimensional space (see Figure 48 above).

6.1.4 The Pattern Recognition Approach

The assumption that detection in visual search tasks is determined by the feature-coding properties of early visual processing has also been challenged recently (Bravo and Nakayama, 1992; He and Nakayama, 1992). Nakayama and his co-workers have found evidence to suggest that visual search has little access to individual feature maps but that a higher level process of surface representation is important for detection. This representation of the visual display as a surface can be processed as a whole when the task for a subject is to detect a previously unknown target and report its distinguishing feature. However, when more information is required of the subject such as to report the shape of the target, then a smaller part of the surface is used and attention must be directed to the appropriate part (Bravo and Nakayama, 1992). He and Nakayama (1992) found that in an experiment where subjects perceived the surfaces of a target and distractors to change in depth (by subtle manipulations of a haploscope) whilst leaving the features of the targets and the distractors intact, it was very difficult to search for the odd target. The target was an either an 'L' shape or a mirror-reversed 'L' shape with several distractors ('L' or reversed 'L') all accompanied by a square. The target and the distractor 'L' were always presented in the same depth plane but were perceived to be in different depth planes relative to the squares. The task was to search for an odd target. The authors found that when the 'L' shape was perceived to be behind the square, this disrupted search efficiency and therefore increased search times. In this case the 'L' shape is perceived as part of a larger surface occluded by the square in front and not as an 'L' shape feature. When the 'L' shape is perceived as being in front of the square, then it is more readily detectable with flat search slopes to the number of items in the display. The authors argue that when the subject no longer perceives an 'L' feature but a square surface that is occluded by another surface in the foreground then search becomes less efficient because the target and the distractor become perceptually more alike and therefore less distinguishable. These results suggested to the authors that visual search is applied at a higher level of visual representation than the early feature detection because subtle changes in the perception of surfaces without altering the features themselves disrupt search efficiency.

6.2 Eye Movement Studies

There has long been the distinction between foveal and peripheral processing. However, it is not yet clear how much information or what sort of information can be used from peripheral vision. Studies which have looked at the pattern of eye movements in a search task can highlight the sort of information that can be picked up and utilised from peripheral vision. Although fixating an object can increase the amount of information available because visual acuity is optimal at the fovea, the evidence is equivocal as to how eye movements and fixations are related to shifts of attention across the visual display. The following section reviews a number of studies which have investigated the nature of the information in peripheral vision which can be used and also the evidence for the relationship between attention and eye movements.

Many studies have monitored the pattern of eye movements to a given task in order to investigate the nature of how they relate to the underlying cognitive processes. Eye movement studies have given insight into the cognitive processes involved in reading (Bouma, 1978; Just and Carpenter, 1976; Rayner, 1983), mental rotation (Just and Carpenter, 1976) visual search tasks (Gould and Dill, 1969; Bouma, 1978; Loftus, 1983) and the integration of information about shapes (Pollatsek et al, 1984; Henderson et al, 1989; Hayhoe et al, 1991; Henderson, 1992). For example, Just and Carpenter (1976) found that the pattern of a subjects eye fixations when conducting the Shepard and Metzler (1971) mental rotation task substantiated Shepard and Metzlers' conclusion that mental rotation is an analogue process because switches in the fixations during the transformation stage indicated that the rotation was monitored in steps of about 50°. An important conclusion from the Just and Carpenter study was that the locus of the eye fixation reflected what was being internally processed.

Fixating on an item in a visual field increases visual acuity and consequently more high resolution information (or fine details) may be processed. This area of the visual field has been referred to as the 'useful field of view' (Loftus, 1983) or the functional visual field (Bouma, 1978) within which an object can be viewed in detail. The structure of the retina is such that visual acuity is best in the fovea which subtends about 2° of visual angle and visibility usually decreases with eccentricity. The properties of the items in the periphery determine the size of this functional or useful visual field. For example, it seems to be smaller for tasks which demand processing of text (because substantial acuity is needed to distinguish one letter from the other) than for processing information about scenes (because features in a scene can often be large enough to be determined in peripheral vision). There may also be an alternative reason why letters and items in a scene have a differential effect on the functional visual field. Because of the relative similarity across letters due to the shared features this may increase lateral interference and fixation would be required to locate the letter in a visual search task. Scenes do not often include items that are similar, therefore an increase in lateral interference would not occur and the target may be detected in peripheral vision. However, information in the periphery may be used to guide eye

movements to fixate on a likely target in a visual display (Gould, 1969; Gould and Dill, 1969; Shepherd, Findlay and Hockey, 1986). Gould (1969) argued that the observer uses three sources of information in order to fixate a target in a visual array; the global preattentive processes, the observer's prior knowledge of the target and the observer's intention or purpose. This information could be further classified into bottom up and top down influences. Treisman argued that eye movements were not necessary in a search for a simple feature which could be found by preattentive processes but that it may be necessary to foveate each item in a conjunction search in order to facilitate the discrimination between the target and the distractors (Treisman and Gormican, 1988). Treisman however made no attempt to monitor eye movements in any of her studies on visual search. It may be therefore that the more complex a visual scene, the more likely the subject is to make eye movements to each item before locating the target. This may be particularly true for high level discriminations such as searching for a shape from among a set of other shapes (Bravo and Nakayama, 1992) or for tasks that demand close scrutiny of the targets (Findlay and Kapoula, 1992).

Gould and Dill (1969) investigated the role of eye movements in a pattern discrimination task. The potential effects of top down processing were removed because novel stimuli were used in the experiments. They found that when subjects were asked to find matches to a standard, unfamiliar pattern from among a set of patterns surrounding it, the more similar the standard pattern was to a surrounding pattern, the more likely the subject was to fixate on it. The patterns consisted of nine asterisks arranged in either a 2×2, 4×4 or 8×8 matrix. The patterns were arranged around the centred, standard pattern in a square array. The visual angle remained constant across trials and it subtended 7° horizontally and 8° vertically. A match to the standard pattern was nearly always fixated foveally and the probability of fixating a distractor depended on its similarity to the standard pattern. They found that the matrix size of the pattern did not affect the results but that the number of elements in a display did cause an increase in fixation times. The authors concluded that fixation times were affected by the relative characteristics of a pattern to the standard pattern rather than the absolute characteristics. Another interesting finding was that subjects tended to skip over distractors that were not very similar to the standard pattern. This suggested that information from peripheral vision was sufficient to detect a distractor but that foveal fixation is usually required to determine whether a pattern is a target.

In a further investigation of what information can be processed from peripheral vision, Pollatsek, Rayner and Collins (1984) found that more higher level information such as the name of an object can also be processed along with visual information. In their experiments subjects were required to fixate on a line drawing of an object that was presented in peripheral vision and then to name the object that was fixated. During the eye movement, the first stimulus could be replaced by another which differed from the initial stimulus according to the experimental conditions. They found a facilitation effect on naming times when the stimuli were identical compared to a control condition where just the location was specified initially. When the original stimulus was replaced by another with the same name

(e.g. two different cows), a facilitation was also observed although not the same as that for the identical picture condition. When the stimuli had different names, then only the visual information across the stimuli produced a facilitation (e.g. a ball replaced by a tomato would cause a facilitation). The authors conclude that both the visual features of the objects and more semantic information such as the name of the object can be picked up from peripheral vision and integrated across saccades.

Similarly, Hayhoe et al. (1991) concluded that information is integrated at a higher, post-categorical level in the visual system and more abstract information can be integrated such as the visual features of an object and its name. In their investigations subjects had to report whether the top angle of three angles shown in a display was acute or obtuse. The three angles defined the angles of a triangle and were either presented in sequence or simultaneously. When presented in sequence, subjects moved their eyes from one angle to the other. However, Hayhoe et al. found no difference in performance between the simultaneous and sequential tasks suggesting that spatial information can be preserved across successive eye movements.

The studies reported above provide evidence that information other than visual information can be picked up from peripheral vision. However, what needs to be determined is how attention relates to the subsequent eye movement to an item in a visual scene. There is a lot of evidence to show that attention can be independent of eye movements (Eriksen and Yeh, 1985; Posner, 1980). Posner (1980) found that subjects could shift attention independently of the direction of their eye movements. Nevertheless, it has been argued that selective attention is a necessary component to the subsequent saccadic eye movement to a target, postulating a direct link between attention and eye movements (Shepherd, Findlay and Hockey, 1986; Findlay and Kapoula, 1992). Posner (1980) proposed a model of spatial attention which was built around the notion of a spotlight which would select the appropriate area of the visual field for processing. Henderson, Pollatsek and Rayner (1989) tested the effects of sequential presentation of four objects arranged in a square array with simultaneous presentation of the four objects on the fixation times of each object. They found that fixation times were significantly less when all objects remained on the display than when only one object at a time was shown. This result substantiated the preview benefit found in the Pollatsek et al. (1984) study. Henderson et al. found that this preview benefit occurred independently of the position of the objects in the display and also the size of the display. The preview benefit remained unchanged as the number of objects viewed from the first fixation increased suggesting that multiple previewing does not have an additive effect on the benefit but has a limited effect. These results suggest that extrafoveal information is picked up from the location which is to be fixated next, to the exclusion of the other locations. They concluded that their results provides evidence that items in the visual scene are attended to in sequence, as in Posner's (1980) searchlight metaphor of attention, and that the oculomotor system is functionally related to visual attention. They argued that under normal search tasks in the natural environment, shifts of attention precede a saccadic eye movement

in order that a potentially interesting item in the scene can be observed.

The studies outlined above give evidence that both visual and semantic information can be detected in peripheral vision. From the investigations into the variables that affect search efficiency, it can be concluded that peripheral information influences search efficiency in a number of ways. For example, if the target is similar to the non-targets then search times are found to increase with display size (Duncan and Humphreys, 1989). Also, if the target is sufficiently dissimilar from the background items, then search times are found to be independent of display size (Treisman, 1986). These results can apply to the real world. For example, searching for a book on a book shelf is probably more difficult than searching for a television in a sitting room. Although different variables within a scene or a display can affect search efficiency, it is interesting to question whether different manipulations of a target would affect search times within a fixed display size. Gould and Dill (1969) found that a pattern that resembled the central pattern was more likely to be fixated suggesting that mapping across patterns is involved in detecting a match. Similarly, in the real world an object that is searched for may be found in a number of different states and an object may be detected on the basis of it being the most likely match to a stored representation in memory. For example, different lighting conditions can change the shading patterns on the surface of the object and the object could also be found in different orientations. The visual system needs to be prepared for such changes and the ability to detect the object should be independent of these changes. However, it has already been established that the recognition of objects is not independent of changes in orientations. The ability to recognise an object and detect an object in a scene may reflect the same visual processes. An investigation into the ability to search for objects in different orientations is described in the following chapter. This investigation was conducted in order to discover whether visual search is not only affected by the relationship between the target and the nontargets but also by changes in the target object.

6.3 Conclusions

From attentional and eye movement studies it seems that the more complex the scene the more that attention and eye movements proceed in a serial manner from item to item. But this is not a random process. Attention and eye movements are guided by either the outputs of the preattentive processes, top-down processes and the information available in peripheral vision. The efficiency with which a target is searched in a visual scene can depend on a variety of factors from the level of similarity between the elements in the scene (Duncan and Humphreys, 1989) to the level of visual information required, for example, reporting the shape of the target object (Bravo and Nakayama, 1992). It seems likely however, that searching in a natural scene it is unlikely that a target would systematically differ from all other items present in the scene such that it could be detected preattentively. Nevertheless, search can be guided to locations where the presence of the target would be highly probable. For example, when searching for a chair, the search could be restricted to larger items in a

room (see Dehaene, 1989). Once a chair has been located, then it can be matched to a representation in memory and identified as such. Search efficiency may also depend on the nature of this representation in memory. For example, if an object is shown in a different orientation in the scene other than a view represented in memory then identification time increases. However, it may also take longer to locate the object if shown in an unusual orientation. This issue is addressed in the following chapter.

Chapter Seven

Visual Search and Object Recognition

7.1 General Introduction

This chapter is concerned with the way in which the visual system detects and identifies familiar objects in a visual scene. A typical visual scene can be very complex but locating and identifying an object seems to occur relatively easy for most people even though the object can vary from one scene to the next in colour, illumination or orientation. For example, a wine bottle may be of green or brown glass, it may be found in a dark cellar or in a dining room and it may be found upright on a table or on its side on a wine rack. Yet when shown a photograph of a scene most people have no difficulty in readily locating and identifying the different objects that make up that scene.

As was argued in the previous chapter, higher order or top-down processing is involved in the detection of an object in a scene (Bravo and Nakayama, 1992). From other studies, it was concluded that an observer uses the overall information from a visual scene to determine where a target is likely to be found (Biederman et al., 1974, 1982). Biederman et al. (1974) found that subjects were less accurate in identifying a target in a cued position in a scene that was jumbled than in a scene that was not jumbled. They concluded that subjects could not only extract information about an individual object from a single glance at a scene but that they could also extract the overall characterisation of the scene. Biederman et al. (1982) later argued that it was not the spatial relations between the objects in a scene that resulted in an overall characterisation but an 'inventory listing' of the objects in a scene. Detection of an object such as a fire hydrant in a kitchen scene is slower and less accurate than if shown in a street scene. However, the relative positions of the objects in a scene is also important. For example, placing the fire hydrant on top of a post box increases search speed as much as including it in a kitchen scene. Biederman's studies show that more top down information such as scene structure can be detected from a single glance at a scene.

Other studies, that have been discussed in the previous chapter, suggested that semantic information can be extracted from peripheral vision (Pollatsek et al, 1984; Henderson et al, 1992). Pollatsek et al. found that both visual and semantic information such as a name of an object presented in peripheral vision can facilitate naming times of a fixated object. Subjects fixated on a central cross when the position of the target was precued. During the saccadic eye movement to the target, the target was replaced by either the same picture, an object with the same name or a visually similar object. They found that both the same picture and the same named object caused a facilitation effect. They concluded that more

higher–order information can be picked up from peripheral vision than just purely visual information. These results may suggest that information from objects that are flanking a target object in a visual scene may influence search times.

Henderson et al. (1989) found that information derived from peripheral vision for recognising an object in a scene is influenced solely by the object in the location to which the eye will move next and not from all of the objects present. They concluded that the identification of an object in a scene is not influenced by all of the objects present in peripheral vision but information is only acquired from the object that the eyes saccade to next. In other words, serial attentive processes work on the next object location and not all locations.

The notion of that selection processes can be guided by both top down and bottom up processing has recently taken significance in visual search (Wolfe et al, 1989; Duncan and Humphreys, 1989 and Bravo and Nakayama, 1992). Treisman's model of attentive processing, on the other hand, is committed to bottom up processing, although she does accept that features can be learned or in other words that familiarity would establish a new perceptual unit (Treisman and Paterson, 1984). If this is the case, then according to the feature integration theory it could be assumed that familiar objects would be coded as discrete perceptual units and that they would be detected in parallel in a visual scene from among a set of homogeneous distractors. However, any additional feature added to a simple feature requires focused attention in order to conjoin the features (Treisman, 1986). For example, if the objects were shown in different orientations in a visual display, then a serial search strategy would be employed in order to find the target. It seems unlikely, however, that observers do an item by item analysis of a natural scene when looking for an object. The role of top–down processing would seem to be a lot stronger than Treisman's model has accounted for.

Nevertheless, some aspects of Treisman's model of search processing would seem to hold. For example, when searching for an object in a 'nonscene', clock-face type display, the number of items in the display affects the search times such that search is less efficient as the number of distractors increase (Biederman et al., 1988). There was no difference found between search times to a target surrounded by distractors which are likely to co-occur in a natural scene or distractors that are highly unlikely to co-occur, suggesting that top down information is only used when searching for an object in a natural scene rather than in a 'nonscene'.

The studies discussed above may suggest that the role of attention can change according to the demands of the search task. It may be that attention can be influenced by top–down processing only in natural scenes. However, some variables may affect search efficiency even in an natural scene. For example, an object may be more difficult to detect in a scene if it is orientated in depth away from the upright. Similarly, immediate detection may be rendered impossible if the target is surrounded by similar distractors such as searching for the proverbial needle in the haystack. Given that objects are unlikely to be found under the same conditions from one scene to the next it was decided to look at the effects that such

changes may have on the object's detection in a visual scene. Objects are often found in different orientations in the environment, therefore, it is interesting to question what sort of information is used in order that an object can be detected irrespective of its orientation.

A task that involves searching for an object that may be disoriented in a scene makes a specific prediction according to Treisman's model (see Figure 47 in the previous chapter). Her model would predict that in a complex array of objects shown in different orientations, the target would not be perceived preattentively. This would be because the similarity among the elements is too great, or rather that the difference between the pooled background activity between trials that contain the target and trials where the target is absent is small, therefore the target will require focused attention in order to determine its identity (see Treisman and Gormican, 1988). The target would be found through serial search processing of the elements in the scene. Treisman would argue that the search for an oriented object requires focal attention to conjoin the features of form and orientation.

According to the Guidance search model, unless the target object was in a different orientation to surrounding objects in a scene, then search could not be guided by any of the information from the other objects (Wolfe et al. 1989; Wolfe et al. 1992). Similarly, if the objects surrounding the target were visually similar and oriented randomly, then the target would be sufficiently similar to the distractors and the distractors sufficiently similar to each other to affect a serial search strategy until the target is found (Duncan and Humphreys, 1989). Thus, according to the major search theories a serial search strategy would be adopted when searching for an oriented object shown among randomly oriented distractor objects.

The aim of the following experiments was to assess the nature of the information that is used to detect an object from among a set of other objects in both homogeneous and heterogeneous displays. Four controlled experiments were run in which subjects had to locate a match to either an object or an object's name from among a set of different objects arranged in a circular array. The following experiments were run in order to determine search efficiency given that objects are not always found under the same conditions in the natural environment. The experiments were particularly interested in the effect of orienting the object on search efficiency and whether orienting the object in depth would be more detrimental to search efficiency than orientations in the picture plane. The recognition time experiments reported in Chapter 3 found that some orientations in depth make the object more difficult to recognise but that objects were equally recognisable across orientations in the picture plane. The ability to recognise an object in different orientations may affect the ability to detect oriented objects.

Orientation in the picture plane and orientation in depth have been established as individual simple features which are detected preattentively because each cause a pop-out effect in search tasks (Treisman and Gelade, 1980; Epstein and Babler, 1989, 1990). However, Wolfe et al. (1992) found that the detection of a target in the presence of randomly oriented distractors, reduces search speed and the target is found only after an item by item search. It

seems therefore that randomly oriented items in a display renders search less efficient. However, some views of objects are difficult to recognise which may effect search efficiency to the extent that each object may require focal attention before the object is recognised (Just and Carpenter, 1976). This result would be expected for orientations that are not represented in visual memory because top–down processing could have no effect on patterns that were not represented in memory. In such a case other bottom–up processing would have to have effect in order to transform the object to match to a memory representation for identification purposes (Bravo and Nakayama, 1992).

The experiments reported in this chapter were based on the experimental paradigm used by both Gould and Dill (1969) and Biederman et al. (1988). The Gould and Dill study is outlined in the previous chapter. They measured search efficiency in terms of eye movements to the correct target in a fixed display size of different targets and distractors which were arranged in a square array around a central shape. Subjects were required to locate the correct match to the central shape. They found that the degree of similarity between the central shape and the items surrounding it affected the item that the eyes initially saccaded to. Biederman et al. (1988) found that the time to locate a familiar object from among a set of other objects arranged in a circular array is not affected by the distractor objects.

The following studies set out to examine the effects of searching for an object when shown in different orientations. Both search times and eye movements were used as a measure of search efficiency.

7.2 Experiment 8

This experiment was designed to test search efficiency of familiar objects in a 'nonscene' display (Biederman, 1988) when the object could be found in one of five different orientations and surrounded by randomly oriented distractors. Two groups of subjects were run on the same experiment. Search times were measured for both groups of subjects and the eye movements of one of the groups were recorded. Henceforth the subject groups will be referred to as the 'search time' group and the 'eye movements' group.

7.2.1 METHOD

Subjects

Search time group

Ten undergraduates from the Department of Psychology, University of Durham participated in this experiment. All subjects had normal or corrected to normal vision. Six of the subjects were male. The ages of the subjects ranged from 19 to 22 years. These subjects had not participated in any of the previous experiments reported in this thesis.

Eye movements group

Six members of the Department of Psychology, University of Durham including the experimenter participated in this experiment. Four of the subjects were male. All subjects had normal or corrected to normal vision. Their ages ranged from 24 to 48 years.

Stimuli

Eight common objects were drawn on a Macintosh IIx computer using the drawing package, Swivel 3D. These objects were a screw, a bottle, a lamp, a glass, a cricket bat, a rolling pin, a light bulb and a frying pan. A clothes peg was included as an extra distractor object and its presence was counter-balanced across all trials. The objects were arranged on a screen in a clock face display i.e. the target object was positioned in the centre of a circle and eight objects were arranged at equidistant points along the circumference, i.e. at the principal axes (North, South, East and West) and at the principal oblique axes (North–East, South–East, North–West and South–West). The display either contained a matching object to the object in the centre or they did not. The proportion of the match–object absent to match–object present trials was 1:2. In the match–object present trials the position of the match–object in a display was counter-balanced across all trials so that the match was equally likely to be found in any of the eight positions.

All objects were drawn so that their elongated axes were the same length with a maximum length of 2 centimetres when fully exposed. The objects were shaded in different tones of grey. The objects were chosen because of their similar structure i.e. all of the objects had well defined elongated axes and minimum number of surface features. Figure 49 below shows an example stimulus from the experiment.

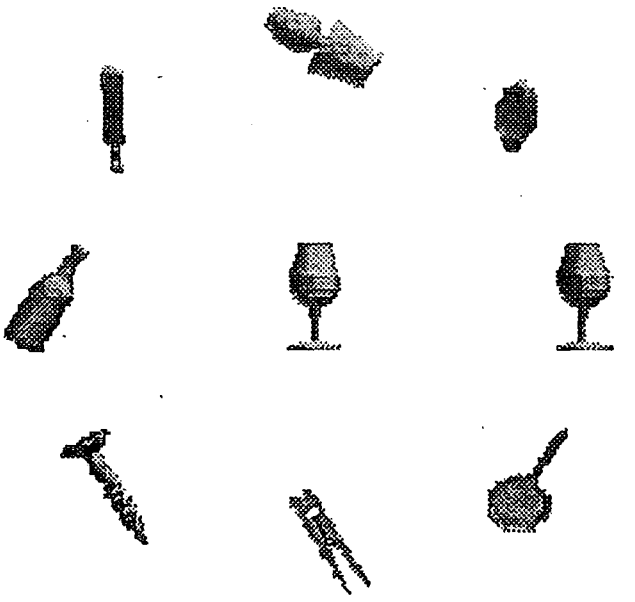

Figure 49: An example of a stimulus used in Experiment 8.

The central, target object was always presented in an upright position (see Figure 49 above). The match object was presented in one of five different orientations (see design section). The distractors were shown in random orientations in each of the displays. Each distractor object was shown either in an upright position or rotated in one of eight ways; 2-D 30°, 2-D 60°, 2-D 120°, 2-D 150°, 3-D 30°, 3-D 60°, 3-D 120°, or 3-D 150°.

The same stimuli were for both the search time subject group and the eye movement subject group. However copies were made of each stimulus and mounted on the slides in the reverse to the original slides for the search time group. It was thought necessary to double the number of slides for this group as 60 slides seemed too few to present in a reaction time experiment. These extra slides were not subjected to later analysis and for the purpose of this study they were subsequently ignored. The proportion of match–present: match–absent trials remained the same as in the eye movement study i.e. 2:1.

Materials and Apparatus

Each trial was photographed from the Macintosh IIx screen and presented in slide form. There were 60 experimental trials in all; 40 match–present trials and 20 match–absent trials.

Each slide was projected onto a 2 way screen which was positioned away from the projector in order that the radius of the search display was 10 cm. The subject sat approximately 114 cm. from the screen. Thus the visual angle from the target object to any of the objects in the array subtended 5°.

A practice block of 12 trials preceded the experimental trials. The practice trial did not include any of the displays of the slides from the experimental block.

Search time group apparatus

Subjects responded to each slide using a labelled two-key response box with a 'MA' (match absent) response key and a 'MP' (match present) response key. Subjects could respond to a match–present trial using their dominant hand. A key press stopped a timer in a BBC micro computer which also registered the reaction times to each trial. A response also triggered the onset of the next slide after a delay of 2 seconds.

Eye movement group apparatus

A scleral eye coil was placed in the right eye of each subject in order to track the position of the eye during the search task (see Collewijn et al., 1975). A description of the eye coil is as follows; A suction ring search coil is placed in the eye of the subject who sits inside two large coils which creates an electromagnetic field that is uniform in the eye region. The eye coil is attached to the larger coils by a small copper wire. A voltage is induced with each movement of the eye but not with lateral head movements. Voltages are measured in both vertical and horizontal channels.

A photocell was placed on the projector screen to trigger a timer on an Alpha computer to record the reaction times. It was placed in such a position on the screen that the light from each slide would fall onto the photocell synchronising the onset of the timer with each slide presentation.

Eye movements were recorded onto the Alpha data recorder. A key press on the response box signalled the end of the eye movement recording for each slide. The same response box was used for this group of subjects as for the previous group. Fixation times and latencies were recorded for each saccade made during the search task.

Design

The experiment was based on a 2 way design with orientation conditions and objects as factors. The position of the match object was a nested factor. The orientations factor included five orientation conditions which were as follows:

1) SAME condition: the matching object was depicted in the same orientation as the target object,

2) 2D30: the matching object was rotated 30° from the original orientation in the Z axis (i.e. the picture plane),

3) 2D60: the matching object was rotated 60° from the original orientation in the Z axis,

4) 3D30: the matching object was rotated 30° from the original orientation in the X axis (i.e. rotated in depth),

5) 3D60: the matching object was rotated 60° from the original orientation in the X axis. See Figure 50 for an illustration of these conditions.

There were eight match objects used in the experiment. There were also eight different positions in a display. The positions of the match objects were counter-balanced across the experiment.

For the purposes of this experiment, each match object was set in a heterogeneous display. In other words the distractor objects were randomly oriented in each display.

The order of presentation of the slides was randomised across all subjects. For the search time group, the reversed displays were always shown after their corresponding experimental slides in order to avoid any practice effects which may transfer across the trials.

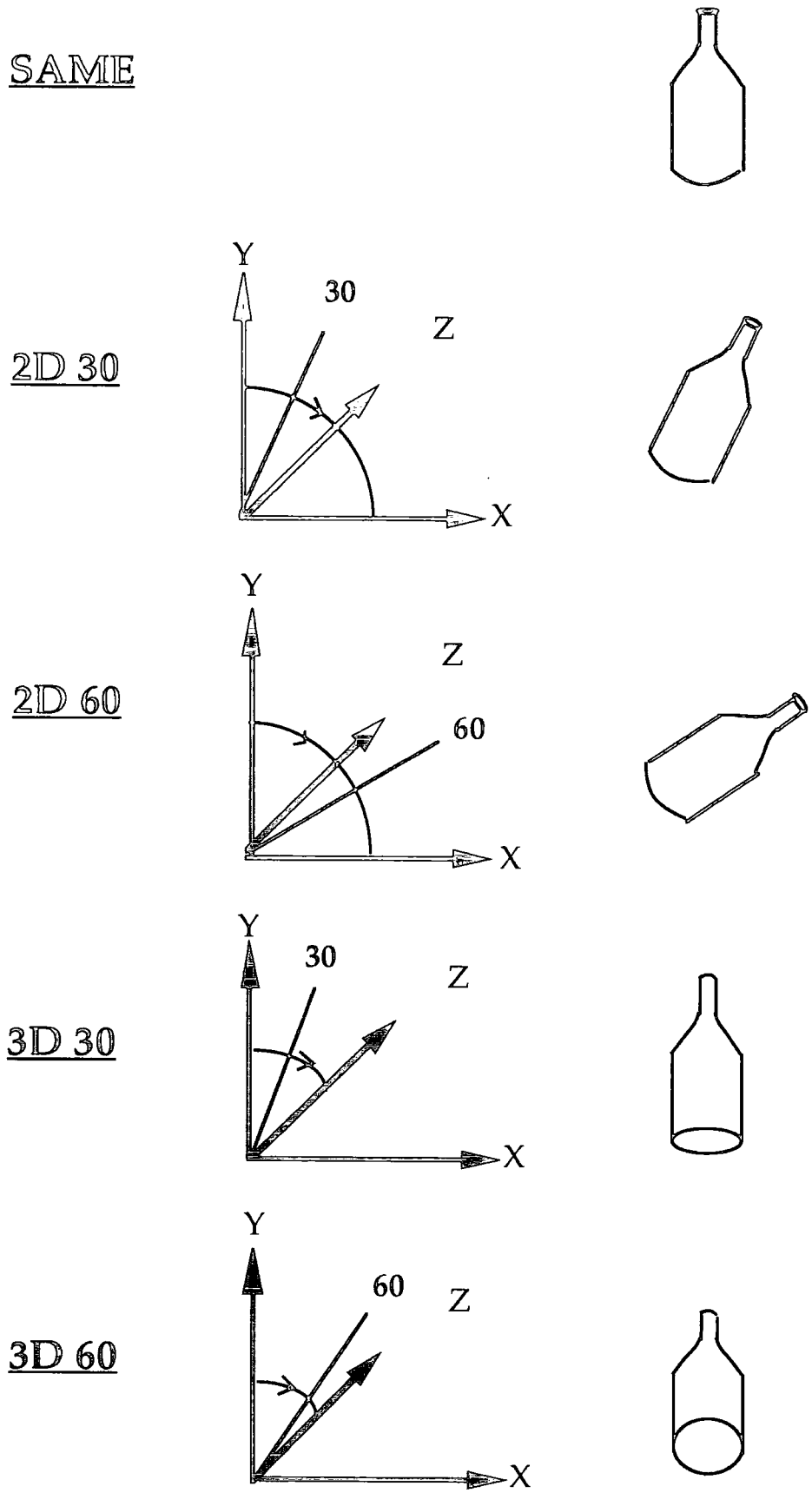

Figure 50: An illustration of the orientations of the match objects tested.

Procedure

Each subject was requested to attend to the screen for the onset of each slide. They were instructed to locate the matching object to the object displayed in the centre of the screen from among the objects surrounding it as fast as possible and to press the appropriate response key, MA (match absent) or MP (match present), as soon as they had made a decision. Both groups of subjects were allowed to move their eyes freely over each display. They were clearly instructed to minimise the number of errors.

A practice block of 12 slides was initially presented to each subject. The experimental block immediately followed the practice trials and there were no breaks taken during the experiment, which lasted approximately 20 minutes for the search time subject group and 10 minutes for the eye movement group.

A response triggered the onset of the next slide after an ISI of 2 seconds. The display remained on the screen until the subject had responded.

For the eye movement group a calibration slide of a 9 point grid preceded the experimental trials. If the eye movements displayed salient step-like patterns showing a linear signal on the oscilloscope then the experiment could proceed. The subjects were asked to fixate on the centre of the screen before the onset of each slide. They were requested to locate the matching object to the target object in the centre of each display as fast as possible and to press the 'match present' key or 'match absent' key as soon as the decision was made. The subjects wore the eye coil for the duration of the experiment.

7.2.2 RESULTS

Only the match–present slides were analysed for both sets of subjects. The analysis for each group is presented below.

Search time subject group.

Errors were less than 5% and there was no evidence of a speed/accuracy trade-off.

Figure 51 depicts the mean search times of all subjects in each condition. An analysis of variance of the search times across all subjects proved significant for condition, $F(4,36)=35.580$, $p=0.0001$ and object $F(7,63)=8.329$, $p=0.0001$. A significant interaction between condition and object was also found, $F(28,252)=2.261$, $p=0.0005$.

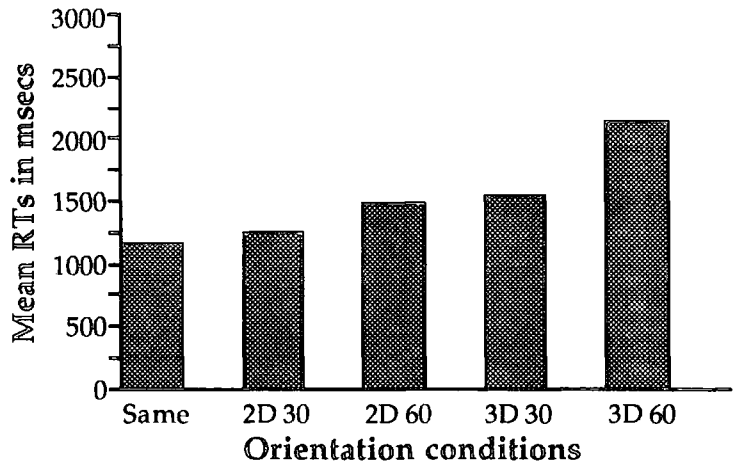

Figure 51; Graph showing mean search times taken to locate object matches in different conditions of orientation by 'search task' subjects.

A Newman-Keuls analysis on each of the main effects highlighted the simple effects at $p<0.01$ level of significance: For the condition effect it was found that 2D60 was significantly different from the SAME condition, the 3D30 was significantly different from both the 2D30 and the SAME condition an finally, that the 3D60 condition was significantly different from all other conditions. The 2D30 condition was also found to be significantly different from the 2D60 condition at $p< 0.05$ level of significance.

The object effect was then subject to a Newman Keuls analysis. A significant difference was found between light bulb and frying pan, screw, glass, lamp and rolling pin at $p<0.01$ level of significance. The light bulb was also significantly different to the bottle at $p<0.05$ level of significance. The cricket bat was found to be significantly different to the frying pan, screw, glass, lamp and rolling pin at $p<0.01$ level of significance and different to the bottle at $p<0.05$ level of significance. Figure 52 below shows the mean search time taken to find each of the match objects.

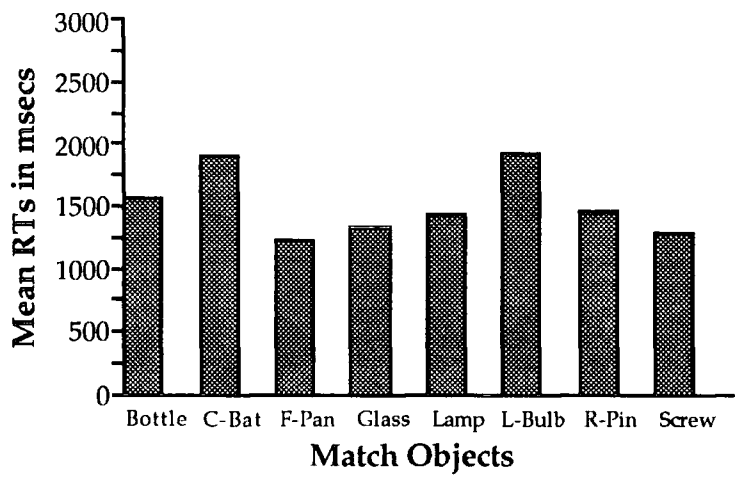

Figure 52: The mean search time taken to locate each of the match objects in the 'search time' experiment.

The time taken to search for a match that was located in a particular position was also subjected to an analysis of variance. This proved highly significant, $F(7,63)= 5.443$, $p=0.0001$. Figure 53 shows the mean search times taken to locate the match in each position.

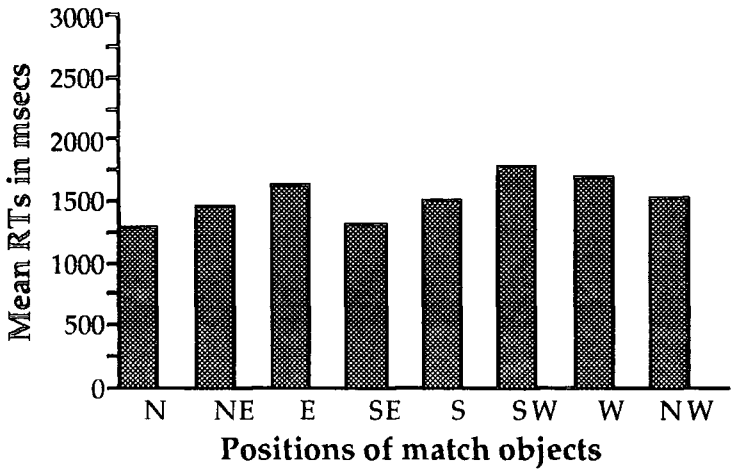

Figure 53; Graph showing mean search times to locate the matching object when shown in different positions in the visual display.

A post hoc Newman-Keuls at $p<0.01$ level of significance revealed that the North position was significantly different from both West and Southwest and that South East was also significantly different from both West and Southwest. At $p<0.05$ level of significance, South West was different to North East, and East was different to North and South East.

Eye movements subject group

The errors were less than 2% across the experiment and there was no evidence of a speed/ accuracy trade off for any of the subjects.

An analysis of variance on the search times taken to locate the matching object for the eye movement subjects was calculated. This showed a main effect of orientation condition $F(4,16)=6.590$, $p=0.0025$ and also a main effect of objects, $F(7,28)=3.466$, $p=0.0085$. No interaction between conditions and objects was found.

Figure 54 shows mean search times of the eye movement subjects for objects rotated in each condition. A Newman-Keuls analysis revealed that there was a significant difference between the 3D60 condition and the SAME, 2D30 and 2D60 conditions at $p<0.01$ level of significance. The 3D60 condition was significantly longer than all other rotations at $p<0.05$ level of significance. No other differences were found.

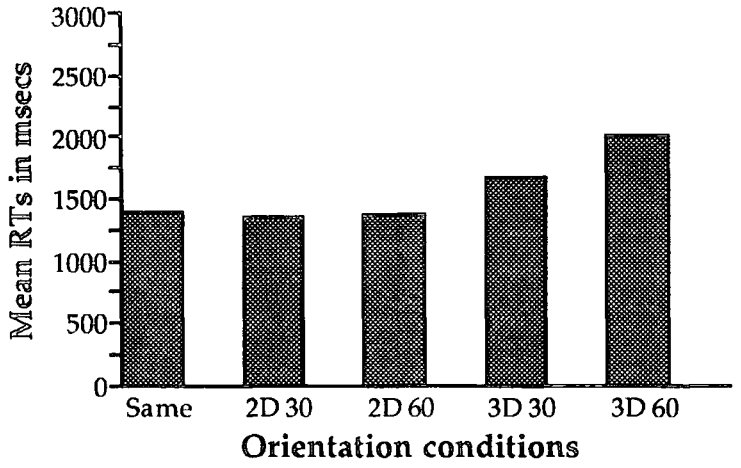

Figure 54; Graph showing mean search times taken to locate object matches in different conditions of orientation by 'eye movement' subjects.

The object effect was subjected to a Newman Keuls analysis which revealed that the cricket bat was significantly different from the screw at $p<0.01$ level of significance. It was also different to the frying pan at $p<0.05$ level of significance. No other differences were found. Figure 55 below shows the mean search times taken to locate each of the match objects.

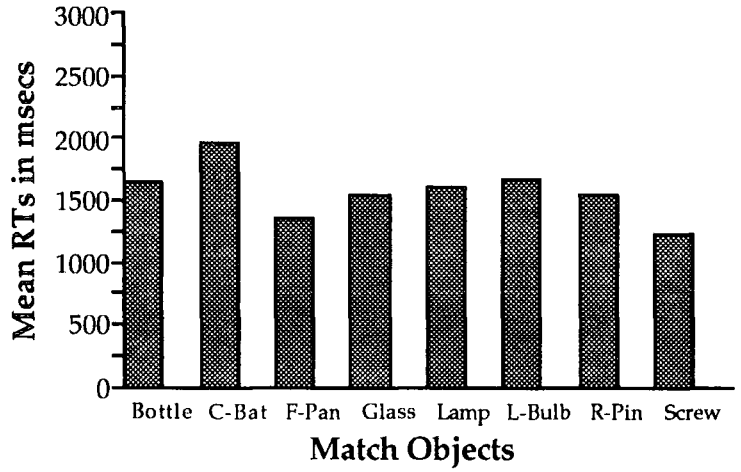

Figure 55: The mean search time taken to locate each of the match objects in the 'eye movement' experiment.

A one-way ANOVA was conducted on the subjects search times to match objects in the different positions in the visual display. There was no effect for position found, $F(7,28)=2.151$, $p=0.0707$. Figure 56 below shows the mean reaction times to the different positions of the objects in the display.

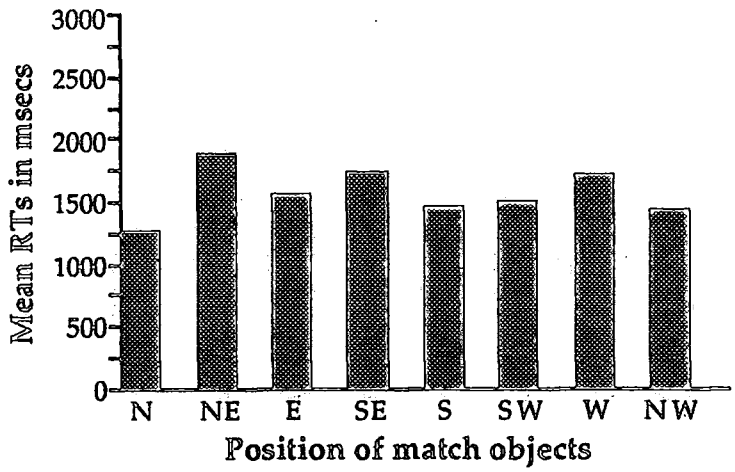

Figure 56; Graph showing mean search times to locate the matching object when shown in different positions in the visual display for 'eye movement' subjects.

All subjects eye movements to each slide were plotted. Figure 57 gives an example of all subjects eye plots to the bottle in the 2D60 condition. It was noted that the target match was not located in 5.5% of the trials. The match was found immediately, i.e. with a single saccade 24.5% of the time across all subjects.

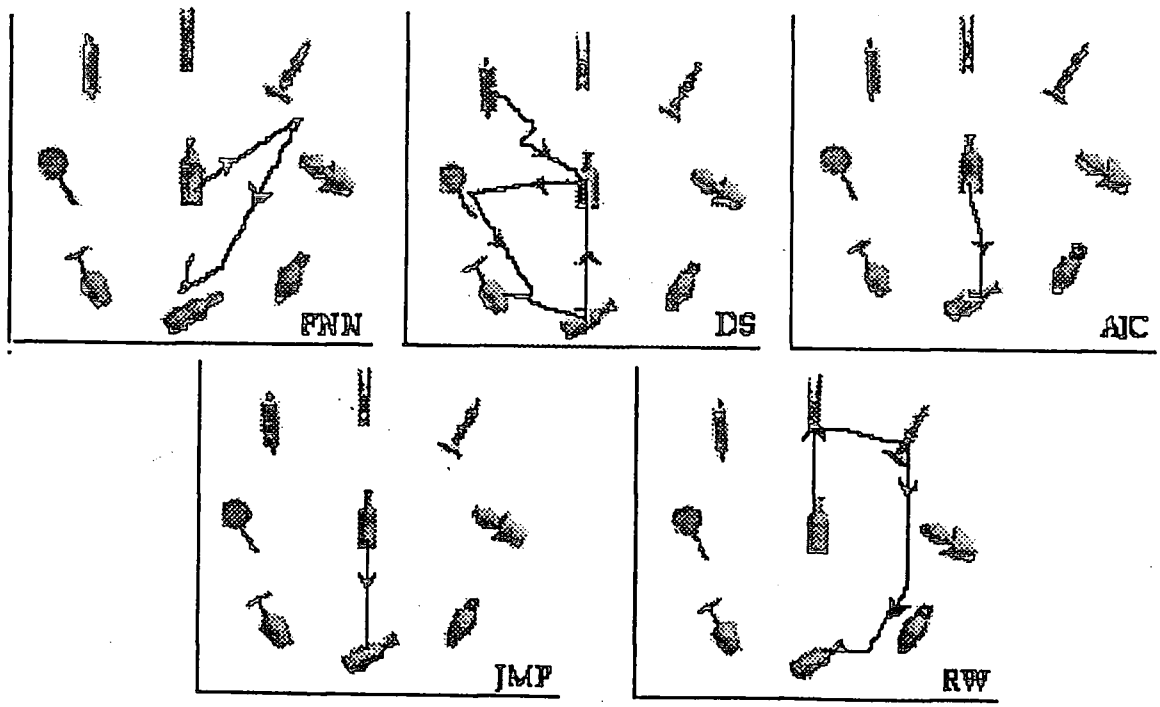

Figure 57: An example of subjects eye movements to a trial in Experiment 8.

It was noted that some subjects directly saccaded to the correct match (subjects AJC and JMF in Figure 57) whereas the other subjects locate the match after a serial-type search from one distractor to the next (subject RW). The number of times all subjects located the correct match by directly saccading to it was subjected to a Friedman analysis of variance. No main effects were found for either condition $Chi_r^2 = 3$, $p = 0.5512$, object $Chi_r^2 = 14$, $p = 0.0533$ or position $Chi_r^2 = 10$, $p = 0.1972$. Figures 58, 59 and 60 below shows the number of direct saccades to the match in each condition, to each object and to each position respectively.

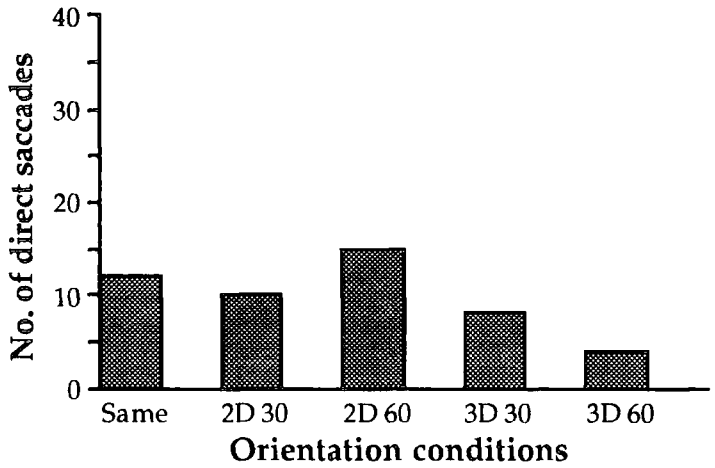

Figure 58; Graph showing total number of direct saccades made to the match object in the different conditions of orientation. The total possible number of direct saccades made in each condition was 40 (Trials x subjects).

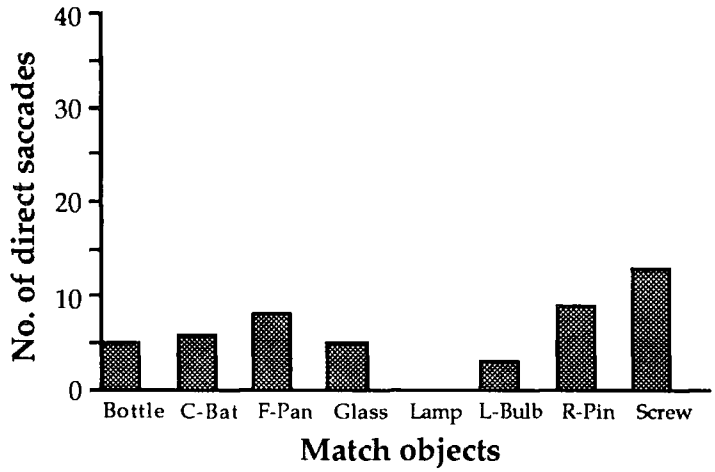

Figure 59; Graph showing total number of direct saccades made to each match object. The total possible number of direct saccades in each condition was 40 (Objects x subjects).

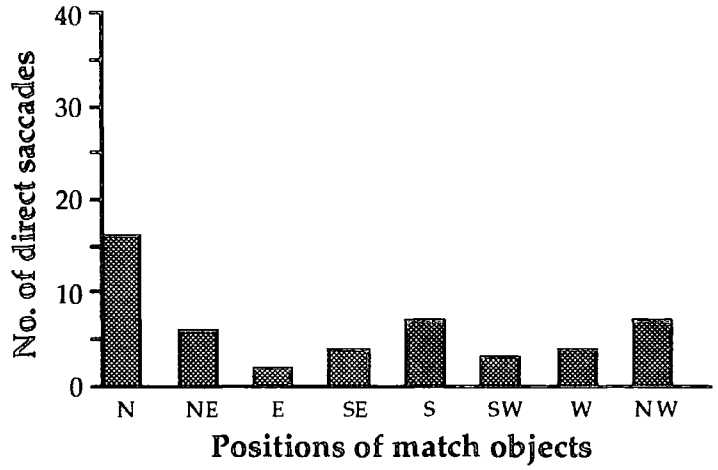

Figure 60; Graph showing total number of direct saccades made to each position. The total possible number of direct saccades to each position was 40 (Positions x subjects).

From Figure 60 above it can be seen that when the match was in the North position it was directly saccaded to more frequently than the other positions. The large number of direct saccades to the North position relative to the other positions is typical of three out of the five subjects (see RW, JMF and DS) in the experiment.

An analysis of variance on the fixation times on the correct match–object for all subjects was carried out. The harmonic mean replaced the missing data cells i.e. the cases where the subject did not find the match. The analysis of this data did not show any main effects of either condition, $F(4,16)=1.194$, $p=0.3513$ or object, $F(6,24)=2.574$, $p=0.0455$. It was noticed that the rolling pin had the most number of missing data cells, especially in the 3D conditions and it was decided to reanalyse the data without including the rolling pin. This still proved insignificant for condition, $F(4,16)=1.889$, $p=0.1615$ and for object, $F(6,24)=2.551$, $p=0.0470$. Figure 61 below shows the mean fixation times to match objects in different orientations.

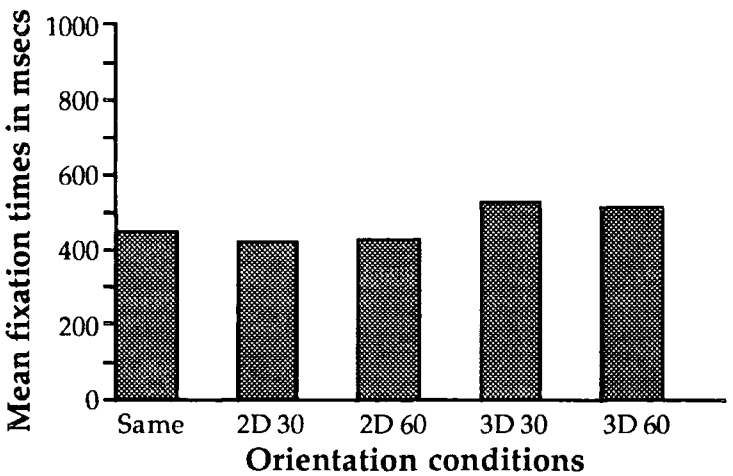

Figure 61; Graph showing mean fixation times to match objects in the different conditions of orientation.

7.2.3 DISCUSSION

The eye movement and search time studies indicated that searching for a disoriented object from among a set of randomly oriented distractor objects was, in general, a difficult task and that a serial search strategy was used to locate the correct match. However, it was found that for both groups of subjects, searching for an object that was rotated 60° in depth (3D60) took longer to locate than all other orientations of the objects. This increase in search times to the 3D60 condition suggests that subjects may have found the correct match only after a careful item by item search. The eye movement data supports this conclusion in that the number of times the 3D60 match object was found with the first saccade was less than chance, although the difference across conditions was not significant. Detection of an object was, therefore, more difficult when the object was rotated in depth by 60°.

It was observed that when the match objects were found in the other orientation conditions then the match was more likely to be found with the first saccade. The search times were faster for these orientations than for the 3D60 condition for both sets of subjects. However a significant increase in the search times from the Same condition to 3D30 and 3D60 conditions was found for the search time group only. This significant increase was not observed for the 2D orientation conditions in either the search time or eye movement group. These data may suggest that searching for an object oriented in depth becomes progressively more difficult as the object is disoriented. It could also be suggested that in searching for an object a 2-dimensional search image is used to map across the features of the objects and therefore matching to a 3D oriented object becomes more difficult. This conclusion is supported by the fact that there was no particular benefit found for the Same condition in the search time data and the eye movement data over the 2D orientations as would be expected if the subject were directly matching the match object with the central target object. In this case, mental rotation effects such as those observed by Shepard and Metzler (1971) would have been found. It could be argued that the objects are identified by matching the match object image with a stored representation of that object and not necessarily with the central object.

The shapes of the objects did not pop out of the array which may have been expected if shape information alone was being used to search for a match. In these experiments it can be argued that shape analysis was not preattentive (Treisman and Paterson 1984). The fact that the correct objects were not detected in peripheral vision suggests that shape is not preattentive in these conditions.

7.3 Experiment 9

The difficulty observed in searching for a match object from among a set of randomly rotated objects may be due to the difficulty of searching for a target in heterogeneous displays

rather than the orientations of the match objects themselves. In other words, the orientation information cannot be used to guide the subject to locate the match object. The search times may therefore be affected by the difficulty of the task rather than the orientations of the match objects themselves. In other words, a serial search strategy may have been used because the displays were heterogeneous over orientations.

The following experiment was designed to test the notion of whether the results found in the previous experiments were affected by the orientation of the match object or whether the random orientations of the surrounding distractors made the task difficult. The orientations of the distractor objects in the following experiment were aligned with the orientation of the match object. The displays were therefore homogeneous across orientations.

7.3.1 METHOD

Subjects

Seven members of the Department of Psychology, University of Durham participated in this study. Five of these subjects were male. Their ages ranges from 25 to 48. All subjects had normal or corrected to normal vision.

Stimuli

See Experiment 8 for a description of the stimuli used. Figure 62 below shows an example of one of the stimuli used in the experiment. The same objects were used in this experiment and the positions of the matching objects were also randomised for this experiment.

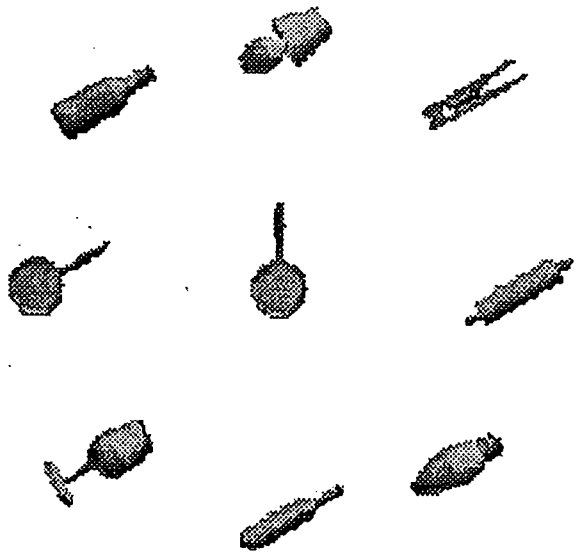

Figure 62: Example of a stimulus from Experiment 9.

Materials and Apparatus

Each stimulus was presented on a Macintosh IIx monitor. There were 80 experimental slides in all; 40 target-present and 40 target-absent.

The radius of the image on the screen was 6 cm. The subject was positioned approximately 57 cm away from the screen which resulted in a visual angle of 6° between the centre object and any of the objects surrounding it.

All of the subjects were required to wear a scleral eye coil which tracked the position of the eye during the search task. As soon as the subject either found the match or decided that a match was absent, the appropriate key on a button box was pressed. The reaction times were recorded by the Macintosh IIx. A delay of 3 seconds followed each response until the presentation of the next stimulus.

The recording apparatus was similar to that used in Experiment 8 except that the Macintosh IIx was used to present and record the data instead of the Alpha computer. Both the reaction times and the eye movements were recorded by the Macintosh.

Design

The experiment was based on a 2 factor, repeated measures design with orientation, condition and objects as factors. The position of the match object was a nested factor. This experiment used the same levels in each factor as in the previous experiment.

The ratio of target present displays to target absent displays was 1:1. The orientations of the distractor objects were aligned to the orientation of the match objects in each trial (see Figure 62 above). The orientations of the objects in the match absent trials were counter balanced across trials. The central, target object in each display remained in the same orientation for all trials i.e. perpendicular to the line of view.

The position of the match in the array was counterbalanced across all objects. The positions of the distractor objects were randomised across all slides.

A practice block of 10 trials preceded the experimental block. The match-present and match-absent conditions were counterbalanced as were the different conditions of orientation across the practice trials. The positions of the objects in the array were randomised across the practice block.

Procedure

Subjects were instructed to locate a match to the object shown in the centre of each display from among the other objects surrounding it as quickly as possible and without making errors. As soon as the match was located they were instructed to press the appropriate 'match present' key or if the match was absent then to press the 'match absent' key on a response box.

A subject was initially presented with a calibration slide which was necessary for a relative measure of the eye movements of each subject. Ten practice trials proceeded the calibration slide. Another calibration followed the practice block. A fixation point preceded each stimulus and subjects were asked to fixate on the centre of the screen before the onset of each slide. A response triggered the offset of each stimulus and the onset of the next stimulus after an inter trial interval of 3 seconds. In the case of a response not made to the stimulus, it would automatically go off after a period of 5 seconds.

The experimental block of trials directly proceeded the practice block. The subject wore the eye coil for the duration of the experiment which lasted about 10 minutes in total. Another calibration was taken after the experimental block. The order of the trials were randomised for each subject.

7.3.2 RESULTS

The mean number of errors made across subjects in each condition was less than 4% and the errors were not subjected to further analysis.

Search times results

A two-factor repeated measures ANOVA was conducted across subject's search times. A significant effect of condition was revealed, $F(4,24)= 7.336$, $p=0.0005$. A significant effect for the objects was also revealed, $F(7,42)= 5.026$, $p=0.0003$. There was a significant interaction between these two variables, $F(28, 168)= 1.586$, $p=0.0402$. Figure 63 shows the mean search times within each of the conditions.

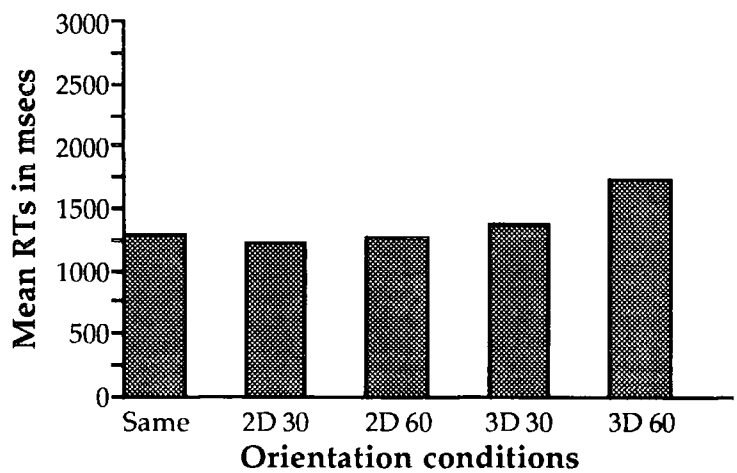

Figure 63: All subjects mean search times within each condition of orientation of the match-objects.

A post-hoc Newman Keuls analysis revealed that search times to the 3D60 condition were significantly longer than search times to all other conditions at $p \leq 0.01$ level

of significance. No other differences were found within this factor.

Figure 64 below shows the mean search times taken to decide whether each of the match–objects was absent or present. The ratio of match–present to match–absent search times was approximately 1:1.5. The effect for objects across the match present trials was analysed and it was found that the search times to the light-bulb were longer than the screw at $p \leq 0.01$ level of significance. The search times to the cricket-bat were also longer than those to the screw at $p \leq 0.01$ level of significance. At $p \leq 0.05$ level of significance search times to the light-bulb were slower than those to the frying-pan, lamp, bottle and rolling pin.

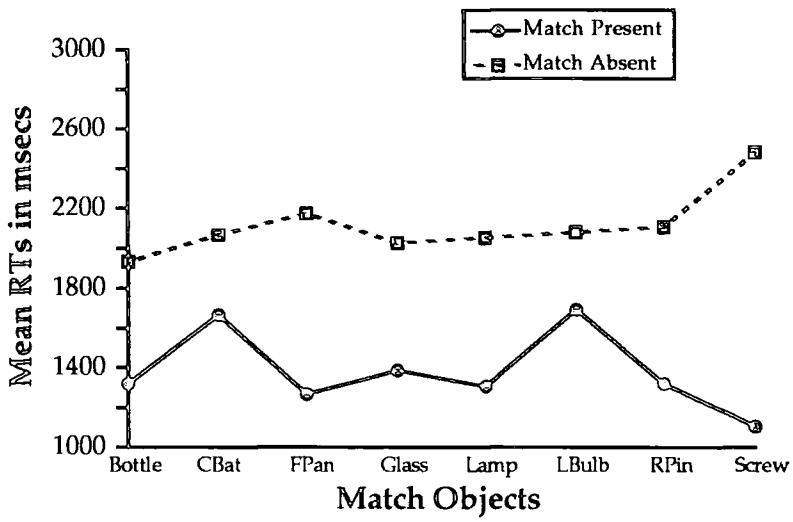

Figure 64: All subjects mean search times to each object in both the match present and the match absent conditions.

The data was also analysed for an effect of the position of the match object. A two-factor, repeated measures ANOVA on the orientation conditions and positions factor proved not significant for position. However, a significant interaction between the conditions and the positions was found, $F(28,168)=3.036$, $p=0.0001$. Figure 65 below shows the mean reaction times to the different positions of the object-matches.

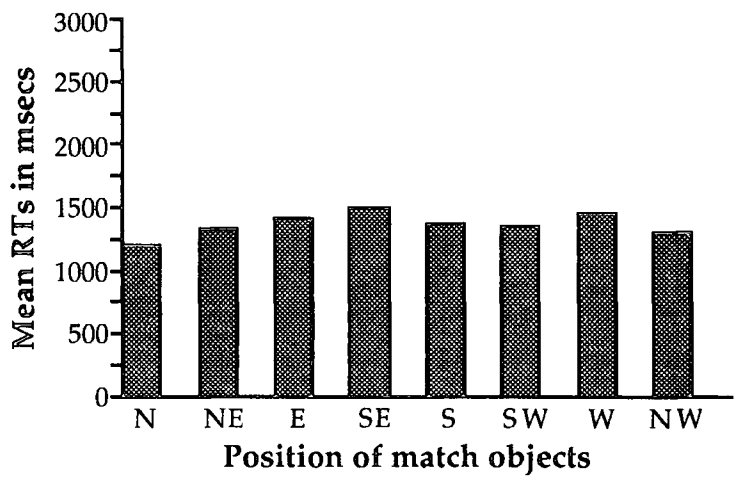

Figure 65: All subjects mean search times to match-objects in each position.

Eye movement results

The total number of times the correct match–object was saccaded to directly in each orientation condition, to each match object and to each position in the display are shown in Figures 66, 67 and 68 below. These data were subjected to a Friedman analysis of variance. There was no difference found in the number of times the match object was saccaded to in each orientation condition, $Chir^2$= 6.575, p= 0.1601. However, a Sign test across each condition showed that 3D60 was significantly less than 3D30,(z=-2.4, p=0.016), 2D60 (z=-2.6, p=0.01), 2D30 (z=-1.961, p=0.049) and Same (z=-2.646, p=0.0082).

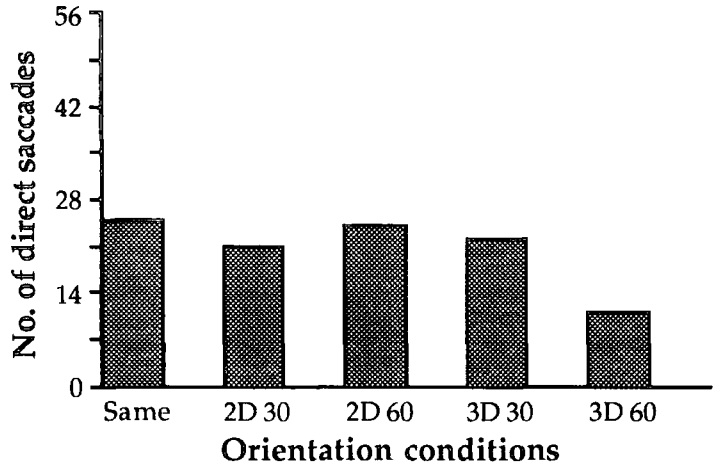

Figure 66: Graph showing total number of direct saccades made to the match object in the different conditions of orientation. The total possible number of direct saccades made in each condition was 56 (Trials per condition ×subjects).

The difference between the number of times each object was directly saccaded proved not to be significant, $Chir^2$=13.05, p=0.0709. Figure 67 below indicates the number of direct saccades made to each object.

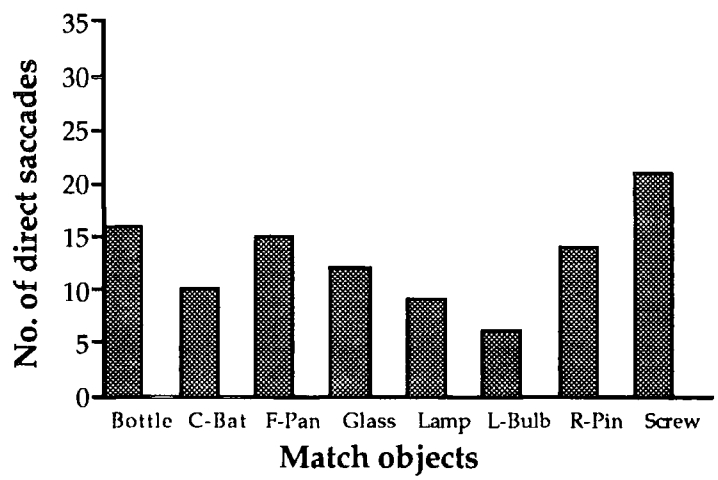

Figure 67: Graph showing total number of direct saccades made to each match object. The total possible number of direct saccades made in each object was 35 (Trials per object ×subjects).

Finally, a significant difference was found between the number of direct saccades to each position, $Chir^2$= 14.283, p=0.0464. Figure 68 below indicates that the North position was saccaded to more often than the other positions and that the South East position was

saccaded to less often than the other positions. This result could have been influenced by one subject who used a similar eye scan for each of the displays by moving the eye to the North position and then clockwise around the display. The subject used this strategy for approximately 70% of the match-present trials.

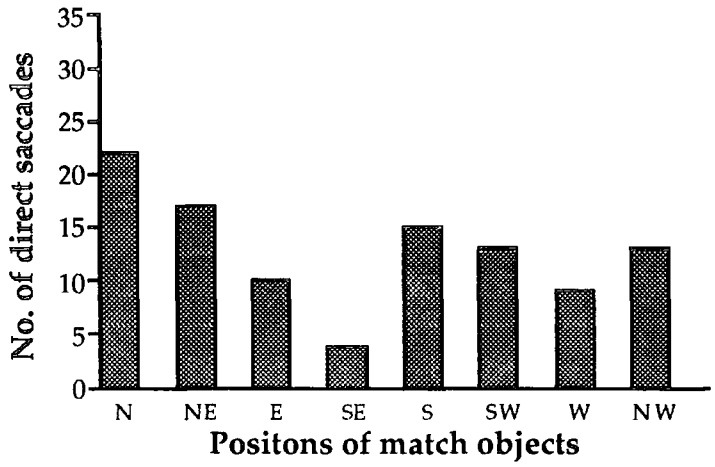

Figure 68: Graph showing total number of direct saccades made to each position of the match object across trials. The total possible number of direct saccades made to each position was 35 (Trials per position ×subjects).

7.3.3 DISCUSSION

The results from this experiment suggest that a search task involving the detection of an object which is aligned with the orientations of the other objects in the display, is difficult and the match is generally found after a serial search through the items in a display. However, it took longer to search for the match object when it was rotated 60° in depth (3D60) and it was less likely to be directly saccaded to suggesting that the overall orientations of the other items in the display cannot guide in the detection of the match object when rotated 60° in depth.

It is interesting to note the effect of reference frames in image analysis: Palmer, Simone and Kube (1988) argued that the overall structure of a configuration drives a perceptual system to choose a reference frame for that image. This idea fits in well with the results of this experiment in that the orientation of the objects in the array can be detected in peripheral vision and the visual system accordingly chooses a reference frame. However, the results show that finding a match to an object which is rotated more than 30° in depth from the upright is more difficult than other orientations. This result supports the previous finding that matching across 2-dimensional transformations is easier than matching across 3-dimensional transformations. The argument that 2-dimensional search representations are used for mapping across different instances of an object is substantiated.

Another conclusion which may be drawn from the results of this experiment and the

previous experiment is that the representations used in searching for the match object are those which are stored in memory and that information other than pure visual information is employed in the search tasks. It could therefore be concluded that the reason why 3D60 orientations are difficult to locate is because either that particular view of the object is not represented in memory or it may be represented but it is difficult to discriminate at that view.

7.4 Experiment 10

In order to establish that the patterns found in the previous experiments are attributed to the representation of the actual objects themselves and not a pattern matching problem per se, the target object in the centre of the array was replaced by the name of the match–object. In other words, the subjects were required to find the matching object to the label that was given. Again the match-objects were seen in each of the five different orientations. The distractors were randomly oriented as in Experiment 8.

7.4.1 METHOD

Subjects

Seven members of the Department of Psychology, University of Durham participated in this experiment. Three of these subjects were female. Their ages ranged from 23 to 33. All subjects had normal or corrected-to-normal vision.

Stimuli

See Experiment 8 for a description of the stimuli used. Each display consisted of the name of an object surrounded by eight objects arranged in a circular array. The objects used were the same as those in the previous experiments. Figure 69 below shows an example of one of the stimuli used in the experiment.

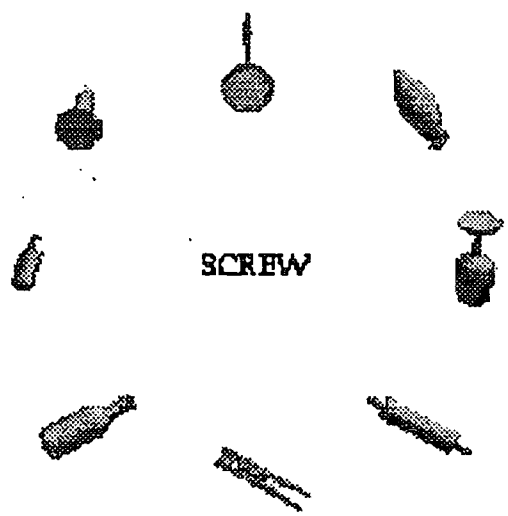

Figure 69: Example of a stimulus from Experiment 10.

As the length of some of the objects names were longer than others (e.g. the word 'frying pan' was longer than the word bottle) the longer words were wrapped around so that they became shorter and did not therefore interfere with the distance between the label and the objects in positions East and West.

Materials and Apparatus

See Experiment 9 for a description of the materials and apparatus used.

Design

The experiment was based on a two factor, repeated measures design. The factors were orientation conditions and objects. The position of the match-objects was a nested factor under objects and orientations. There were equal numbers of match–present and match–absent trials across the experiment. The orientations and the position of the match-object was counterbalanced across all trials. The orientations and positions of the distractors were randomised across all trials.

Procedure

Subjects were instructed to locate a matching object to the label shown in the centre of each display from among a set objects surrounding it as quickly as possible and without making too many errors. The subjects were requested to respond by pressing the 'match present' key as soon as the matching object was located in the array, or the 'match absent' key if the match was not present. The correct/incorrect responses, the reaction times and the eye movements were recorded on the Macintosh IIx computer. The subjects wore a scleral eye coil for the duration of the experiment which lasted about 10 minutes in total.

All subjects were initially presented with a calibration trial of a nine point grid which was necessary as a relative measure of the eye movements of each subject. Ten practice trials proceeded the calibration slide. Another calibration trial followed the practice block. Each objects label was seen at least once in the practice block. There were equal numbers of match-present and match-absent trials in the practice blocks and five different positions were given as an example.

The experimental block of 80 trials proceeded the practice block. The order of the trials was randomised for each subject. A fixation point preceded each stimulus and subjects were asked to fixate on the centre of the screen before the onset of each slide. A response triggered the offset of each stimulus and the onset of the next stimulus after an inter trial interval of 3 seconds. In the case of a response not made to the stimulus, it would automatically go off after a period of 5 seconds.

7.4.2 RESULTS

The mean number of errors made was 7.5% across the match present trials. The errors were not subjected to further analysis.

Search Times results

The mean search times taken to find the match objects in the target only condition are shown in Figure 70. A two way analysis of variance using conditions and objects as factors, revealed that there was no significant difference between the search times in each condition, $F(4, 24)=2.232$, $p=0.0957$. A significant difference was found between the objects $F(4,42)=6.421$, $p=0.2271$.

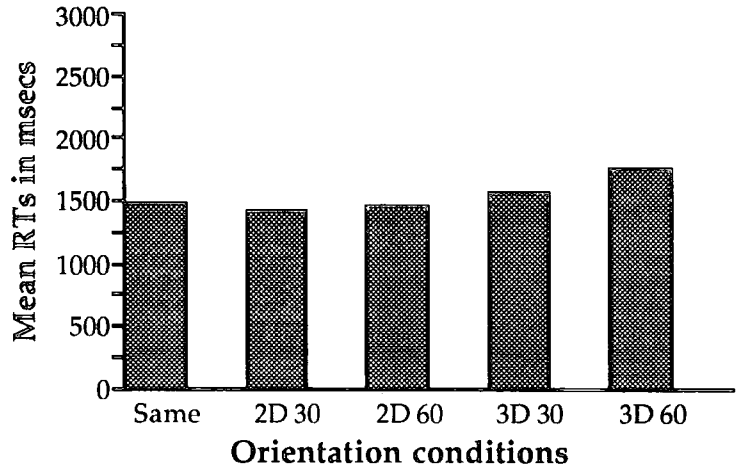

Figure 70; Mean search times to all objects oriented in the different conditions.

A post-hoc Newman-Keuls analysis revealed that the search times to locate the cricket bat were significantly slower than search times to the screw, frying pan and lamp and also search times to the light bulb were significantly slower than those to the screw at $p<0.01$ level of significance. At $p<0.05$ level of significance, the search times to the glass and bottle were also significantly slower than the screw.

The mean search times to the different objects when the match was present and when it was absent are shown in Figure 71 below. The proportion of 'match present' search times to 'match absent' search times was 1:1.5.

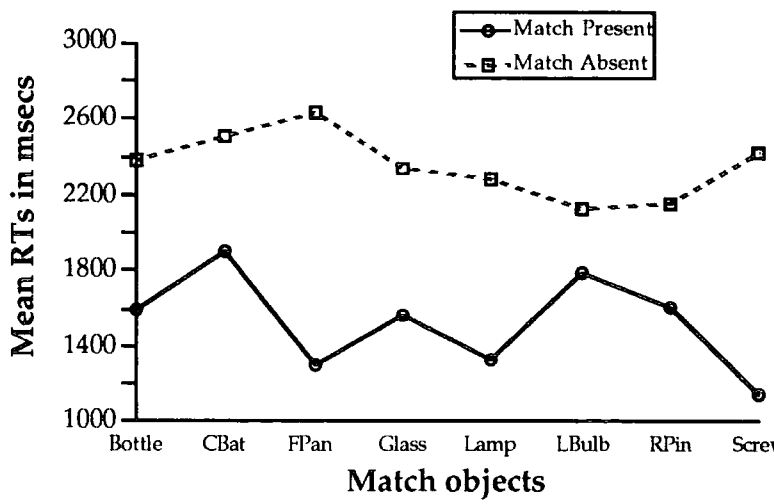

Figure 71: Mean search times to object-matches across the match present and match absent conditions.

A one-way ANOVA was conducted on the search times to the different positions of the match objects. There was no significant effect found for position of the object, F(7,39)=1.0, p=0.6648. Figure 72 below gives the mean search times across all match-present trials when the match was given in each of the positions in the array.

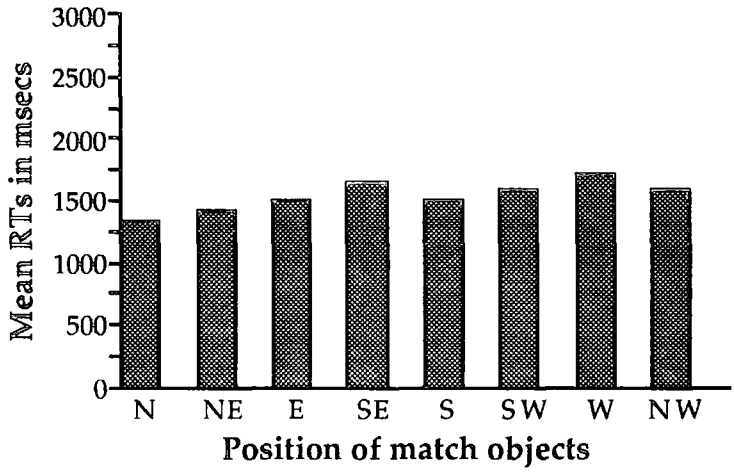

Figure 72: Mean search times to object-matches shown in different positions in the visual display.

Eye movement results

The total number of times the correct match-object was saccaded to directly in each orientation condition, to each match object and to each position in the display are shown in Figures 73, 74 and 75 below. These data were each subjected to a Friedman analysis of variance. There was no difference found in the number of times the match object was saccaded across orientation conditions, $Chi_r^2 = 5.375$, p= 0.2509 (see Figure 73 below). However, a Sign test proved that 3D60 was significantly less than 2D60, z=-2.357, p=0.0184.

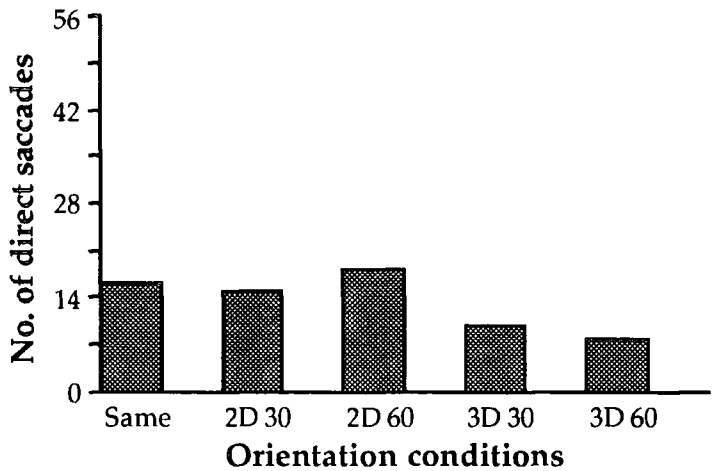

Figure 73: Graph showing total number of direct saccades made to the match object in the different conditions of orientation. The total possible number of direct saccades made in each condition was 56 (Trials per condition ×subjects).

The difference between the number of times each object was directly saccaded also did not prove to be significant, $Chir^2 = 9.95$, $p = 0.1914$. Figure 74 below shows the number of times each match object was found in the first saccade.

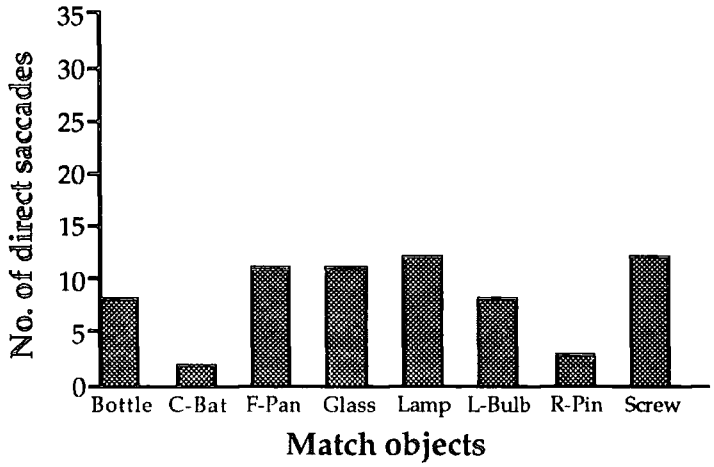

Figure 74: Graph showing total number of direct saccades made to each of the match objects. The total possible number of direct saccades made to each object was 35 (Trials per object ×subjects).

Finally, there was no significant difference found between the number of direct saccades to each position, $Chir^2 = 11.783$, $p=0.1079$. Figure 75 below shows the number of times the match object was saccaded to in each of the different positions.

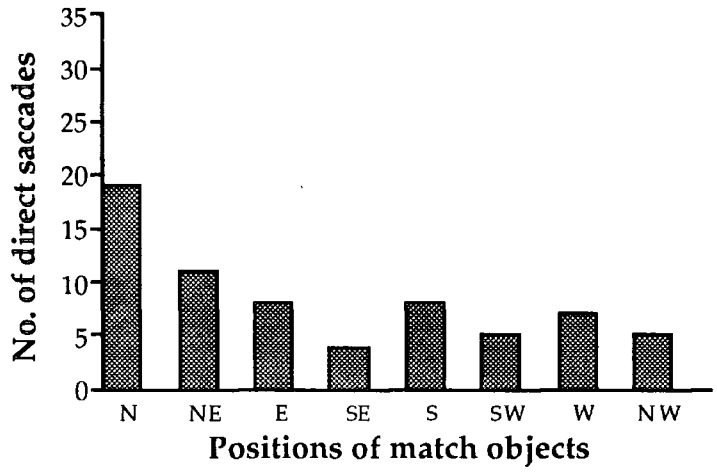

Figure 75: Graph showing total number of direct saccades made to each of the positions match objects. The total possible number of direct saccades made to each position was 35 (Trials per position ×subjects).

7.4.3 DISCUSSION

The results reveal no difference between the search times to objects in the different conditions of orientations. The search times to the match objects that were rotated 60° in depth were not significantly slower than search times to objects in the other conditions of

orientations as was found in the previous experiments. Although not significant, Figure 70 shows that the search times were slower to the 3D60 condition however and that the data followed the same trend as in the previous experiments.

7.5 Experiment 11

For the same reasons outlined in Experiment 9, a study of the efficiency of searching for an object–match to a given label when the object is found in different orientations and surrounded by similarly oriented objects was conducted. It was expected that if the representations of the object was accessed by the label, then searching for objects in different orientations would be more efficient for those orientations that are represented and less efficient for orientations that are not represented. The results of the previous experiment showed that searching for objects orientated 60° in depth are somewhat less efficient than other orientations suggesting that search is inefficient when matching memory representations to objects rotated 60° in depth. The task, however, may have been more difficult given that the distractors were randomly orientated (see Duncan and Humphreys, 1989) and that a search may initially be conducted on objects that are not orientated in depth. A task where the distractors are aligned with the orientation of the match object should yield similar search patterns, i.e. 60° in depth takes longer to find if a memory representation does not directly match that orientation. However, if the results of the previous experiments reflected a preference for saccading to some orientations over others, then no difference should be found between the different orientations. The results from Experiment 10 above suggested that a search based on preferences for orientation was not in fact employed. However, the following experiment was run in order to substantiate the conclusions drawn from the previous experiments that search representations match information in 2-dimensions rather than 3-dimensions.

7.5.1 METHOD

Subjects

Six subjects who were all members of the University of Durham participated in this experiment. There were 4 male and 2 female subjects. Their ages ranges from 23 to 33 years. All subjects had normal or corrected-to-normal vision.

Stimuli

The same set of objects used in the previous experiments were again used in this experiment but the positions of the distractors and the match objects were randomised for this experiment. Figure 76 below shows an example of a stimulus used in the experiment. The name of the match–object was in the centre of the display. For half of the trials the match– object was not present in the display. The positions of the match–objects were counterbalanced. The

positions of the distractors in the array were randomised.

The orientations of the distractors were aligned with the orientations of the match object in a match trial. When the match object was not present in the array the orientations of the distractors was counter balanced across the match-absent trials. The orientations of the distractors in a match-absent trial were aligned with each other and counter-balanced across the match-absent trials.

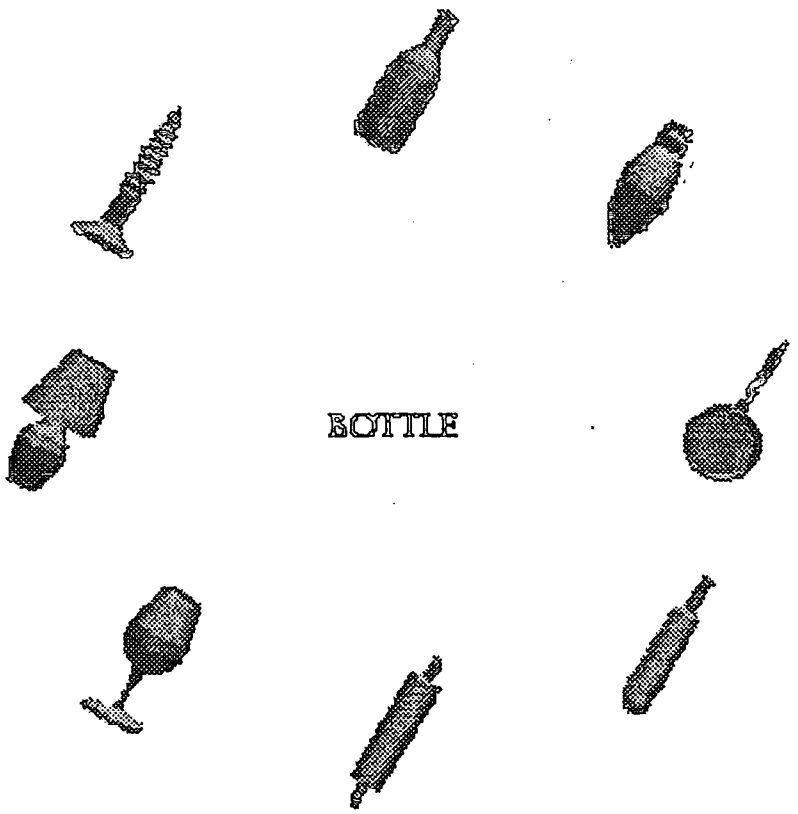

Figure 76: An example of a stimulus from Experiment 11.

Materials and Apparatus

See Experiment 10 for a description of the materials and apparatus used.

Design

The experiment was based on a two-factor repeated measures design with condition of orientation and objects as factors. The position of the match objects was a nested factor. The orientations used were the same as those used in the previous 'search' experiments. The orientations were counter-balanced across all objects. The orientations of the match object and the distractors were the same in all trials. The orientations of the objects across the match

absent arrays were counter-balanced. The position of the objects in the array was randomised therefore a single object was never shown twice in the one position. The positions of the distractors were randomised across all trials.

Procedure

Calibration trials for the eye movement records preceded each of the practice and experimental blocks and the experiment ended with a calibration trial.

Following the first calibration trial the subjects were then presented with a practice block of 10 trials which consisted of 5 match-present and 5 match-absent trials. Each object label was seen at least once in the practice block.

The experimental block which consisted of 40 match-present and 40 match-absent trials immediately proceeded the practice block. Subjects were instructed to locate the object in the array that matches the label given in the centre of the display. They were asked to respond as fast as possible without making too many mistakes. The subjects indicated when they had found the match-object by depressing the 'present' button on the response box and the 'absent' button if the match-object was not present. The subject's response triggered the offset of the trial and the onset of the next trial after an inter-trial interval of 3 seconds. The search times were measured from the onset of a stimulus to the subjects response. The eye movements were only recorded during the onset of each trial. The eye movements and the search times were recorded by the Macintosh IIx.

7.5.2 RESULTS

The percentage number of errors made to the match-present trials across all subjects was 7.083%. The errors were not subjected to further analysis.

Search Times analysis

The search times to the trials where the match-object was present in the display were subjected to a two-way repeated measures ANOVA with orientation conditions and objects as factors. A main effect of condition $F(4,20)=4.652$, $p=0.0081$ and of objects, $F(7,35)=4.118$, $p=0.0021$ was found. There was a significant interaction between the two factors, $F(28,140)=2.113$, $p=0.0024$.

Figure 77 below shows the mean search times to the different conditions of orientations of the objects. A post-hoc, Newman-Keuls analysis on the condition effect revealed a significant difference between the search times to objects oriented in the 3D60 condition and those orientated in the 2D30 condition at $p<0.01$ level of significance. The search times to the 3D60 condition were significantly longer than search times to any other condition at $p<0.05$ level of significance.

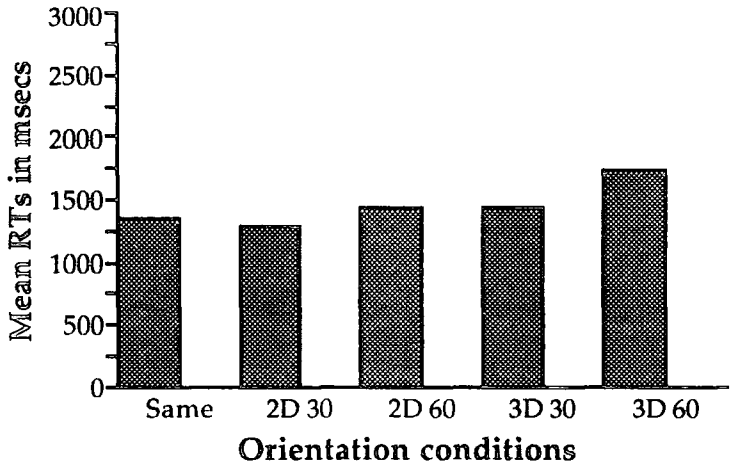

Figure 77: Plot of mean search times for objects shown in the different conditions of orientation.

A post-hoc Newman-Keuls analysis on the objects effect revealed that search times for the light bulb were significantly slower than search times for the screw and the frying pan at $p \leq 0.01$ level of significance. There was a significant difference between search times for the cricket bat and for both the screw and the frying pan at $p \leq 0.05$ level of significance. Figure 78 below gives the mean search times to the different objects in both the match–present and the match–absent conditions.

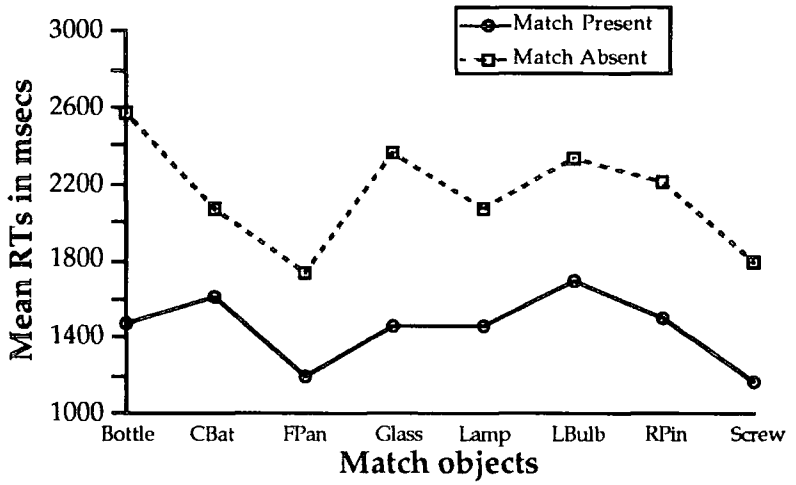

Figure 78: Plot showing mean search times to the different objects in both the match–present and the match–absent conditions.

The mean search times across the different objects in the match present condition was significantly faster than the mean search times to the target absent condition, $F(1,5)=22.014$, $p=0.0054$.

The search times to the match objects in the different positions were subjected to a one-way ANOVA. There was no effect found for the position of the match object in the array, $F(7,28)=1.815$, $p=0.1237$. Figure 79 below shows the mean search times taken to locate the

match object in each of the different positions in the array.

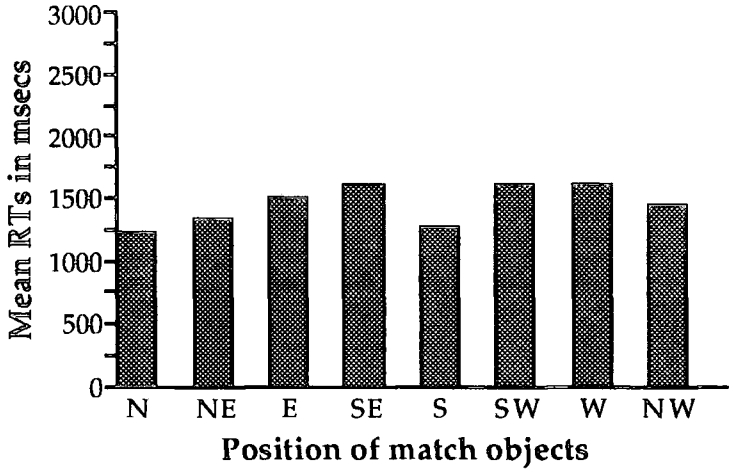

Figure 79: Plot showing mean search times to the different positions of the match-objects.

Eye movement analysis

The total number of times the correct match object was saccaded to directly in each orientation condition, to each match object and to each position in the display are shown in Figures 80, 81 and 82 below. These data were subjected to a Friedman analysis of variance. There was no difference found in the number of times the match object was saccaded to in each condition, $Chi^2 = 1.475$, $p = 0.8311$ (see Figure 80 below).

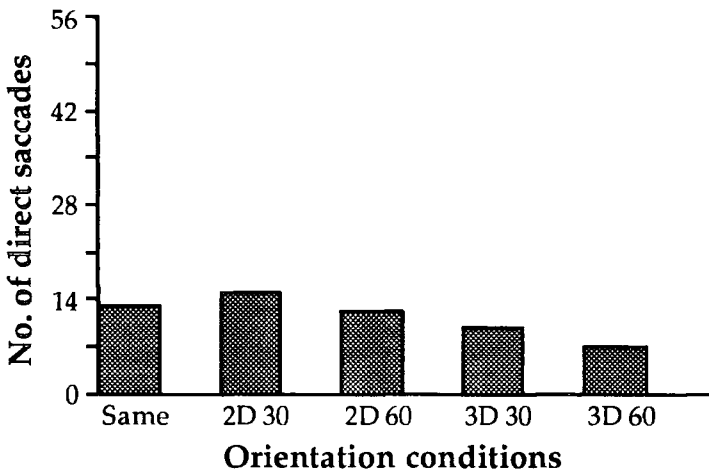

Figure 80: Graph showing total number of direct saccades made to the match object in the different conditions of orientation. The total possible number of direct saccades made in each condition was 48 (Trials per condition ×subjects).

The difference between the number of times each object was directly saccaded also did not prove to be significant, $Chi^2 = 7.567$, $p=0.3723$. Figure 81 below shows the total number of times each object was directly saccaded to.

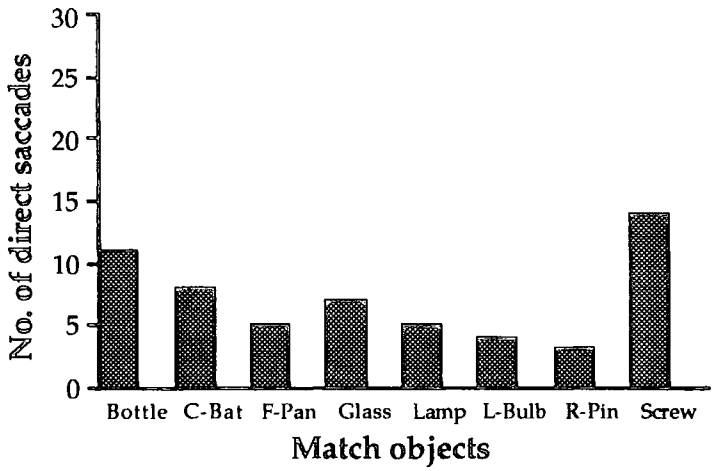

Figure 81: Graph showing total number of direct saccades made to each match object. The total possible number of direct saccades made to each object was 30 (Trials per object ×subjects).

Finally, a significant difference was found between the number of direct saccades to the match object in each position, $Chir^2 = 16.2$, $p = 0.0234$. Figure 82 below indicates that the North position was saccaded to more often than the other positions and that the South East position was saccaded to less often than the other positions. This result could have been influenced by one subject who used a similar eye scan for each of the displays by moving their eyes to the North position and then clockwise around the display in order to locate the match–object. The subject correctly found the match object in the North position 80% of the time.

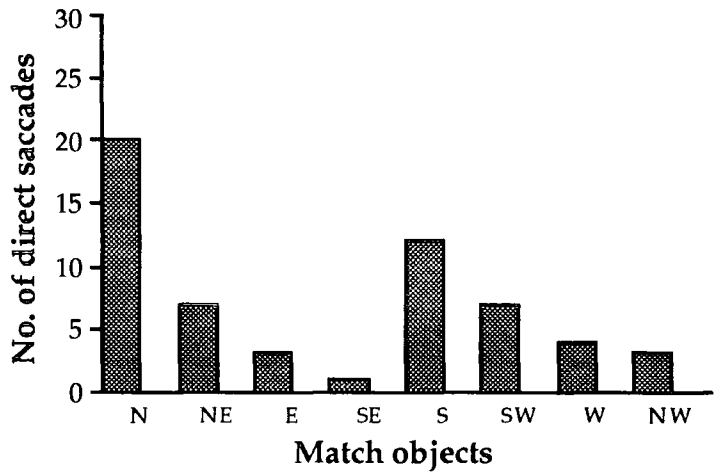

Figure 82: Graph showing total number of direct saccades made to each position of the match objects. The total possible number of direct saccades made to each position of match object was 30 (Trials per position ×subjects).

7.5.3 DISCUSSION

The results of Experiment 11 suggest that searching for an object–match to a given

label that is rotated in depth by 60° is less efficient than search for any other orientation of the object. The time to search for an object rotated 60° in depth is slower than the time to find an object rotated either 30° or 60° in the picture plane or 30° in depth and is also probably less likely to be found in the initial saccade although not significantly so. It could therefore be argued that the information used to match an object seen in peripheral vision to a representation in memory is 2-dimensional rather than 3-dimensional because objects orientated in depth are less readily found.

The results however are confounded by the fact that there was an effect of position found in the first saccade. It could be suggested that when the target is in the North position that it is more likely to be found in the first saccade than if it was in any other position. However, the match objects were found in the other positions more often than chance suggesting that the subjects could use information presented in peripheral vision to guide the eye movements.

7.6 General Discussion

The experiments presented in this chapter looked at what information is used in detecting and identifying an object from among a set of other similar objects that are randomly oriented. The results from the four experiments were consistent. It was found that objects that are rotated in depth by 60° are slower to be detected and somewhat less likely to be directly saccaded to than other orientations of the objects such as the upright, picture plane orientations of 30° and 60° and 30° in depth orientation. These same results were found when the subjects were told to locate a match to a picture of an object and to the name of the object. The results were also consistent over experiments with homogeneous displays or heterogeneous displays.

In sum, the mapping of representations used in search tasks to a picture–match of an object is generally easier for orientations in a 2-dimensional plane than orientations in a 3–dimensional plane. The results for the search time subjects in Experiment 8 however found that objects rotated by 60° in the picture plane took longer to be located than orientations that were the same as the central object or 30° in the picture plane. A difference between the Same, 3D30 and 3D60 conditions was also found for this group of subjects. This difference was not found for the eye movement subjects in Experiment 8 nor were they found for the subjects in the other 3 experiments. There was a larger number of subjects tested in the search time task in Experiment 8 although the difference in subject numbers between this group and the other experiments was, at most, four. As these effects did not generalise across the experiments, they could be considered not representative. It could therefore be argued that search representations are 2-dimensional rather than 3-dimensional because matching was found to be faster in most of the experiments when the 2-dimensional information was preserved. This

conclusion has obvious implications for the nature of the memory representations that are used in search tasks. In the Same condition the match–objects were in the same, upright orientation as the central target object in Experiments 8 and 9. However, there was no obvious benefit found for the detection of the match–objects in this condition relative to other conditions except the 3D60 condition. This result indicated that search did not proceed on purely bottom–up information but that more higher level, top down information influenced search. This conclusion is supported by the fact that the same results were found for experiments where the subject had to locate a match to a picture and to a label suggesting that the same representations of the objects were used across all experiments. It seemed that the same perceptual representations were activated by both a picture of an object and by the objects name. Indeed, the relative search speed between locating a match to a picture and a match to a label were approximately equal across experiments.

When there was sufficient overlap between the representation activated by the central object and the match object in the display then identification was facilitated. Objects that were rotated in the picture plane or by 30° in depth had sufficient information available in order that they were recognised more readily than objects rotated 60° in depth. This result may reflect the nature of the information held in memory representations. The objects rotated in the Same, 2D30, 2D60 and 3D30 orientations were more easily detectable in peripheral vision than objects rotated by 60° in depth. In other words, the objects rotated by 60° in depth were found by an item by item search and detected after fixating on the match object. Objects rotated in depth may need to be fixated longer in order that transformational processes may work on the object to align it to the nearest stored representation (see Just and Carpenter, 1976). Although the fixation times in Experiment 8 did not reveal any significant differences between the fixation times to the different orientation conditions across all match trials, the fixation times to the 3D rotations were longer than to other orientations.

Caution however, must be exercised in interpreting the results of the experiments in terms of the nature of memory representations. It may be that in the latter case, serial search is required because the image is too compressed to decide that it is a match in peripheral vision. In all of the experiments reported above, the image of the match-object rotated 60° into the picture plane was compressed. There was no compression of the image when orientated in the picture plane and slight compression when rotated 30° into the picture plane. The differential effect of search times in detecting a match object rotated 60° in depth may reflect the difference of some metric function such as relative size between the different conditions rather than an orientation difference. Gould and Dill (1969) found that search times were influenced by the relative characteristics of a scene rather than the absolute characteristics. Also Duncan and Humphrey's (1989) found an effect of the ratio of item size to eccentricity in that small letters were a lot more difficult to locate than large letters at increasing eccentricity. Both the relative size difference and the larger size:eccentricity ratio may have been the cause of the longer search times in the 3D60 condition. Although there

was no explicit control for relative size, in Experiments 9 and 11 the displays were homogeneous, i.e. the orientations of the distractors were aligned with the orientations of the match object, therefore there was no difference in the relative sizes between the different objects rotated 60° in depth. The same results were found for these experiments as for experiments with heterogeneous displays or randomly oriented objects (Experiments 8 and 10) suggesting that the relative size did not affect the time taken to detect a match rotated 60° in depth.

The search time ratio between the match–present and the match–absent trials was approximately 1:1.5 and did not conform to the traditional serial search slope ratio of 1:2 although the latter ratio is a function of display size. Search for objects that were found in either a scene containing randomly oriented objects or objects aligned with the match object was not very efficient, however, an item by item search was not required to locate the correct match. This result may indicate that not all items needed to be scanned before deciding that the match–object was not present. A strictly serial search pattern was predicted by Treisman's search model (Treisman, 1986), the Wolfe et al (1989) guidance search model and by Duncan and Humphreys(1989) similarity theory. However, search was not strictly serial suggesting that processes other than feature extraction were employed. The results support the notion that top–down processing was indeed used in searching for familiar objects in both homogeneous and heterogeneous displays and that objects were classified according to more semantic information rather than purely visual information (see Pollatsek et al. 1984).

The eye movement data supported the notion that identification required fixation because all of the subjects fixated the match–object when a correct response was made. The finding that the correct match–object was sometimes directly saccaded across the different orientations suggests that some perceptual information was accessed from peripheral vision. All of the match–objects required fixation before the subject made a response suggesting that identification required focal attention. This result was, however, confounded by the fact the some of the subjects used the same search strategy in looking for the match object across many of the trials. However, the match objects were found more often than chance in the different positions across the experiments which supports the notion that perceptual information can be used from peripheral vision, even information that does not directly match the central target–object in terms of orientation, but that a fixation on the match was required before a response was made.

Chapter Eight

Discussion and Conclusions

8.1 General Overview

The main focus of this thesis has been on the effect of orientation on the recognition of objects. A number of experiments were run in order to determine the nature of the object representations in memory, particularly whether representations are orientation invariant or orientation specific. Another important issue that the experiments addressed was the nature of the information that is extracted from the object image in order to match the image with a stored representation. Finally the nature of the process involved in matching an image with a stored view was also investigated. It was argued that any model of object recognition would need to specify the nature of the stored information and the processes involved in matching between the inputted image and the stored representation for recognition.

Marr (1982) argued that recognising objects across different orientations proceeds by extracting orientation invariant information from the object image to build a 3–dimensional, object–centred representation. These invariant properties are found in the occluding contour of the image. A 3–dimensional object model is built around the principal axis of the object, which is resolved from the information about the edges of the object (Marr and Nishihara, 1978). This model makes a specific prediction on the recognition of objects across different orientations: As long as the time taken to resolve the principal axis remains constant, recognition is invariant over different views. Biederman (1987) also postulated that recognition was invariant over orientation using a different model of recognition than that proposed by Marr. According to Biederman's model, the visual system extracts information from the image about the non–accidental properties of that object such as parallel edges etc.. It is these non–accidental properties that are used to define the basic 3–dimensional components of an object called geons. Object representations are then built from the spatial arrangement of these geons, which in turn determines the identity of the object.

An alternative model of object recognition to the invariant properties model espoused by Marr (1982) and Biederman (1987), proposes that objects are represented as a collection of stored, characteristic views and recognition is fastest to these views (Jolicoeur, 1992). According to a view centred approach, the recognition of novel views is achieved by transforming the image to match the nearest stored view. The transformation process involved in matching an inputted image to the nearest stored view has often been identified as mental rotation (Tarr and Pinker, 1989; Jolicoeur, 1985; Jolicoeur, 1992) although other transformations such as interpolation have also been proposed (Poggio and Edelman 1990;

Edelman and Weinshall 1991; Bülthoff and Edelman 1992).

Much of the evidence used in support of the view–independent and view–dependent models of object recognition involved either 2–dimensional line drawings of the familiar objects (Bartram 1976; Jolicoeur 1985; Biederman 1987; Biederman and Gerhardstein 1992), or 3–dimensional nonsense objects (Rock, DiVita et al. 1981; Rock and DiVita 1987; Tarr and Pinker 1989; Edelman and Bülthoff 1990; Bülthoff and Edelman 1992; Cutzu and Edelman 1992), or a limited number of orientations (Palmer, Rosch et al. 1981; Humphreys 1984; Jolicoeur 1985; Koriat and Norman 1985; Ellis, Allport et al. 1989; Jolicoeur 1990). This thesis attempted to provide an examination of the object–centred and the view–centred approaches by testing the recognition times of computer generated 3–dimensional images of objects shown in a variety of different views along each of the major axis of rotation and their combinations. The initial experiments attempted to provide evidence for the canonical view model proposed by Palmer et al. (1981).

A secondary aim of the thesis was to test the nature of the information accessed from peripheral vision in detecting and identifying an object under different orientation conditions. Previous work on object detection in peripheral vision had found that both visual and semantic information can be accessed (Pollatsek et al., 1984, Biederman et al., 1974, 1982). Four experiments reported in this thesis investigated the effects on search efficiency of detecting an object that could be found in a number of different orientations. Search efficiency was measured for matches to a picture of an object and to the name of an object.

8.2 An Outline of the Main Findings

8.2.1 Effects of Orientation on Recognition and Detection

An initial examination of the nature of the stored representations of objects with particular reference to the canonical view was undertaken (see Palmer, 1981). A number of experiments were run in order to explore the nature of the canonical view of a set of elongated objects particularly, whether there was a single canonical view for each object. Three experiments were run in which the subject had to match a given label with a picture of a familiar object shown in different orientations around the 3 major axes of rotation and their combinations. In all three experiments there was a strong, consistent effect of orientation on recognition times. In Experiment 1 orientation times were shown to increase once the object's foreshortening exceeds a critical value. However, there was no evidence found that recognition times were fastest for any single view of the individual objects. In fact, the results suggested that recognition was fastest to a number of views and were only slower to views where the object was considerably foreshortened. These results were shown to be independent of practice effects (Experiment 2). More importantly, a facilitation effect for the recognition of objects that have strong gravitational uprights when shown in an upright orientation was

not found for objects rotated in depth (Experiment 3) nor for objects rotated in the picture plane (Experiment 4). Previous studies have found that the upright orientation is recognised faster than other orientations in the picture plane (Koriat and Norman 1984; Jolicoeur 1985; Corballis 1988). A comparison between the results of Experiments 3 and 4 and the findings from previous studies such as those reported by Jolicoeur (1985) is discussed in detail in section 8.2.3.

The orientation function was found to be highly consistent across the initial three experiments. On closer inspection of this function it was found that recognition times were the same to a number of different views of the objects but were significantly slower to orientations 30°off the foreshortened view and slowest of all to the foreshortened view. Collectively the results suggest that recognition times were fastest for views that maximise the amount of information about the object. This conclusion resembles the Palmer et al. (1985) definition of a canonical view. However, there is an important difference between the results found in Experiments 1, 2 and 3 and those found in the Palmer et al. study. There was no evidence found that a single canonical view exists for any of the objects. On the contrary, the results support the notion that a number of views are favoured in recognition.

That recognition times were found to be dependent on orientation rejects the notion that representations are orientation invariant and object–centred (Marr and Nishihara 1978; Marr 1982, ; Biederman 1987; Biederman and Gerhardstein 1992). However, the object–centred model must not be rejected on this statement alone without further examination of the findings. The Marr and Nishihara (1978) model postulates that the principal axis needs to be resolved before a 3–dimensional object model that is invariant across viewpoints is built. The information about the axis is derived from the view–centred, $2^1/_2$–D sketch. However, some views of an object may render resolving the axis more difficult and could consequently be a time consuming process. The findings of Experiment 1, 2 and 3 therefore, may have been due to the time taken to resolve the principal axis from the view–centred image. It will be argued however that this is not the case: If the time taken to recognise an object in different orientations reflects the ease at which the principal axis is derived from the $2^1/_2$–D sketch, then the recognition times should increase in a monotonic fashion from the view in which the axis is fully exposed to the view with the axis foreshortened. Recognition times should therefore be fastest for views which have the principal axis fully exposed. A similar prediction can be made from Biederman's 'geon' approach to object recognition: Some views of objects would contain information which would lead to direct access of the representation of the object because the 'geons' would be more easily resolvable in views where they are neither occluded nor accreted. For elongated objects, these views would probably correspond to the views with the axis fully exposed because the information from the edges of objects is maximised in these views. An increasing, monotonic function was not found in any of the experiments and there was no particular benefit for views with the axis fully exposed, therefore object–centred models of recognition can be rejected.

Further evidence that recognition does not proceed according to the model suggested by Marr and Nishihara (1978) was provided in Experiment 6. It was found that priming the orientation of the object had no effect on the orientation function. This finding is potentially damaging to Marr's axis–based model. As was stated previously, according to Marr 3–D representations of objects are built around a reference frame that is intrinsic to the object which is the elongated axis for elongated objects. This reference frame is resolved before the object model is built (see Jolicoeur, 1992) therefore prior presentation of this frame should result in orientation invariant recognition across all orientations of an elongated object if the elongated axis is used as a reference frame to maintain constancy over the different views. As the orientation effect in the priming condition was no different than in the non–primed condition, then the evidence did not support an axis based model of recognition.

As in the recognition time experiments, view specific effects were also found in peripheral detection tasks (Experiments 8, 9, 10 and 11). The subjects task in the search experiments was to locate a match to a target object or name from among a set of similar objects. A highly consistent result was found across four different experiments; search efficiency was independent of orientation unless the object was rotated more than 30° in depth from the upright orientation. There was no significant difference between the search times to objects found upright, rotated 30° or 60° in the picture plane or rotated not more than 30° in depth. It was also found that objects rotated more than 30° in depth were less likely to be found with the initial saccade than other orientations of the object although this result was not, on the whole, significant. The orientation effect on search efficiency was found across both heterogeneous (distractors randomly orientated) and homogeneous (distractors aligned with the orientation of the match object) displays. The same effect was observed whether the subjects had to locate a match to a picture of an object or to the name of the object. This finding suggested that similar representations were accessed by both a picture of the object and the name of the object (see Pollatsek et al., 1984).

The results of the object naming experiments (Experiments 1, 2, 3 and 4) and the object detection experiments (Experiments 8, 9, 10 and 11) are similar in the sense that orientations where the object is considerably foreshortened take longer to recognise and are more difficult to detect. It could be suggested that because objects take longer to recognise when rotated in depth 60° away from the upright than other views that are less foreshortened, then the time to locate an object rotated 60° in depth would also be delayed. In other words, it may be that the search times effect was not due to the lack of the object's detectability among similar distractors when rotated 60° in depth but because it is less readily recognisable in that orientation. Moreover, the fact that the same effects were observed when matching to a name rather than a picture suggests that a memory representation of the shape of the object was accessed. The results of the recognition time experiments suggested that the time delay in recognising objects rotated 30° from the fully foreshortened view (the equivalent orientation in the search experiments was 60° in depth) was due to a transformation process to align the

view with a stored representation. A time consuming transformation process to match the object rotated 60° in depth with a stored representation may therefore have contributed to the orientation effect in the search experiments.

8.2.2 2–Dimensional and 3–Dimensional Image Information

Contrary to the idea that representations are built using surface rather than edge information (see (Biederman and Ju 1988)), Experiment 5 found that there was no difference in recognition times between shaded versions and silhouetted versions of objects. It was concluded that depth cues such as shading information may be another route to object recognition but that there is sufficient information available in the edges of images in order to build a representation of the object.

The results from the recognition time experiments and the search time experiments show that a large 3–dimensional transformation of objects makes them more difficult to recognise and detect. Indeed it was found in Experiment 4 that recognition speed was not altered by a 2–dimensional transformation. These results may suggest that a matching process between the image and the stored representation proceeds by mapping across 2–dimensional features in the image (Intrator, Gold et al., 1991; Ullman 1989; Cutzu and Edelman 1992) and that recognition times are dependent not only on the presence of the features but also on their 3–dimensional orientation. The term 'features' does not necessarily correspond to the sort of early features that Treisman proposes but rather to a cluster of visual contours which is peculiar to each object (see Warrington and James, 1986). Although this proposal is post–hoc and speculative, it is nevertheless supported by the present results. The features in the contour of the image are equally accessible from shaded images of objects and from silhouetted version since the information needed is contained in the edges of the image. In the object detection tasks (Experiments 8 to 11) it could be argued that the results reflect the ease of mapping across features. If mapping occurs across 2–dimensional features without the need for transformations, then it would be expected that 3–dimensional rotations of an image would make 2–dimensional feature mapping more difficult. Indeed it was found that 3–dimensional rotation did make detection more difficult provided the object was sufficiently rotated in depth. Perhaps the reason why rotations of 30° in depth from the upright did not have a delaying effect on search times was because there was sufficient overlap between the represented features and the 2–dimensional projection of the image for the object representation to be directly accessed.

Other studies on the nature of representations have concluded that 2–dimensional rather than three dimensional information is stored as representations of objects (Bülthoff, 1992; Ullman, 1990 and Jolicoeur, 1992). Ullman and Basri (1990) proposed that recognition proceeds by a linear combination of the views of the object (see Bülthoff et al., 1992). In other words the 2–dimensional co–ordinates of a projected image of an object can be represented by a linear combination of the co–ordinates of the corresponding points in a small number of fixed

2–dimensional views of the same object. The required number of views needed to represent an object depends on the 3–dimensional transformations that are allowed by the visual system.

8.2.3 *The rôle of Familiarity on Recognition*

A model of object recognition needs to account for the sort of information that is used to determine the stored representations. According to Marr's model for example, the objects principal axis is a strong determinant of the representation. Palmer et al. (1981) on the other hand argued that information content determines the representation and that therefore a view that maximises the amount of salient information about the object is the view most likely to be represented. However, other investigations have concluded that the familiarity of the view of an object strongly influences the nature of the representation (Jolicoeur 1985; Koriat and Norman 1985; Larsen 1985; Tarr and Pinker 1989). Bülthoff and Edelman (1992) found that recognition times to novel objects were initially strongly dependent on the previously trained views which suggested to them that objects are represented by a collection of stored, familiar views. Jolicoeur (1985) found the recognition times of rotated natural objects increased as the object was rotated away from the upright view. The objects used in Jolicoeur's experiments all had strong gravitational uprights. The results therefore suggest that recognition is dependent on the familiarity of the view because the upright was more familiar than other orientations in the picture plane.

Practice with other views of the objects has been shown to reduce the effect of the most familiar view (Jolicoeur, 1985 and Tarr and Pinker, 1989). Conversely, views that are highly over–learned or familiar would be recognised more efficiently than other views of the object. Faces, for example, are mostly seen upright and are therefore a typical example of highly over–learned view of a stimulus. In recognition experiments where subjects are asked to recognise inverted faces, there is a particular disadvantage observed for the recognition of inverted faces over other stimuli such as landscapes (Diamond and Carey, 1986 and Yin, 1969). Valentine (1988) argued that this inversion effect was due to the effect of familiarity of a stimulus class rather than the specific representation of faces. Recognition times depend on the expertise of processing different views rather than the unique processing of different stimuli. Indeed Diamond and Carey (1986) found that dog breeders were affected by inversion of dogs more than non-dog breeders suggesting that expertise with a certain view of a stimuli reduces the ability to recognise that stimulus in different orientations. Thus support for the notion of familiarity as a determinant of the representations stored of the object can also be found in the literature.

Prior to the recognition time experiments, subjects rated a set of objects according to their typical uprightness in the environment. However, there was no differential recognition effect found for the upright view of objects with strong gravitational uprights for either depth rotated familiar objects (Experiment 3) or objects rotated in the picture plane (Experiment 4). Instead, recognition times were fastest to a number of views in which the

information about the objects was maximised. In other words, there was an effect of canonical aspect observed (i.e. a number of views spanning a range of orientations) rather than a canonical view. These results seemed to suggest that recognition was independent of the most familiar views and to contradict the findings from previous studies that the upright view is the most easily recognisable view (Jolicoeur 1985; Koriat and Norman 1985).

It could be argued that the discrepancy between Jolicoeur's findings and the findings reported in Experiments 3 and 4 is due to the nature of the stimuli used between the experiments. Jolicoeur used line drawings of objects in most of his experiments which are unfamiliar versions of objects (Bartram, 1976)[1] . Bartram (1976) found that subjects were slower to match different views of line drawings of objects than photographs of objects. He concluded that line drawings are coded differently to photographs. Hence the effects observed in Jolicoeur's study may have been due to the unique encoding of line drawings or to the effects of learning to recognise that particular version of the stimuli (see Tarr and Pinker, 1989).

There are links however, between previous studies on the effects of orientation on recognition. Both Jolicoeur (1985) and Tarr and Pinker (1989) found that practice causes recognition time to become more uniform across different orientations. This finding can be applied to the results of Experiment 3 and 4: The objects used were objects with which the subjects were already familiar and hence the observed effects could be affected by this familiarity. In the Tarr and Pinker (1989) study, unfamiliar objects were used and the initial orientation effects observed were due to the unfamiliarity of the different orientations. Similarly, Jolicoeur (1985) found an initial difference between the time to name disoriented familiar objects and disoriented unfamiliar objects (see Experiment 3, Jolicoeur, 1985). The results found in Experiment 3 and 4 therefore, may be dependent on the most familiar views of the objects although the effect of familiarity was not necessarily tested in these experiments.

The effect of previously trained views on recognition times to different views was tested in Experiment 8. It was found that familiarity of the views plays an initial rôle in determining the representations of novel objects but that as other views of the objects become more familiar during the course of the experiment, recognition times become more uniform around those views which maximise the salient information about the objects. Familiarity with the different views of the novel objects in Experiment 8 produced the same effect on recognition as the different views of the familiar objects in Experiments 1 to 6. The results seem to indicate that although representations are initially determined by the most familiar view of the object, equal exposure of other views results in representations which store the maximum information about the object. According to this model therefore, foreshortened views of an object would not be stored as representations unless they were a familiar view (see Experiment 8).

[1] Jolicoeur also tested the effects on naming times of rotated watercolour drawings of objects but the effects observed could have been attributed to orientation invariant information such as the colour of the objects (see Experiment 1, Jolicoeur, 1985).

8.3 Implications of the findings for Theories of Object Recognition

8.3.1 The notion of Characteristic Views

In general the results from the experimental investigation into the recognition of objects from different orientations reported in this thesis suggest that recognition is not orientation independent, nor is it view–dependent in the sense that a single view or canonical view maximises recognition efficiency. Instead, recognition times were found to be fastest to a number of views which were characteristic in that they were views which contained the maximum amount of visual information about the object (Palmer, 1981). These stored views therefore collectively characterise the canonical aspect of the objects.

The notion of characteristic views has been well documented in the literature (Thomas, 1990; Harries et al, 1991; Bülthoff, 1992; Perrett, 1992). Harries et al (1991) found that inspection times to different views of modelled heads were preferential to a small number of these views which corresponded to the face and near-profile views. Similarly, Thomas et al (1991) found that the same stimuli were recognised most efficiently at the single view between face and profile. Perrett et al. (1991) reported finding cells in the macaque visual cortex which are tuned to respond to characteristic views of faces namely the full face and profile views (see also Yamane et al, 1988).

However, the conclusion that elongated objects are represented as a collection of characteristic views centred around the views that maximise the salient information about the object, namely the views which are not foreshortened, needs to be qualified with the observation that not all of the individual objects followed the same effects of orientation. The results of Experiment 1 and 3 for example, suggest that different objects have different characteristic views which are recognised most efficiently. It could be argued that in general most elongated objects are represented by characteristic views in which the elongated axis is not foreshortened. Thus, elongated objects are represented by a collection of views which are on average the most informative views. This is true for both familiar and unfamiliar objects. Further research is needed however in order to generalise the results across a larger set of elongated objects. It may also be the case that non–elongated objects do not have a single canonical aspect but that characteristic views are either distributed more evenly across different orientations or contingent on properties such as an axis of symmetry. Again, this suggestion needs to be investigated with samples of non–elongated objects which can be either symmetrical or non–symmetrical.

Perrett and his coworkers have suggested that a small number of viewer centred descriptions could be used to construct the object centred description and also for direct recognition of the object properties (Perrett and Harries, 1988; Perrett et al, 1992). This ties in with the Humphreys and Riddoch (1984) notion of prototypical views and Palmer's et al. (1981) idea of canonical views of objects i.e. those views that directly access information about the objects properties in order that the object is recognised. A view that does not

directly access this information would take longer to recognise because an object–centred description would have to be constructed before the information is made explicit.

Perrett and Harries (1988) found that in a learning task where the subject needed to acquire enough information about a tetrahedral object or a potato in order to discriminate it from other examples, subjects preferred to look at both end–on and side–on views more than any other views of the objects. That the end–on view was a preferred view is inconsistent with Marr's idea of the elongated axis being important to build an object–centred description of the object. However, Perrett and Harries suggest that different view–centred descriptions are needed in order to build the 3–dimensional object description. In other words, view–centred descriptions are needed to a) determine the principal axis and b) build an object–centred description based on the principal axis. It is during the recognition stage that the object–centred description is accessed and therefore views that include information about the principal axis directly tap this description. However, in Perrett and Harries study, discrimination between the objects was based on the surface patterns of the objects (i.e. squiggly lines drawn on the surface). Therefore the preferential views found in the learning stage may be due to the learning of local, surface features and not the global shape characteristics needed to build 3D object representations. This argument is particularly true for discrimination between the tetrahedra all of which had the same dimensions. The potatoes on the other hand had different shapes. Both these types of stimuli yielded similar results which suggests that the same processes were working on both objects. It is argued that this process involves local feature detection rather than shape information.

Other investigations have found that characteristic, view–dependent descriptions without the need for object–centred descriptions are sufficient as representations of objects provided a transformation is allowed on novel views of objects in order to match them with the relevant stored views (Tarr and Pinker 1989; Edelman and Bülthoff 1990; Edelman and Weinshall 1991; Harries, Perrett et al. 1991; Jolicoeur 1992, Ullman and Basri, 1990). The results of the experiments reported in this thesis suggest that object–centred descriptions are not accessed for recognition purposes (see discussion in 'Effects of orientation on recognition and detection' above).

8.3.2 The Nature of the Transformation Process

Storing a number of 2–dimensional views of the object as opposed to a 3–dimensional model makes the prediction that recognition times will be fastest for views that fall within the space spanned by the stored set of views but will be slower for views that are not stored and therefore need to be transformed to match the nearest stored view. The amount of transformation required to perform this alignment process would depend on how close the image is to a stored view. The time to transform the object could have a linear effect on recognition times dependent on the distance of the image to a stored view. For example, a model that incorporated mental rotation as a transformation would predict such a linear

effect. Linearity is therefore dependent on the nature of the transformation process used. However, other transformations such as view interpolation would have a non–linear effect on the recognition times of novel views of objects (Edelman and Bülthoff 1990; Edelman and Weinshall 1991; Bülthoff and Edelman 1992; Cutzu and Edelman 1992). Furthermore, a model which incorporates feature alignment between the image and the stored representation would also predict a non–linear transformation process (Ullman, 1989). The pattern of the orientation function yielded from the recognition times to the different orientations of objects was non–linear rather than linear (see Experiments 1, 2, 3, 5, 6 and 7).

However, although previous studies have argued that clear linear increases in reaction times with orientation away from a reference orientation such as the upright is evidence for a mental rotation transformation (Jolicoeur 1985; Tarr and Pinker 1989; Jolicoeur 1992), other studies have argued that mental rotation is not strictly associated with linearity (Koriat and Norman, 1985; Cooper and Shepard, 1973). Koriat and Norman (1985) for example, argued that representations of familiar patterns are broadly tuned to orientation such that slight deviations from a stored view does not affect recognition times. In this case results would follow a curvilinear trend rather than a strict linear trend. Indeed Perrett et al. (1991) found that cells that respond to faces can tolerate 60° of rotation in depth before response is reduced to half of its optimal rate. This broad orientation tuning may have contributed to the non–linear effect observed in the initial seven experiments in this thesis. Mental rotation cannot therefore be rejected as a candidate transformation process to match the inputted image with a stored view.

However, there are alternative non–linear transformations which could also have produced the above mentioned orientation effects. For example, Cutzu and Edelman (1992) argued that an interpolation process would produce non–linear effects on response times to novel orientations of views. They found that recognition times of novel views of objects were significantly correlated with the 2–dimensional distances between the features in the image and those in the stored representations. They concluded that non–linear changes in the image plane are a better model of object recognition than transformations including mental rotation. Further evidence from neuropsychological work has suggested that mental rotation is not necessarily the transformation used to align novel views with stored views. Farah and Hammond (1988) reported that a patient, RT, could identify the majority of a set of inverted objects but whose performance in mental rotation tasks was very poor. They concluded that the patients preserved ability to recognise orientated objects despite his deficit in mental rotation does not support the idea that mental rotation is used to transform novel views to match a stored view.

This thesis however, did not set out to establish the nature of this transformation process. Future experimental work is therefore required in order to test the proposal that a transformation other than mental rotation is involved in the recognition of views that do not have corresponding stored counterparts.

8.3.3 Towards a Model of Object Recognition

A model of object recognition can be proposed based on the findings of the experiments reported in this thesis although such a model would be tentative. This model is illustrated in Figure 83 below.

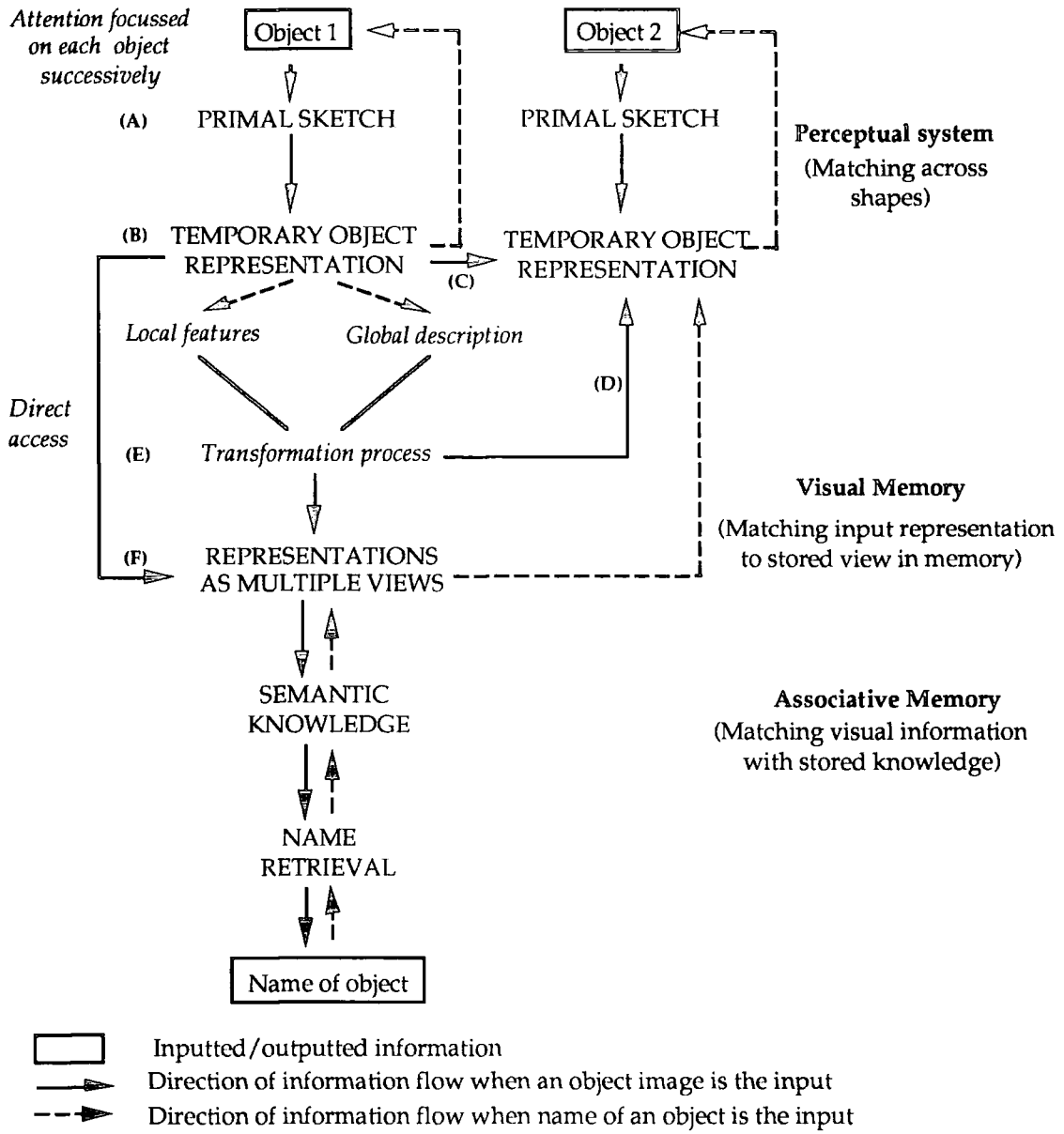

Figure 83: A model of the visual recognition system based on the findings from the experimental investigation of object recognition reported in this thesis. Recognition can be influenced by both top–down and bottom–up information. The direction of information flow outlined in the model should be in both directions rather than in one direction although for clarity, only one direction is given per input type (i.e. object or name of object).

The initial stages of visual recognition involve building a description of the object from the information available in the image. It was found that all subjects fixated onto the match object before responding to its presence in the search tasks (Experiment 8 to 11). This observation suggested that focussed attention may be required in order to identify an object when surrounded by similar distractors and also that attention is capacity limited when identifying an object from among other objects (see Nakayama, 1989). It could be argued that attention is required in order to extract the features of an object, for example, information about the edges such that a temporary object representation is built. It has been argued that representations are built from the information in the edges or occluding contour of the image (Marr, 1980; Hoffman and Richards, 1985; Biederman, 1988 and Warrington and James, 1986. See also Experiment 5). Thus the primal sketch (A in Figure 83) is built from attending to the image and thus extracting the information about the edges of the object.

Treisman argued that a temporary object representation is built from the raw sketch but Treisman's object file consists of an aggregate of features which are different from the sort of features proposed in the present model (see Treisman et al., 1990). At this stage (B) the object has not been matched to a stored representation in memory and is therefore not yet identified. However, it is proposed that such temporary object representations can be matched to each other without the need to identify the objects. The objects can either be directly matched (C in Figure 83) or matched after a transformation of one temporary object representation relative to another (D). In the neuropsychological literature, some agnosic patients have been found that can match across objects without the ability to identify them (associative agnosics) whereas other patients cannot match across shapes of objects (apperceptive agnosia) (see Humphreys and Riddoch, 1984).

The various transformation processes are available at this stage also such that rotated objects or images can be matched without identifying the objects (E in Figure 83). It was argued that the results from the recognition time experiments suggest that a collection of characteristic views are stored as representations of objects and that novel views are in some way transformed to match the nearest stored view. This transformation process could involve either a 3–dimensional transformation such as mental rotation of the global description of the object's image based on say the major axis of the object or 2–dimensional feature mapping across the features of the image and the stored view. Indeed it has been found that either of these two processes may be involved in matching images to stored representations (Humphreys and Riddoch 1984; Warrington and James 1986; Ellis, Allport et al. 1989). Jolicoeur (1992) has recently proposed a dual–systems theory that either feature mapping and mental rotation can be employed for matching an image to a stored representation. Nevertheless, the transformation is a time consuming process and recognition times and detection times of foreshortened objects (e.g. 0°, 180°, 30°, 150°, 210°, 330° views in the recognition time experiments and 3D60 view in the search tasks) are delayed. Identity therefore involves either direct access to the stored representations or a mediating transformation process.

The final stage involves matching the temporary object representation with its stored counterpart (F in Figure 83) which in turn has direct accesses to higher order, semantic information such as the name of the object. These stored representations are collections of different views of objects. The number of views stored per object is limited (Tarr and Pinker, 1989). The nature of the stored representation is initially influenced by the familiarity of the view but when all views are equally familiar then views which hold the maximum amount of information about the object are then stored. In this sense, the memory system is self–organising. The temporary object representations can either directly access one of the stored views of the objects or access can be mediated by a transformation if the view of the image is novel. In order to reduce the search space within which the temporary object representation locates a match, Ullman (1989) proposed that matching can proceed on the basis of minimal information such as a small number of corresponding feature points in the image and the stored representations. The temporary object representations and the stored representations stages of visual memory do not only receive inputs but it can also give feedback to other visual areas. For example, different pictures of familiar objects were found to be recognised equally fast as the same views of object (Bartram, 1976). It has also been found that contextual information in a visual scene can affect the recognition times of an object (Biederman, 1974) and prior information such as an objects name can facilitate the recognition of a picture of an object (Pollatsek et al, 1984). Such findings suggest that recognition can be influenced by both bottom–up and top–down processing (Bravo and Nakayama, 1992).

8.4 Future Research in Object Recognition

The advent of computer packages which allow the design and careful manipulation of 3–dimensional objects is beginning to influence psychological investigations into object recognition. These packages allow careful transformations of object images in more than one dimension. For example, a stimulus can be oriented, positioned, reduced or enlarged precisely whereas previous manipulations of stimuli along these dimensions proved time consuming and tedious. Furthermore, computer generated displays can allow stimulus presentation under controlled conditions such as surface shading and direction of light source. It is expected that computer generated images will be utilised much more in the investigation of theories of object recognition.

This thesis concentrated on the effect of different orientations of elongated objects on recognition times. It was found that a limited number of characteristic views are stored as representations and that these views cluster around the view in which the elongated object is fully exposed. As the characteristic views include information about the axis it could therefore be argued that the axis is a salient feature of elongated objects. The results therefore may not generalise across different shapes of objects that do not have a salient axis. Symmetrical objects may prove to show the same orientation function as elongated objects although further experimentation across a broader selection of objects is needed before this

conclusion can be met.

The results from the reaction time experiments suggested that novel views of objects are transformed to match the nearest stored view. The views of objects 30° away from the foreshortened views were less readily recognised than other less foreshortened views because, it was argued, a time consuming transformation process operated on the image in order to align it to the nearest stored view. Although no attempt was made to identify the transformation process, a number of transformation candidates proposed by other workers were discussed (Tarr and Pinker, 1989; Jolicoeur, 1985; 1992; Shepard and Metzler, 1971; Poggio and Edelman 1990; Edelman and Weinshall 1991; Ullman, 1989). Each of these transformations make different predictions on the recognition times of novel views of objects. It is believed that these predictions can be tested using the object–name matching paradigm used in this thesis and through a more detailed examination of the orientation function found in the experiments in the thesis.

The results of Experiment 4 found that there was no facilitation in recognition times to objects shown in the upright orientation. This result contrasted with the findings reported by Jolicoeur (1985). It was argued however, that his findings may have been affected by the nature of the stimuli that he used, namely line drawings. It would therefore be interesting to test whether there is indeed a difference in the coding of line drawings, silhouettes and shaded objects. It has already been established that silhouettes are recognised as fast as shaded drawings. However, line drawings are not only degraded images of objects but are also unfamiliar versions of objects and as such may produce different effects on recognition times.

Finally, it was found in the search time experiments that the detection of objects rotated 60° in depth from the upright is slower than other rotations of objects. One of the problems discussed in Chapter 7 was that the images of the objects are compressed in the 3D60 condition and that this compression may have affected the results. Consequently, an investigation of the effects of rotations on search times whilst controlling for the size of the images is envisaged.

8.5 Conclusions

Recent discussions of object representations in visual memory have made two main distinctions. Marr (1982) and Biederman (1987) have argued that representations are 3–dimensional, object–centred models which are invariant across orientations while others have proposed that object representations are a collection of 2–dimensional, view–centred views of an object (Tarr and Pinker, 1989; Bülthoff, 1992; Cutzu, 1992; Edelman, 1991; Perrett and Harries,1988; Rock, 1987). The results of a number of experiments that measured the recognition speed and detection speed of a set of elongated objects shown in different orientations suggest that the latter approach may be the more appropriate model of object

recognition. It was shown that objects are represented by a limited number of views which collectively characterise the maximum amount of salient information about the object. For views of objects that were not stored as representations, a time consuming transformation was involved to match novel views to the nearest stored view. Although mental rotation has been a popular candidate for the transformation process on matching the image to a stored representation, the orientation function was non–linear and it was concluded that some other transformation such as 2–dimensional feature mapping was involved. Indeed it was found that recognition was equally efficient for silhouettes of objects as for shaded objects. It was not possible however, to argue definitely for any of the transformations that may be involved.

References

Arbib, M. A. and A. R. Hanson (1987); Vision in Perspective. *Vision, Brain, and Cooperative Computation: An Overview.* MIT Press. Cambridge, Mass.

Arbib, M. A. and A. R. Hanson, Ed. (1987); *Vision, Brain and Cooperative Computation.* Cambridge, Mass., MIT Press.

Arnoult, M. D. (1954); "Shape discrimination as a function of the angular orientation of the stimuli." *Journal of Experimental Psychology* **47** (5): 323-328.

Barlow, H., C. Blakemore and Weston-Smith, M. (1990); *Images and Understanding: Thoughts about images, ideas about understanding.* Cambridge University Press.

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

Bartram, D. J. (1976); "Levels of coding in picture-picture comparison tasks." *Memory and Cognition* **4** (5): 593-602.

Biederman, I. (1987); "Recognition-by-Components: A Theory of Human Image Understanding." *Psychological Review.* **94** (2): 115-147.

Biederman, I., Blickle T. W., Teitelbaum, R.C. and Klatsky, G. (1988); "Object search in nonscene displays." *Journal of Experimental Psychology: Learning, Memory and Cognition.* **14** (3): 456-467.

Biederman, I. and Cooper, E.E. (1991); "Priming contour-deleted images: Evidence for intermediate representations in visual object recognition." *Cognitive Psychology* **23**: 393–419.

Biederman, I. and Cooper, E.E. (1992); "Size Invariance in Visual Object Priming." *Journal of Experimental Psychology; Human Perception and Performance,* **18** (1), 121-133.

Biederman, I. and Gerhardstein P. C. (1992); Recognising Depth-Rotated Objects: Evidence for 3D Viewpoint Invariance. University of Southern California.

Biederman, I., Gerhardstein, P. C., Cooper, E.E. and Nelson, C.A (1992); "High level object recognition without a temporal lobe." *Investigative Ophthalmology and Visual Science* **33** (4): 956-979.

Biederman, I. and Ju G. (1988); "Surface versus Edge-Based Determinants of Visual Recognition." *Cognitive Psychology* **20**: 38-64.

Biederman, I., Mezzanotte R. J. and Rabinowitz, J.C. (1982); "Scene perception: Detecting an judging objects undergoing relational violations." *Cognitive Psychology* **14**: 143-177.

Biederman, I., Rabinowitz J. C., Glass, A.L. and Stacy, E.W. (1974); "On the information extracted from a glance at a scene." *Journal of Experimental Psychology* **103** (3): 597-600.

Blake, A. and T. Troscianko (1990); *AI and the Eye.* John Wiley and Sons Ltd., Chichester, U.K.

Blakemore, C. (1989); *Visual Coding and Efficiency.* Cambridge University Press, U.K.

Bouma, H. (1978); Visual Search and reading: Eye movements and functional visual field: A tutorial review. *Attention and Performance VII.* LEA. 115-147. Hillsdale, U.S.A.

Bravo, M. J. and K. Nakayama (1992); "The role of attention in different visual-search tasks." *Perception and Psychophysics.* **51** (5): 465-472.

Brogan, D. (1990). Visual Search. *1st International Conference on Visual Search.*, Durham, U.K., Taylor and Francis.

Brown, J. M., N. Weisstein, and May, J. (1992); "Visual Search for Simple Volumetric Shapes." *Perception and Psychophysics.* **51** (1): 40-48.

Bruce, V. and P. R. Green (1990); *Visual Perception: Physiology, psychology and ecology. 2nd. ed.* LEA, Hove.

Bülthoff, H. and S. Edelman (1992); "Psychophysical Support for a Two-dimensional View Interpolation Theory of Object Recognition." *Proceedings of the National Academy of Sciences U.S.A.* **89**: 60-64.

Caudil, M. (1990); Introduction to Neural Networks. AI Expert: Neural networks primer. 2–9.

Cave, K. R. and J. M. Wolfe (1990); "Modelling the role of parallel processing in visual search." *Cognitive Psychology* **22** (225-271):

Cooper, L. A. (1990); "Mental Representation of Three-Dimensional Objects in Visual Problems Solving and Recognition." *J.E.P. ; L,M&C.* **16** (6): 1097-1106.

Cooper, L. A. and R. N. Shepard (1973); "The time required to prepare for a rotated stimulus." *Memory and Cognition* **1**: 246-250.

Corballis, M. C. (1988); "Recognition of Disoriented Shapes." *Psychological Review* **95** (1): 115-123.

Cowey, A. (1985); Aspects of Cortical Organisation Related to Selective Attention and Selective Impairments of Visual Perception: A Tutorial Review. *Attention and Performance XI.* L.E.A. 41-62.

Cowey, A. (1991); Grasping the Essentials. Nature. **349**, 365-366.

Cutzu, F. and S. Edelman (1992); Viewpoint-Dependence of Response Time in Object Recognition. Weizmann Institute of Science.

De Renzi, E. (1986); Current issues in prosopagnosia. *Aspects of face processing.* Martinus Nijhoff. 243-252. Dordrecht.

Dehaene, S. (1989); "Discriminability and dimensionality effects in visual search for featural conjunctions: A functional pop-out." *Perception and Psychophysics* **46** (1): 72-80.

Desimone, R. (1992); "The Physiology of Memory - Recordings of Things Past." *Science,* **258** (5080), 245-246.

Diamond, R. and S. Carey (1986); "Why Faces are and are not Special: An Effect of Expertise." *Journal of Experimental 4; General.* **115** (2): 107-117.

Donnelly, N., G. W. Humphreys, and Riddoch, J.M. (1991); "Parallel Computation of

Primitive Shape Descriptions." *D.E.P. Human Perception and Performance.* **17** (2): 561-570.

Duncan, J. (1981); "Direction attention in the visual field; Notes and Comment." *Perception and Psychophysics* **30** (1): 90-93.

Duncan, J. (1983); "Category Effects in Visual Search: A Failure to replicate the "oh-zero" Phenomenon." Perception and Psychophysics, 34 (3), 221-232.

Duncan, J. and G. W. Humphreys (1989); "Visual Search and Stimulus Similarity." *Psychological Review.* **96** (3): 433-458.

Edelman, S. and H. H. Bülthoff (1990); Viewpoint-specific representations in three-dimensional object recognition. A.I. Memo No. 1239. Massachusetts Institute of Technology.

Edelman, S., H. H. Bülthoff, and Weinshall, D. (1989); Stimulus familiarity determines recognition strategy for novel 3D objects. A.I. Memo No. 1138. Massachusetts Institute of Technology.

Edelman, S. and D. Weinshall (1991); "A Self-Organising Multiple-View Representation of 3D Objects." *Biological Cybernetics.* **64**: 209-219.

Efron, R. (1968); What is perception? *Boston studies in the philosophy of science, 4.* Reidel, Dordrecht.

Ellis, A. W. and Young, A. W. (1988); *Human Cognitive Neuropsychology.* LEA, Hove.

Ellis, R., Allport D. A., Humphreys, G.W. and Collis, J. (1989); "Varieties of Object Constancy." *The Quarterly Journal of Experimental Psychology.* **41A** (4): 775-796.

Engel, A.K., Konig, P., Kreiter, A.K., Schillen, T.B. and Singer, W. (1992) " Temporal coding in the visual cortex: New vistas on integration in the nervous system." Trends in Neurosciences, **15**, 218-226.

Enns, J. T. (1988); "Three-Dimensional Features that Pop-Out in Visual Search." *In Brogan,D.(Ed.) Visual Search : Proceedings of First International Conference on Visual Search.* Taylor and Francis, London. :

Epstein, W. and Babler T. (1989); "Perception of Slant-in-Depth is Automatic." *Perception and Psychophysics.* **45** (1): 31-33.

Epstein, W. and Babler T. (1990); "In Search of Depth." *Perception and Psychophysics.* **48** (1): 68-76.

Eriksen, C.W. and Yeh, Y.Y. (1985); "Allocation of attention in the visual field." *Journal of Experimental Psychology; Human Perception and Performance;* **11** (5) 583-597.

Ettlinger, G. (1990); "'Object Vision' and 'Spatial Vision' : The Neuropsychological Evidence for the Distinction." *Cortex.* **26**: 319-341.

Farah, M. J. (1991); "Patterns of co-occurrence among the associative agnosias: implications for visual object representation." *Cognitive Neuroscience* **8**: 1-19.

Farah, M. J. and Hammond K. M. (1988); "Mental rotation and orientation invariant object recognition: dissociable processes." *Cognition* **29** (29-46):

Feldman, J. A. (1985); "Four frames suffice: A provisional model of vision and space." *The Behavioural and Brain Sciences* **8**: 265-289.

Findlay, J. M. and Kapoula Z. (1992); "Scrutinization, spatial attention and the spatial programming of saccadic eye movements." *Quarterly Journal of Experimental Psychology.* **45A** (4): 633-647.

Goodale,M.A., and Milner, A.D. (1992) "Separate visual pathways for perception and action." Trends in Neuroscience, **15**, 20-25.

Gould, J. D. (1969); Eye movements during visual search. IBM Thomas J. Watson Research Centre.

Gould, J. D. and A. B. Dill (1969); "Eye-movement parameters and pattern recognition." *Perception and Psychophysics* **6** (5): 311-320.

Gross, C. G. (1978); "Inferior temporal lesions do not impair discrimination of rotated patterns in monkeys." *Journal of Comparative and Physiological Psychology.* **92** (6): 1095-1109.

Gross, C. G., Rocha-Miranda, C. E. and Bender, D.B. (1972); "Visual properties of neurons in inferotemporal cortex of the macaque." *Journal of Neurophysiology* **35**: 96-111.

Haber, R. N. and M. Hershenson (1980); *The Psychology of Visual Perception.* Holt, Rinehart and Winston, New York.

Harries, M. H., D. I. Perrett and Lavender, A. (1991); "Preferential inspection of views of 3-D model heads." *Perception* **20**: 669-680.

Hayhoe, M., J. Lachter, and Feldman, J. (1991); "Integration across saccadic eye movements." *Perception* **20**: 393-402.

He, Z. J. and K. Nakayama (1992); Surfaces versus features in visual search. Nature. **359** 231-233.

Henderson, J.M., Pollatsek A., and Rayner, K. (1989); "Covert visual attention and extrafoveal information use during object identification." *Perception and Psychophysics.* **45** (3): 196-208.

Henderson, J.M. (1992); "Identifying objects across saccades; Effects of extrafoveal preview and flanker object context." *Journal of Experimental Psychology; Learning, Memory and Cognition,* **18** (3), 521-530.

Hinton, G. E. (1981). A parallel computation that assigns canonical object-based frames of reference. *7th International Joint Conference on Artificial Intelligence,* Vancouver, Canada,

Hinton, G. E. and Parsons, L.M. (1981); Frames of reference and mental imagery. *Attention and Performance.* LEA. 261-277. Hillsdale, N.J.

Hoffman, D. D. and Richards W.A. (1985); Parts of Recognition. *Visual Cognition.* MIT Press. 65-96. Cambridge, Mass.

Holmes, E. J. and Gross C. G. (1984); "Effects of inferior temporal lesions on discrimination of stimuli differing in orientation." *The Journal of Neuroscience* **4** (12): 3063-3068.

Holmes, E. J. and Gross C. G. (1984); "Stimulus equivalence after inferior temporal lesions in monkeys." *Behavioural Neuroscience* **98** (5): 898-901.

Hubel, D. H. and Wiesel T. N. (1962); "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." *Journal of Physiology* **160**: 106-154.

Hubel, D.H. and Wiesel, T.N. (1977); "Orientation columns in Macaque monkeys visual cortex

demonstrated by the 2-deoxyglucose autoradiographic technique." *Nature*, **269** (5626) 328-330.

Hummel, J. E. and Biederman I. (1992); "Dynamic binding in a neural network for shape recognition." *Psychological Review* **99** (3): 480-517.

Humphreys, G. W. (1983); "Reference frames and shape perception." *Cognitive Psychology*, **15**: 151-196.

Humphreys, G. W. (1984); "Shape constancy: The effects of changing shape orientation and the effects of changing focal features." *Perception and Psychophysics* **36**: 50-64.

Humphreys, G. W., Ed. (1992); *Understanding Vision.* Blackwell Publs., Cambridge, Mass..

Humphreys, G. W. and Bruce V. (1989); *Visual Cognition; Computational, Experimental and Neuropsychological perspectives.* LEA, Hove, U.K.

Humphreys, G. W. and Quinlan P. T. (1987); Normal and pathological processes in visual object constancy. *Visual Object Processing.* LEA. 43-99. Hove, U.K.

Humphreys, G. W. and Quinlan P. T. (1988); "Priming effects between two-dimensional shapes." *Journal of Experimental Psychology; Human Perception and Performance.* **14** (2): 203-220.

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

Humphreys, G. W. and Riddoch, M. J. (1987); The fractionation of visual agnosia. *Visual object processing: a cognitive neuropsychological approach.* Lawrence Erlbaum. 281-306. London.

Humphreys, G. W. and Riddoch, M. J. (1987); *To see but not to see: a case study of visual agnosia.* Lawrence Erlbaum, London.

Humphreys, G. W. and Riddoch, M. J. , Ed. (1987); *Visual Object Processing.* Hove, U.K., LEA.

Humphreys, G. W. and Riddoch, M. J. (1991); Interactions between object and space systems revealed through neuropsychology. *Attention and Performance, XIV.* Lawrence Erlbaum. Hillsdale, New Jersey.

Ingle, D. J., M. A. Goodale, and Mansfield, R.J.W., Ed. (1982); *Analysis of Visual Behaviour.* Cambridge, MA., MIT Press.

Intrator, N., Gold J. I., Bulthoff, H.H. and Edelman, S. (1992); "3-D object recognition using unsupervised feature extraction." *Neural Computation* **4** (1): 98-107.

Jolicoeur, P. (1985); "The time to name disoriented natural objects." *Memory and Cognition* **13** (4): 289-303.

Jolicoeur, P. (1990); "Orientation congruency effects on the identification of disoriented shapes." *Journal of Experimental Psychology: Human Perception and Performance.* **16** (2): 351-364.

Jolicoeur, P. (1992); "Identification of Disoriented Objects: A Dual-systems Theory." In Humphreys, G.W. (Ed.)*Understanding Vision.* Blackwell. 180-198. Cambridge, Mass.

Jolicoeur, P., Gluck, M. A. and Kosslyn, S.M. (1984); "Pictures and names: Making the

connection." *Cognitive Psychology* **16**: 243-275.

Jolicoeur, P. and Kosslyn S.M. (1983); "Coordinate systems in the long term memory representations of three-dimensional shapes." *Cognitive psychology* **15**: 301-345.
Jolicoeur, P. and Landau M.J. (1984); "Effects of orientation on the identification of simple visual patterns." *Canadian Journal of Psychology* **38** (1): 80-93.

Just, M. A. and Carpenter P.A. (1976); "Eye fixations and cognitive processes." *Cognitive Psychology* **8**: 441-480.

Koriat, A. and Norman J. (1984); "What is rotated in mental rotation?" *Journal of Experimental Psychology: Learning, Memory and Cognition.* **10** (3): 421-434.

Koriat, A. and Norman J. (1985); "Mental rotation and visual familiarity." *Perception and Psychophysics.* **37**: 429-439.

Kosslyn, S. M. (1980); *Image and Mind.* Harvard University Press, Cambridge, Mass.

Kubovy, M. and Podgorny P. (1981); "Does pattern matching require the normalisation of size and orientation?" *Perception and Psychophysics* **30** (1): 24-28.

Larsen, A. (1985); "Pattern matching: Effects of size ratio, angular difference in orientation and familiarity." *Perception and Psychophysics* **38**: 63-68.

Levine, D. N. and Calvanio R. (1989); "Prosopagnosia: a defect in visual configural processing." *Brain and Cognition* **10**: 149-170.

Lissauer, H. (1890); "Ein Fall von Seelenblindheit nebst einem Beitrage zur Theorie derselben." *Archiv für Psychiatrie und Nervenkrankheiten* **21**: 222-270.

Loftus, G. R. (1983); Eye fixations on text and scenes. *Eye movements in reading: Perceptual and Language processes.* Academic Press. 359-376. New York.

Long, J. and Baddeley, A. D. Eds. (1981); *Attention and Performance.* Hillsdale, N.J., LEA.

Lowe, D.G. (1985); Perceptual Organisation and Visual Recognition. Kluwer Academic, Boston, Mass., U.S.A.

Marr, D. (1982); *Vision.* W.H. Freeman and Co., San Francisco.

Marr, D. and Hildreth, E. (1980). "Theory of Edge Detection." *Royal Society of London, Series B,* **207**, 187-216.

Marr, D. and Nishihara, H. K. (1978); "Representation and recognition of the spatial organisation of three-dimensional shapes." *Royal Society of London, B,* **200**, 269-294.

McCarthy, R. A. and Warrington, E. K.(1990); *Cognitive neuropsychology: a Clinical Introduction.* Academic Press, London.

McClelland, J. L. and Rumelhart, D. E. (1985); "Distributed memory and the representation of general and specific information." *Journal of Experimental Psychology: General* **114**: 159-188.

McClelland, J. L. and Rumelhart, D. E. (1986); *Foundations.* MIT Press, Cambridge, Mass.

Medland, A. J. and Mullineux G.(1988); *Principles of CAD; A coursebook.* Kogan Page Ltd., London.

Milner, A.D., Perrett, D.I., Johnston, R., Benson, P.J., Jordan, T.R., Heeley, D.W., Bettucci, D.,

Mortara, F., Mutani, R., Terazzi, E., and Davidson, D.L.W. (1991); Perception and Action in 'Visual Form Agnosia'. *Brain*, **114**, 405-428.

Mishkin, M. and Appenzeller T. (1987); The Anatomy of Memory. Scientific American. 80-89.

Murphy, G. L. (1991); "Parts in object concepts: Experiments with artificial categories." *Memory and Cognition.* **19** (5): 423-438.

Nakayama, K. (1989); The iconic bottleneck and the tenuous link between early visual processing and perception. *Visual Coding and Efficiency.* Cambridge University Press. 411-422.

Nakayama, K. and Silverman, G. H. (1986); Serial and parallel processing of visual feature conjunctions. Nature. **320** 264-265.

Newcombe, F. and Ratcliff, G. (1982) "Agnosia: A disorder of object recognition." *Les syndromes de disconnexion calleuse chez l'homme.* Colloque International de Lyon.

Palmer, S. E. (1977); "Hierarchical structure in perceptual representation." *Cognitive Psychology* **9**: 441-474.

Palmer, S. E. (1989); "Reference frames in the perception of shape and orientation." In Shepp, B.E. and Ballesteros, S.(eds.) *Object perception: Structure and Process..* LEA. 121-161. Hillsdale, N.J.

Palmer, S. E., Rosch, E. and Chase, P. (1981); Canonical perspective and the perception of objects. In Long, J. and Baddeley, A.D. *Attention and Performance.* LEA. Hillsdale, N.J.

Palmer, S. E., Simone, E. and Kube, P. (1988); "Reference frame effects on shape perception in 2 versus 3 dimensions." *Perception* **17**: 147-163.

Parker, D. M. (1989); "Simultaneous processing of features may not be possible." *Behavioural and Brain Sciences.* **12** (3): 411.

Pentland, A. (1986); "Perceptual organisation and the representation of natural form." *Artifical Intelligence* **28**: 293–331.

Perrett, D. and Harries, M. H. (1988); "Characteristic views and the visual inspection of simple faceted and smooth objects: 'Tetrahedra and potatoes'." *Perception* **17**: 703-720.

Perrett, D., Harries, M. H., Bevan, R., Thomas, S., Benson, P.J., Mistlin, A.J., Chitty, J.K., Hietanen, J.K., Ortega, J.E.. (1989); "Frameworks of analysis for the neural representation of animate objects and actions." *Journal of Experimental Biology* **146**: 87-133.

Perrett, D. I., Harries, M. H., and Looker, S. (1992); "Use of preferential inspection to define the viewing sphere and characteristic views of an arbitrary machined tool part." *Perception* **21**: 497-515.

Perrett, D. I., Oram, M. W., Harries, M.H., Bevan, R., Hietanen, J.K., Benson, P.J., Thomas, S.; (1991); "Viewer-centred and object-centred coding of heads in the macaque temporal cortex." *Experimental Brain Research* **86**: 159-173.

Perrett, D.I., Mistlin, A.J. and Chitty, A.J. (1987); "Visual neurons responsive to faces." *Trends in Neurosciences*, **10** (9) 358-364.

Perrett, D. I., Rolls, E. T. and Caan, W. (1982); "Visual neurons responsive to faces in the monkey temporal cortex." *Experimental Brain Research* **47**: 329-342.

Pinker, S. (1985); *Visual Cognition.* MIT Press, Cambridge, Mass.

Pinker, S. (1985); Visual Cognition: An Introduction. *Visual Cognition.* MIT Press. 1-64. Cambridge, Mass.
Poggio, T. and Edelman S. (1990); A network that learns to recognise three-dimensional objects. Nature. **343** 263-266.

Pohl, W. (1973); Dissociation of spatial discrimination deficits following frontal right and parietal lesions in monkeys. *Journal of Comparative and Physiological Psychology*, **82**, 227-239.

Pollatsek, A., Rayner, K. and Collins, W.E., (1984); "Integrating pictorial information across eye movements." *Journal of Experimental Psychology: General.* **113** (3): 426-442.

Posner, M. I. (1980); "Orienting of attention." *Quarterly Journal of Experimental Psychology* **32**: 3-25.

Price, C. J. and Humphreys G. W., (1989); "The effects of surface detail on object categorisation and naming." *The Quarterly Journal of Experimental Psychology* **41A** (4): 797-828.

Pylyshyn, Z. W. (1973); "What the mind's eye tells the mind's brain: A critique of mental imagery." *Psychological Bulletin* **80**: 1-24.

Quinlan, P. (1991); *Connectionism and Psychology: A psychological perspective on new connectionist research.* Harvester Wheatsheaf, London.

Quinlan, P. T. (1991); "Differing approaches to two-dimensional shape recognition." *Psychological Bulletin.* **109**: 224-241.

Rayner, K., Ed. (1983); *Eye movements in reading: Perceptual and Language processes.* Perspectives in Neurolinguistics, Neuropsychology and Psycholinguistics. New York, Academic Press.

Requin, J., Ed. (1978); *Attention and Performance.* Hillsdale, U.S.A., LEA.

Robertson, L. C., Palmer, S. E. and Gomez, L.M. (1987); "Reference frames in mental rotation." *Journal of Experimental Psychology: Learning, Memory and Cognition.* **13** (3): 368-379.

Rock, I. (1973); *Orientation and Form.* Academic Press, U.S.A.

Rock, I. and DiVita J. (1987); "A case of viewer-centred object perception." *Cognitive Psychology* **19**: 280-293.

Rock, I., DiVita, J. and 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** (4): 719-732.

Rogers, B. J. (1991); Why the eye doesn't shape up. *Nature.* **349**, 365-366.

Rosch, E. (1973); "Natural categories." *Cognitive Psychology* **4**: 328-350.

Rosch, E. (1977); Classification of real-world objects: Origins and representations in recognition. *Thinking: Readings in Cognitive Science.* Cambridge University Press.

Rosch, E., Mervis, C. B., Gray, W.D., Johnson, D.M., and Boyes-Braem, P., (1976); "Basic objects in natural categories." *Cognitive Psychology* **8**: 382-439.

Schneider, W. and Shiffrin R. M. (1977); "Controlled and automatic human information

processing: I: Detection, search and attention." *Psychological Review* **84** (1): 1-66.

Shepard, R. N. and Hurwitz S., (1985); Upward direction, mental rotation, and discrimination of left and right turns in maps. *Visual Cognition.* MIT Press. 161-194. Cambridge, Mass.

Shepard, R. N. and Metzler J. (1971); "Mental rotation of three-dimensional objects." *Science* **171**: 701-703.

Shepherd, M., Findlay, J. M. and Hockey, R.J. (1986); "The relationship between eye movements and spatial attention." *Quarterly Journal of Experimental Psychology* **38** : 475-491.

Shepp, B. E. and Ballesteros, S. Eds. (1989); *Object Perception: Structure and Process.* Hillsdale, N.J., LEA.

Snodgrass, J. G. and Feenan K. (1990); "Priming effects in picture fragment completion: Support for the perceptual closure hypothesis." *Journal of Experimental Psychology: General* **119** (3): 276-296.

Snodgrass, J. G. and Vanderwart M. (1980); "A standardised set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity." *Journal of Experimental Psychology: Human Learning and Memory* **6** (2): 174-215.

Spillman, L. and Werner, J. S. Ed. (1990); *Visual perception: The neurophysiological foundations.* USA, Academic Press.

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

Thomas, S., Perrett, D. I., Davis, D.N. and Harries, M.H.. (1991); Effect of Perspective View on Recognition of Faces. Unpublished document, University of St. Andrews.

Todd, J. T. and Mingolla E. (1983); "Perception of surface curvature and direction of illumination from patterns of shading." *Journal of Experimental Psychology: Human Perception and Performance.* **9** (4): 583-595.

Treisman, A. (1986); Features and objects in visual processing. Scientific American. 106-115.

Treisman, A., Cavanagh, P., Fischer, B., Ramachandran, V. and von der Heydt, R.. (1990); Form perception and attention - Striate cortex and beyond. *Visual Perception: The Neurophysiological Foundations.* Academic Press. USA.

Treisman, A. and Gelade, G. (1980); "A feature integration theory of attention." *Cognitive Psychology,* 12 (1), 97-136.

Treisman, A. and Gormican S., (1988); "Feature analysis in early vision: Evidence from search asymmetries." *Psychological Review* **95** (1): 15-48.

Treisman, A. and Paterson R. (1984); "Emergent features, attention and object perception." *Journal of Experimental Psychology: Human Perception and Performance.* **10** (1): 12-31.

Treisman, A and Schmidt, H. (1982); "Illusory conjunctions in the perception of objects." *Cognitive Psychology,* **14** (1), 107-141.

Tversky, B. and Hemenway, K. (1984); "Objects, parts and categories." *Journal of Experimental Psychology; General.* **113**: 169-193.

Ullman, S. (1985); Visual Routines. *Visual Cognition.* MIT Press. 97-160. Cambridge, Mass.

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

Ullman, S. and Basri R. (1990); Recognition by linear combination of models. Massachusetts Institution of Technology.

Ungerleider, L. G. and Mishkin M. (1982); Two cortical visual systems. *Analysis of Visual Behaviour.* MIT Press. 549-586. Cambridge, MA.

Vaina, L. M. and Zlateva S. D. (1990); "The largest convex patches: A boundary based method for obtaining object parts." *Biological Cybernetics* **62**: 225-236.

Valentine, T. (1988); "Upside-down faces: a review of the effect of inversion upon face recognition." *British Journal of Psychology* **79**: 471-491.

Valentine, T. (1991); Representation and process in face recognition. *Pattern recognition by man and machine.* Macmillan. Volume 14 in J.R. Cronly-Dillon (Ed.), Vision and visual dysfunction, ed. 107-124. Basingstoke.

Warren, C. E. J. and Morton J. (1982); "The effects of priming on picture recognition." *British Journal of Psychology* **73** (117-130):

Warren, C. E. J. and Morton J. (1982); "The effects of priming on picture recognition." *British Journal of Psychology.* **73**: 117-130.

Warrington, E. K. (1982); "Neuropsychological studies of object recognition." *Philosophical Transactions of the Royal Society of London* **B** (298): 15-33.

Warrington, E. K. (1985); Agnosia: the impairment of object recognition. *Handbook of clinical neurology, 1: clinical neuropsychology.* Elsevier. 333-349. Amsterdam.

Warrington, E. K. and James M. (1986); "Visual object recognition in patients with right-hemisphere lesions: axes or features?" *Perception* **15**: 355-366.

Watt, R. (1988); *Visual Processing: Computational, Psychophysical and Cognitive Research.* LEA, Hove.

Weiskrantz, L. and Saunders R. C. (1984); "Impairments of visual object transforms in monkeys." *Brain* **107**: 1033-1072.

Wolfe, J. M., Cave K. R., and Franzel, S.L.. (1989); "Guided search: An alternative to the feature integration model of visual search." *Journal of Experimental Psychology: Human Perception and Performance.* **15**: 419-433.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M.I. and O'Connell, K.. (1992); "The role of categorisation in visual search for orientation." *Journal of Experimental Psychology: Human Perception and Performance.* **18** (1): 34-49.

Wolfe, J. M., Yu, K., Stewart, M.I., Shorter, A.D., Friedman-Hill, S.R. and Cave, K.. (1990); "Limitations on the parallel guidance of visual search: Colour X colour and orientation X orientation conjunctions." *Journal of Experimental Psychology: Human Perception and Performance.* **16** (4): 879-892.

Yamane, S., S. Kaji, and Kawano, K. (1988); "What facial features activate face neurons in Inferotemporal cortex of monkeys." *Experimental Brain Research.* **73**: 209-214.

Yin, R. K. (1969); "Looking at upside-down faces." *Journal of Experimental Psychology* **81** : 141-145.

Young, A. W., Hellawell D. J., and Hay, D.C. (1987); "Configural information in face perception." *Perception* **16**: 747-759.

Young, M. P. (1992); Objective analysis of the topological organisation of the primate cortical visual system. Nature. **358** 152-155.

Zeki, S. (1992); The visual image in mind and brain. Scientific American. **267:** 43-50.