DOCTOR OF PHILOSOPHY

Eye movements and the visual perception of shape

Davitt, Lina

Award date:
2012

Awarding institution:
Bangor University

Eye movements and the visual perception of shape

Lina I. Davitt

Submitted for the degree of Doctor of Philosophy

2012

School of Psychology

Bangor University

Supervisor: Professor Charles Leek
School of Psychology
Bangor University
UK

External examiner: Professor Toby Lloyd -Jones
School of Psychology
Swansea University
UK

Internal examiner: Dr. Paloma Mari-Beffa
School of Psychology
Bangor University
UK

Declarations

This work has not been previously accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed …….... (Candidate)

Date ……....

STATEMENT 1

This dissertation is the result of my own independent work/investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A list of references is appended.

Signed …….... (Candidate)

Date ……....

STATEMENT 2

I hereby give consent for my dissertation, if accepted, to be available for photocopying and for inter-library loans, and for the title and summary to be made available to outside organisations.

Signed …….... (Candidate)

Date ……

Acknowledgments

First and foremost I want to express my gratitude to Prof. Charles Leek for giving me the opportunity to work under his supervision. He has provided the vision, encouragement and all the necessary advice and support which helped me to progress and reach this stage of my PhD (although I have to pinch myself to make sure that I am not dreaming). Having such an enthusiastic and refreshingly human supervisor has helped me to develop my research skills in a friendly environment and despite that the PhD research has proven to be the most difficult task I have ever faced, I have enjoyed every minute of the journey.

Special thanks to my committee, Dr. James Intriligator and Dr. Simon Watt, for their support, guidance and helpful suggestions.

I want to extend a huge thank you to Dr. Filipe Cristino, for the vast help with MatLab programming which speeded up the analyses and allowed me to finish my PhD on time; and for his friendly invaluable advice. I am also grateful to Wendy Adams and Steve Davitt for having the dubious pleasure for proof reading my thesis. Last, but not least I want to thank my brilliant office mates, Pamela Arnold (in the first two years), and Julie Kerai (during the final year), for their cheerful presence at all time, and great fun during our extended lunches.

I also wish to thank my family for their love, constant support and encouragement. Finally Steve, thank you for your caring love, and all the fabulous meals you have cook for me every day. Meeting you has transformed my life in a wonderful way, and I am looking forward spending more time with you.

Lina Davitt
Bangor University
July, 2012

Contents

List of Tables

List of Figures

Abstract

This thesis reports the results of five novel studies that used eye movement patterns to elucidate the role of shape information content of object shape representation in human visual perception. In Experiments 1, and 2 eye movements were recorded while observers either actively memorised or passively viewed different sets of novel objects, and during a subsequent recognition memory task. Fixation data were contrasted against different models of shape analyses based on surface curvature bounding vs. internal contour and low level image visual saliency. The results showed a preference for fixation at regions of internal local features (either concave or/and convex) during both active memorisation and passive viewing of object shape. This pattern changed during the recognition phase where there was a fixation preference towards regions containing concave surface curvature minima. It is proposed that the preference of fixation at regions of concavity reflect the operation of a depth-sensitive view interpolation process that is constrained by key points encoding regions of concave curvature minima.

Experiments 3 and 4 examined the extent to which fixation-based local shape analysis patterns are influenced by the perceptual expertise of the observer and the level of stimulus classification required by the task. These studies were based on the paradigm developed by Wong, Palmeri & Gauthier (2009) in which observers are extensively trained to categorize sets of novel objects (Ziggerins) at either a basic or subordinate level of classification. The effects of training were measured by comparing performance between a pre- and post-test sequential shape matching task that required either basic- or subordinate-level judgements. In addition, we also recorded fixation patterns during the pre- and post-tests.

The results showed significant effects of training on shape matching in the post-tests. In particular, participants showed evidence of perceptual expertise in making basic and subordinate-level shape classification judgements. We also found that the acquisition of perceptual expertise did not result in significant changes in the local spatial distributions of fixation patterns observed. However, there was a tendency for fixations located at areas with concave curvature minima regardless of level of classification. This finding provides evidence that the preference for fixating at concave regions generalises across levels of stimulus classification in recognition tasks.

The last study examined how eye movement patterns can be used to elucidate shape analyses strategies across tasks of object recognition and planning prehensile movement where participants were asked to memorise and later recognise an object and to imagine picking up an object by using their thumb and a forefinger. The results showed a significantly different fixation pattern between the recognition and motor imagery task, thus providing support for differential processing during shape perception influenced by task demands.

The main empirical findings in this thesis show: 1) How eye movements can elucidate properties of internal mental representations of shape; 2) Consistent fixation pattern to concave areas that generalises across tasks; and 3) Different fixation patterns during recognition and motor imagery task.

Chapter 1

1.1. Introduction

'The eye sees only what the mind is prepared to comprehend'
Henri-Louis Bergson, 1859-1941

Our ability to detect, classify and recognise objects despite their high variability and changes in scale, size, position, illumination, and configuration seems effortless and yet, no computer system has been devised to emulate the robustness of object recognition in human visual system. A vast body of research has attempted to answer the question how are we able to perceive the three-dimensional (3D) world around us -so quickly and effortlessly- from the two-dimensional (2D) pattern of retinal images. This fundamental question has provided the focal point for investigations from a variety of disciplines such as psychology, physiology, biology, neuropsychology and computer science, and has opened the way for many more questions in the course of the investigation. Such as how do we perceive and represent object shape, one of the most informative properties for object recognition.

The general aim of this thesis is to investigate the nature of object shape representation. It is generally accepted that amongst all the information required to recognise object shape, edges and contours have the primary shape defining properties; the first stage of visual processing (V1) seems to respond to object edges (e.g., Hubel & Wiesel, 1968; Lee, 2003) followed by other informative properties (e.g. surfaces). A great deal of interest from a variety of disciplines has been given to the information located along object edges and contour curvature (e.g., Attneave, 1954; Feldman & Singh, 2005; Hoffman & Richards, 1984). Attneave (1954) was one of the

first to suggest that contour curvature magnitude carries more shape relevant information for perception than regions of straight contour segments. Subsequent research (Barenholtz, Cohen, Feldman & Singh, 2003; Cohen, Barenholtz, Singh & Feldman, 2005; Cohen & Singh, 2007; De Winter & Wagemans, 2006; Feldman & Singh, 2005; Hoffman & Richards, 1984; Hoffman & Singh, 1997) has shown that the sign of curvature, and not just the magnitude, carries essential informational content for shape perception. An Illustration of the sign of curvature see Figure 1 below.

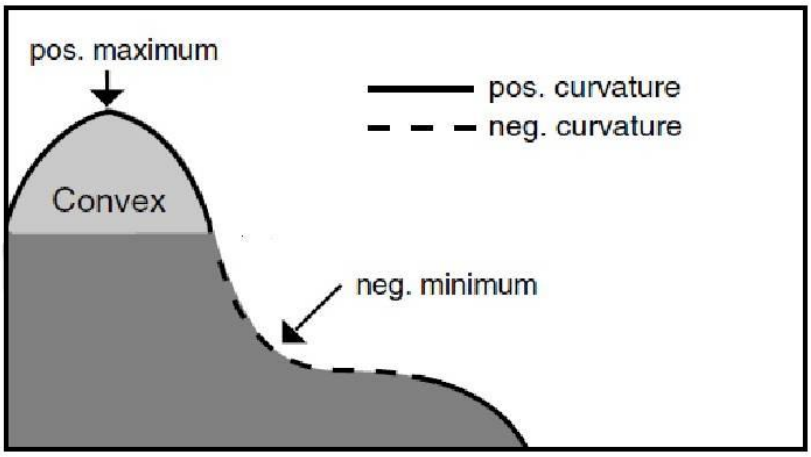


Figure 1 An example of positive (concave) and negative (convex) curvature extreme. From Cohen and Singh (2007).

However, the functional significance of these curvature extrema to shape perception remains the subject of on-going debate (Bertamini, 2008). Despite a large body of empirical research examining curvature in the context of contour-defined 2D images such as polygons and line drawings (Cohen, Barenholtz, Singh & Feldman, 2005; Cohen & Singh, 2007; De Winter & Wagemans, 2006) there is relatively little data examining the role of curvature singularities defined by changes in surface (rather than contour) curvature polarity. Moreover, such an investigation is further motivated by recent evidence that saccadic eye movements during the spontaneous exploration of visual images follow surface depth gradients (Wexler & Ouarti, 2008).

A recent study (Lloyd-Jones, Gehrke, & Lauder, 2010) used eye movements to examine the contributions of outline contour and salient individual futures in animal-non-animal recognition tasks, employing line drawings and silhouette stimuli. The authors proposed that animal recognition is mediated by the salient features located in the outline contour as the latter facilitates the access to stored visual representations. Despite the fact that the significance of local shape curvature in shape perception is subject to a lot of attention and under multidisciplinary research investigation (e.g. Renninger, Coughlan & Verghese, 2007; Lloyd-Jones, Gehrke, & Lauder, 2010) it is rather surprising that no one (to our knowledge) has yet attempted to investigate in detail 3D shape representation with eye-tracking methodology.

1.2. Aim of the thesis

The aim of this thesis is to examine the extent to which local shape analyses, as shown by fixation patterns, can be predicted by the presence of different types of local shape curvature in 3D object recognition. It should be noted that although eye movement pattern analyses are employed to investigate object shape representation I do not imply that these patterns emulate everyday object recognition. The rationale behind this thesis investigation stems from recent studies (e.g. Najemnik & Geisler, 2005; Renninger, Coughlan & Verghese, 2007) employing eye movement patterns to investigate 2D pattern recognition, showing that eye movement patterns can be highly informative in shape processing during perception. For example, Najimek and Geisler (2005) demonstrated that fixations during simple pattern visual search tasks are selected in order to reduce uncertainty, rather than choosing a possible target location. Similarly, Renninger et al., (2007) provided evidence that observers tend to fixate object areas with high information content, and proposed that reducing local

uncertainty can be a very good predictor of human fixations during shape learning tasks.

In regard to the organisation of this thesis, chapters are written as standalone sections and are parts of published papers (Leek, Cristino, Conlan (now Davitt), Patterson, Rodrigues & Johnson, 2012; Cristino, Conlan (now Davitt), Leek, 2012); hence some repetition of material is possible.

Chapter 1 provides a brief overview of the neuropsychology of the human visual system and a general overview of dominant theories of object recognition alongside approaches using different kinds of shape information (local vs. global) across a broad range of spatial scales obtainable from the sensory output. Chapter 2 comprises a brief overview of the current and historical eye tracking methods used to investigate the human visual system, and an overview of studies employing eye tracking methods. In Chapter 3, I present Fixation Region Overlap Analyses (FROA) - the methodology used to analyse the data in this thesis.

In Chapter 4, I examine fixation pattern analyses to investigate whether fixation patterns can be used to elucidate local shape analyses during object perception. Chapter 5 comprises of Experiment 2 that investigates the robustness of the results from the first study. Chapter 6 includes a categorisation experiment where I investigate fixation patterns during basic and subordinate levels of object categorisation in participants trained to be experts, or untrained novices in recognising novel objects at either basic or subordinate levels of categorisation. In Chapter 7, I present a categorisation experiment with an extended training regime and investigate further whether fixation patterns change as a function of training. In Chapter 8, I present a study examining how eye movement patterns relate to the perception of

shape during object recognition and motor-imagery. Finally, in Chapter 9 I discuss all the results in a more general context.

1.3. Background

Overview of the visual system

The eyes were referred to as a 'windowes of the mind' by the French anatomist and medical scientist Du Laurens (1596, p.19). This view provides a simple and true description of a very sophisticated and extremely complicated mechanism which enables us to see and interpret the world around us. Perhaps the true significance of the visual system can be demonstrated by the fact that about half of the cerebral cortex is involved in analysing the visual input of the world. Moreover, no computer visual system has yet been able to reproduce precisely the processes in the human visual system (Bear, Connors & Paradiso, 2007).

Pre-cortical processing

The fundamental requirement that allows us to see the world is the electromagnetic radiation around us, also known as *light*. The characteristics of this electromagnetic radiation (also described as a 'wave of energy', Bear et al., 2007, p.279) are: wavelength (the distance between peaks), frequency (the number of waves per second), and amplitude (the difference between a trough and a peak). The human eye has a very specific design which enables us to detect, capture and analyse a small part of the electromagnetic wave spectrum 400-700 nanometers. Before entering the eye, visible light travels in a straight line until it interacts with obstacles in the environment, such as objects, water, or atoms and molecules in the atmosphere. Consequently the light refracted by the cornea and the lens projects a picture from the environment on the retina. Following light absorption, the photoreceptors exhibit a complex chemical mechanism thus converting the retinal image into neural signals.

There are two types of photoreceptors cells, rods and cones. Rods are more sensitive to low intensity light and can function under dim light conditions; therefore, rods cannot distinguish colour, but enable us to see at night. Conversely, cones are responsible for colour vision and to function correctly require a high level of light. Cones are highly concentrated at the fovea, which contributes to high visual acuity. For an illustration of foveal vision see the distribution of rods and cones in Figure 2 below. The neural signals from the photoreceptors are transmitted to the ganglion cells which then, via the optic tract, convey the signals to the lateral geniculate nucleus (LGN) of the thalamus (the primary circuit for transmitting visual information).

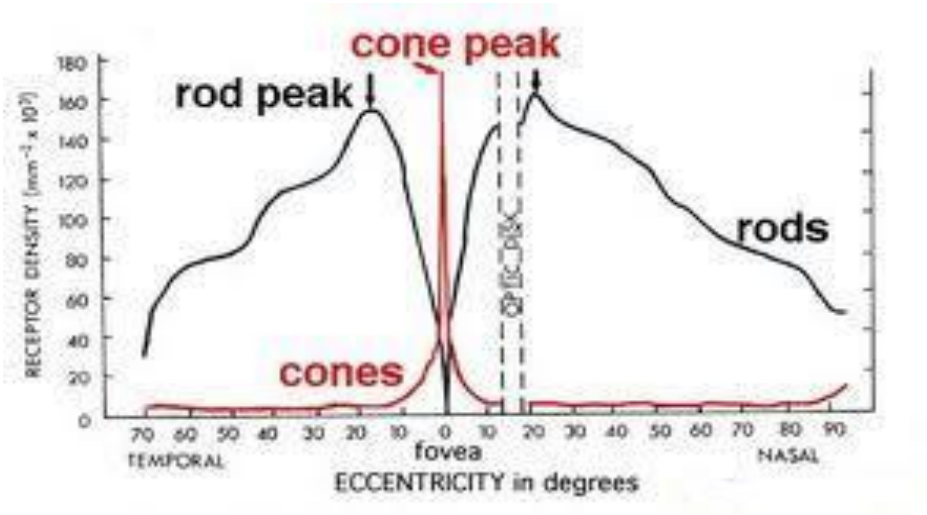


Figure 2 An illustration of the distribution of rods and cones in the retina. From Osterberg (1935).

Cortical processing

The LGN transmits sensory information in the form of neural impulses from the retina to the primary visual cortex V1 which is positioned in the occipital lobe, around the calcarine fissure and responds mainly to simple features and differently oriented lines (Hubel & Wiesel, 1959). V1 in each hemisphere transmits visual information through two distinct cortical systems, the ventral and the dorsal pathway.

The ventral pathway begins with V1, passes through V2, V4 (areas considered to be involved in preferential responding to more complex stimuli, Pasupathy & Connor 1999) and then to inferior temporal cortex (IT). The dorsal pathway begins with V1, passes through V2, followed by dorsomedial area MT (V5) and then to Posterior parietal cortex. The ventral pathway is considered to be involved in object recognition whereas the dorsal pathway is considered to be essential in spatial vision (Haxby et al., 1991) and involved in guiding the visual system for reaching and grasping tasks.

A large proportion of the human cortex beyond V1 is involved in the visual processing of variety of features and stimuli, thus distinct processing areas have been identified (Wandell, Dumoulin & Brewer, 2007). For example, inferotemporal cortex (IT) and the lateral occipital complex (LOC) areas associated predominantly with object recognition are suggested to play a role in encoding complex shapes (Tanaka et al., 1991), object parts (Hayworth & Biederman, 2006), non-accidental object features (Kayaert et al., 2003), and shape curvature (Kayaert et al., 2005). Furthermore, object selective areas in the ventral stream (LOC) appear to be less sensitive to low-level image features such as contrast (Avidan et al., 2002) and recent studies suggest that these areas are involved in responding more to object shape than object contours (e.g., Andrews et al., 2002; Kourtzi & Kanwisher, 2000). Another current study (Haushofer et al., 2008) has demonstrated that LOC has a preferential encoding of positive shape curvature (convex), suggesting that curvature may play an essential role in object shape representation[1]. A single cell recording study (Pasupathy & Connor, 2001) found that a subpopulation of neurons in the area of V4 is sensitive to boundary curvature (concave and convex). The authors proposed that shape representation in area V4 is distributed and a number of single cells encode small parts of larger objects.

[1] Representation is a way of encoding visual properties from the outside world which are organised in a useful and structured way to assist recognition processes.

This finding was interpreted as an indication for part-based shape processing in V4, which is consistent with theoretical accounts for object shape processing, supporting the notion of shape representation as a combination of smaller features and primitives in a hierarchical manner (e.g. Biederman, 1987; Marr & Nishihara, 1978; Hoffman & Richards, 1984). However, despite recent evidence concerning functional specialisation in the cortex, the exact content, organisation and structure still remain unknown. Thus, the complexity and the abstractness of the visual representation of shape beyond pre-cortical processing should not be underestimated.

In the next section I will give a brief overview of several classes of object recognition model which are the structural description models (Biederman, 1987; Marr & Nishihara, 1978; Leek, Reppa & Arguin, 2005), Image- based models (Bülthoff et al., 1995; Bülthoff & Edelman, 1992; Edelman, 1995; Edelman & Weinshall, 1991), Feature-based models as a hierarchy of fragments (Lowe, 2004; Mikolajczyk & Schmid, 2005; Ullman, 2007; Ullman et al., 2002), and Hybrid models (e.g. Hummel & Stankiewitz, 1996, 1998; Thoma, Davidoff & Hummel, 2007; Thoma, Hummel & Davidoff, 2004).

1.4. Overview of object recognition models

1.4.1. Structural description models

Structural description models (SDM) share the view that objects are represented by decomposing them into simple units, but make dissimilar claims about the exact arrangement and components used. For example, one SDM proposes that objects are represented as an arrangement of elementary viewpoint invariant 3D parts (Biederman, 1987; Biederman & Cooper, 1991; Hummel & Biederman, 1992), which are cylinders, bricks, wedges, or cones, with specified interrelations and spatial

configurations. Another structural description model (Leek et al. 2005) proposed that surfaces (bounded 2D polygons) and their spatial configuration mediate 3D object shape representation. The final structural description approach (Marr & Nishihara, 1978) suggests that object parts are mentally represented as 3D generalised cylinders and that objects can be hierarchically reconstructed of local image features into more complex descriptions relating the spatial position from one part to another. The fundamental difference between the claims of these accounts is the type of components used to build the volumetric object parts. For example, Biederman (1987) proposed sets of non- accidental invariant contour features (NAPs, collinearity, symmetry, parallelism, curvature, and co-termination) as the necessary building blocks of volumetric parts. Marr and Nishihara (1978) added the implication of depth, as they suggested that lines are grouped into contours, which are grouped into surfaces and the surfaces are grouped into volumetric parts. Finally, Leek et al., (2005) proposed that the units mediating 3D shape representation consist of object surfaces and their spatial configuration.

1.4.1.1. Recognition by components

One influential SDM proposed by Biederman (Biederman, 1985; Biederman & Cooper, 1991, 1992) is the 'recognition by components' model (RBC). RBC claims that humans have mental representations of a restricted set of 36 3D volumetric parts (e.g., blocks and cylinders) called geons with specified interrelations and spatial configurations which are viewpoint invariant under certain conditions (see Figure 3).

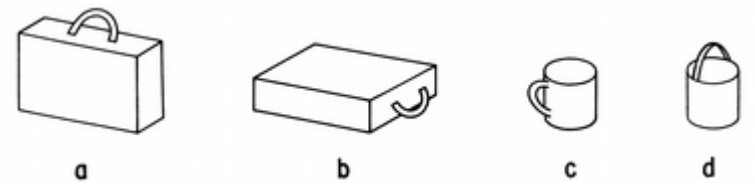


Figure 3 Different arrangements of the same components can produce different objects. From Biederman (1987).

This viewpoint invariance is suggested to be reliant on the stable local contour configurations (edges or junctions) which remain visible despite different rotations in depth. RBC proposed that these sets of non-accidental invariant contour features (NAPs) (collinearity, symmetry, parallelism, curvature, and co-termination; see Figure 4) are necessary to distinguish the existing geons (the building blocks for object representation) in an object, therefore the representations distinguished by the geons possess the same invariance (up to occlusion).

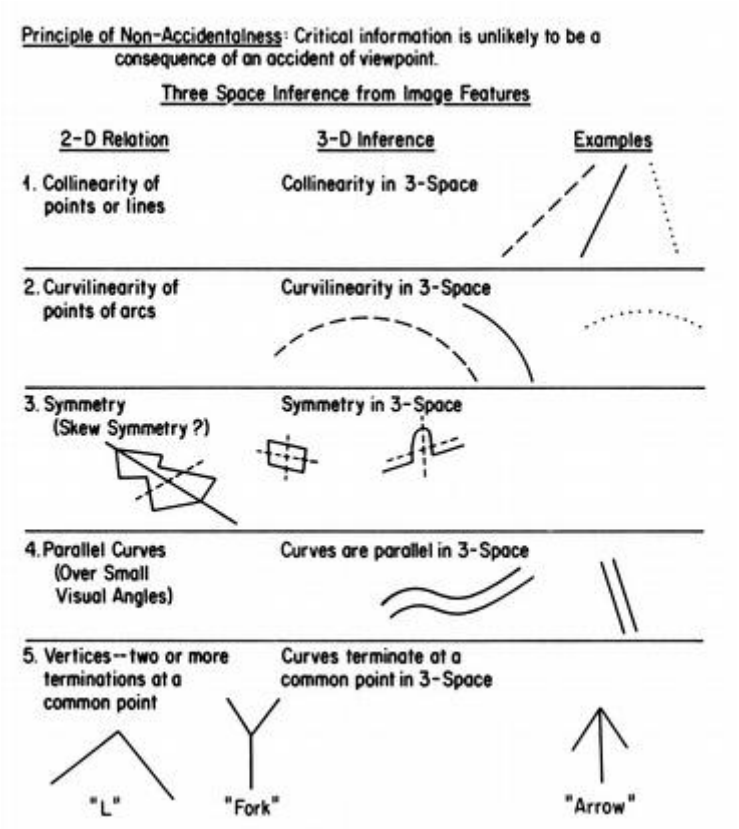


Figure 4 The non-accidental properties of image features. From Biederman (1987).

This theory suggests that as long as two or more geons are extracted from an object, the representations of this object will be nearly always be successful regardless of the viewpoint or rotation in depth (but not in picture plan rotations).

More specifically, RBC proposed that contour description of a given object is achieved after the edge extraction stage, and object characteristics such as luminance, texture, colour, and stereo information do not play any role in the task (see Figure 5). This is followed by detecting the image edges (NAPs collinearity, symmetry) which is assumed to be executed alongside with image segmentation, predominantly at regions containing deep concavities. The next stage is matching the primitive components against representations in memory which is supposed to occur in parallel and is assumed to have unlimited capacity (Biederman, 1987). The final stage involves object identification.

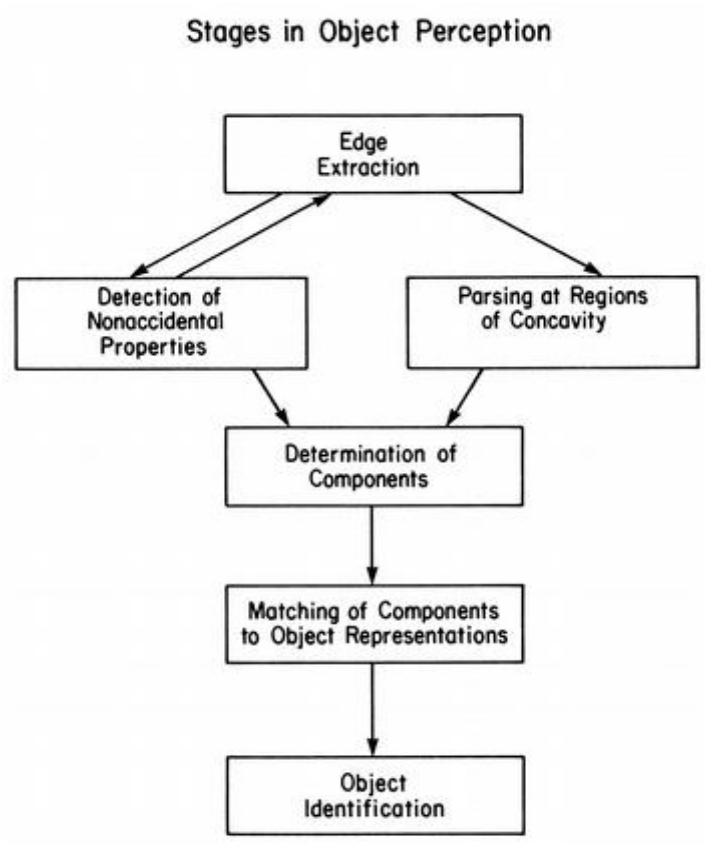


Figure 5 Biederman (1987) suggested stages of object processing. From Biederman (1987).

There are number of studies providing experimental support for RBC hypotheses although they have not been left unchallenged. For example, Biederman, Ju, and Clapper (1985) tested the prediction that two or three geons would be sufficient for quick identification of an object. In this experiment the participants had to name briefly presented complete objects, and objects lacking some of their components, with the restriction of presenting a minimum of two components per object. The authors found that participant's accuracy and RT performance increased as a function of the numbers of components present in the objects, which supported their initial hypothesis.

In several other experiments Biederman and Blickle, (1985, as cited in Biederman, 1987) provided evidence for the prediction that some of the contours present in an image play a vital role for identifying an object. They hypothesised that if object contours were deleted at vertices of edges (termed non-recoverable degradation) this will disrupt the recovery of object components and result in difficulty recognising it, compared to when the same amount of contour is deleted from a midsection of a curve (termed recoverable degradation) (see Figure 6 below). The objects were also modified in vertex areas along with altering the clarity of symmetry and parallelism. The results confirmed that non-recoverable versions of the objects were practically non-identifiable and the mean error rate was very high. The non-recoverable objects were identified only in instances when some of the components were not deleted and the object exposure was 200 ms in duration. Conversely, the recoverable objects were named with a high degree of accuracy at the 200 ms exposure duration.

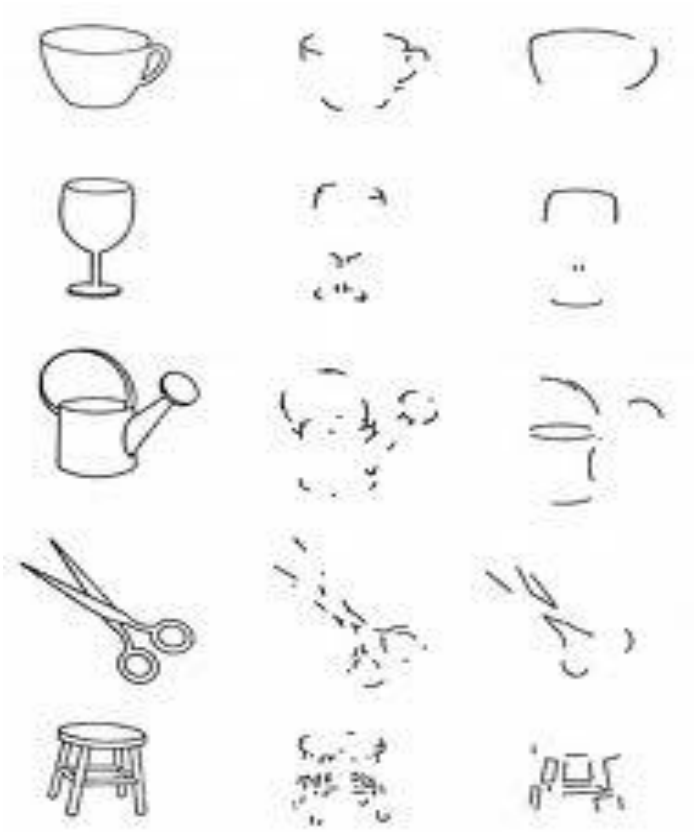


Figure 6 Example of the objects used in Biederman (1987) experiment. Left column depicts the intact objects. Middle column shows the recoverable version of the objects. The right column shows the non-recoverable version of the objects. From Biederman (1987).

In another study by Biederman (1987) the participants were presented with 18 line drawings of objects with varying amounts (25%, 45%, and 65%) of internal and external contours deleted for 100, 200, or 750ms. The deleted contours were either at the vertices or at mid-segments, but without bridging the components of collinearity or curvature which was the case in the non-recoverable condition in the previous experiment. The results showed that when the object exposure duration was 100 ms coupled with 65% contour deletion, the removal of vertices resulted in higher error rates compared to contour removal of the mid segment. When contour deletion was less and/or the exposure was longer, the decrease in naming accuracy was reduced. Overall, the authors concluded that filling-in of contours at mid segment and vertex

can be completed within one second. However, when there is a misleading component breaking the concavity, the image produced cannot index the components of the original object, regardless of the length of time given to view the image.

According to Biederman, the final stage of the aforementioned processes is object recognition, which is suggested to occur at the basic level category level (e.g., bird), but does not account for the process by which we may compute or determine either the superordinate (e.g., animal) or subordinate level (e.g. eagle) of recognition (Peissig & Tarr, 2007).

1.4.1.2. Surface -based structural description model

Although experimental evidence seems to support the role of volumetric parts in object shape representation, this theory is not indisputable. Leek et al., (2005) used a whole-part matching paradigm (Palmer, 1977) to test the 'special' status assigned to the volumetric components. The authors hypothesised that if volumetric parts play an imperative role in 3D shape representation, then whole-part matching should be better for part stimuli with volumetric components, rather than for part stimuli with non-volumetric configurations of contours on surfaces. In a number of experiments the authors examined performance when matching shapes of whole objects to subsets of shapes containing (a) volumetric components, (b) regions containing open or closed non-volumetric edge contour, and (c) regions of edge contour corresponding to object surfaces. The results showed an advantage for matching edge contour of volumetric components and surfaces over open or closed non-volumetric regions of edge contour. These findings were used to motivate a surface-based model of object representation where the generation of the surface boundaries was obtained from surface object discontinuities grouped in to 2D closed regions, thus not using volumetric primitives or volumetric image decomposition. In another study, Leek, Reppa, Rodrigues and

Arguin (2009) examined the assumption held by volumetric models that object recognition involves volumetric completion for the non-visible elements due to viewpoint or occlusion. The authors used a masked repetition priming paradigm to examine the extent to which occluded surfaces of volumetric parts, or visible surfaces yield greater priming effects. The rationale behind the design was that magnitude of the priming effects should be a function of the overlap between the shape information in the prime and the object information encoded in test performance. Participants initially memorized shapes of novel objects and later distinguished them from previously unseen objects. The two main prime conditions included (1) occluded volumetric primes vs. visible surface primes, and (2) occluded surface primes vs. visible surface primes. The results showed priming effects which did not support the volumetric theories, but are consistent with Leek et al., (2005) surface based model of 3D shape representation, as priming effects were greater for visible surface part primes, compared with occluded image properties.

1.4.1.3. Marr's stage model

Following Binford's (1971) pioneering work, Marr and Nishihara (1978) proposed a sequence of representations that could facilitate a progressive recovery of 3D geometric information from 2D images. Their framework comprised three representational stages: 'Primal sketch', '2½-D sketch', and '3D model representation'.

The proposed purpose of the primal sketch was to encode information about representations of significant gradients of light intensity from a relatively noisy image. This primal sketch itself consisted of three main stages (see Figure 7). The first stage was called 'the detection of zero-crossings' and was a process involved in the detection of significant intensity changes in a given image. The authors (Marr &

Hildreth, 1980) proposed that the best way for locating these features is by detecting areas of zero-crossings, created by the Laplacian of Gaussian second derivative gradient operator, which is considered to be an orientation independent operator. During the second stage termed as 'formation of the raw primal sketch' information such as edges, bars, blobs, alongside attributes of orientation, contrast, length, and position are obtained from the zero crossings. The last stage of the primal sketch is the grouping and formation of a higher level construct with clearer description of spatial organisation of the intensity changes, and is formed from the aforementioned primitive points obtained.

Marr proposed that the second representational stage is the formation of the '2½-D sketch'. This is described as a rather complicated matter as the information needed is collected from different processes such as stereopsis, optical flow, texture and shading, occluding contours, surface contours and motion parallax. Marr proposed that the 2 ½-D sketch intends to represent the orientation and the depth of the visible surfaces, and discontinuities from a specified in a viewer-centred coordinate system**.**

The third and final stage of processing required for object recognition is the conversion of the viewer-centred 2 ½-D sketch' to an object-centred '3D model representation' to enable the recognition from different viewpoints. Marr suggested that the necessary criteria for suitable 3D shape representations are accessibility, scope and uniqueness, and stability and sensitivity.

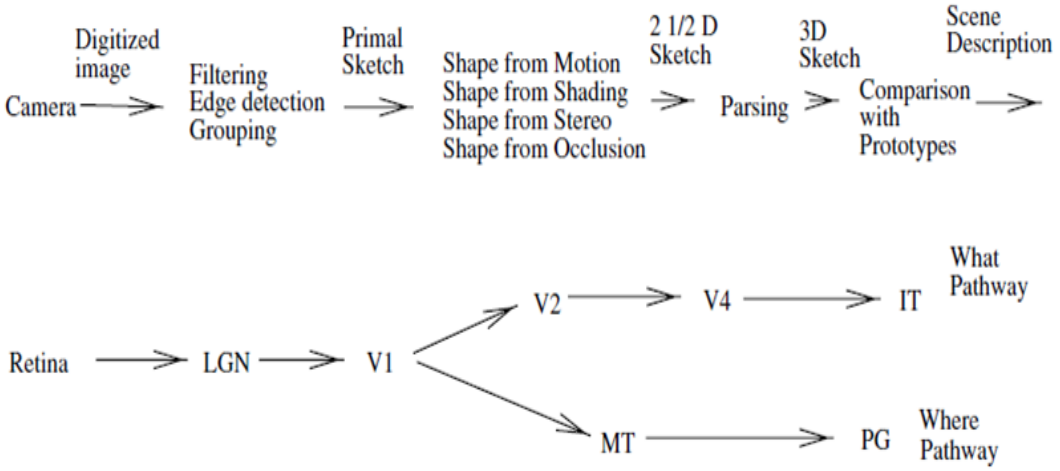


Figure 7 Marr's flow chart of visual processing, explained in hierarchical manner. Local edges and tokens are computed and grouped to infer surface orientations of objects. The 2.5D sketch is then parsed and matched against stored 3D prototypes. From Lee (2003).

The organisation of '3D model representation' (see Figure 8) is based on arranging object parts as generalised cones in a hierarchical manner in order to encode increasing complexity of object structure. Marr suggested that observers use the information from the object's bounding contour (concavities and convexities) to identify main component parts and locate the major axis of elongation.

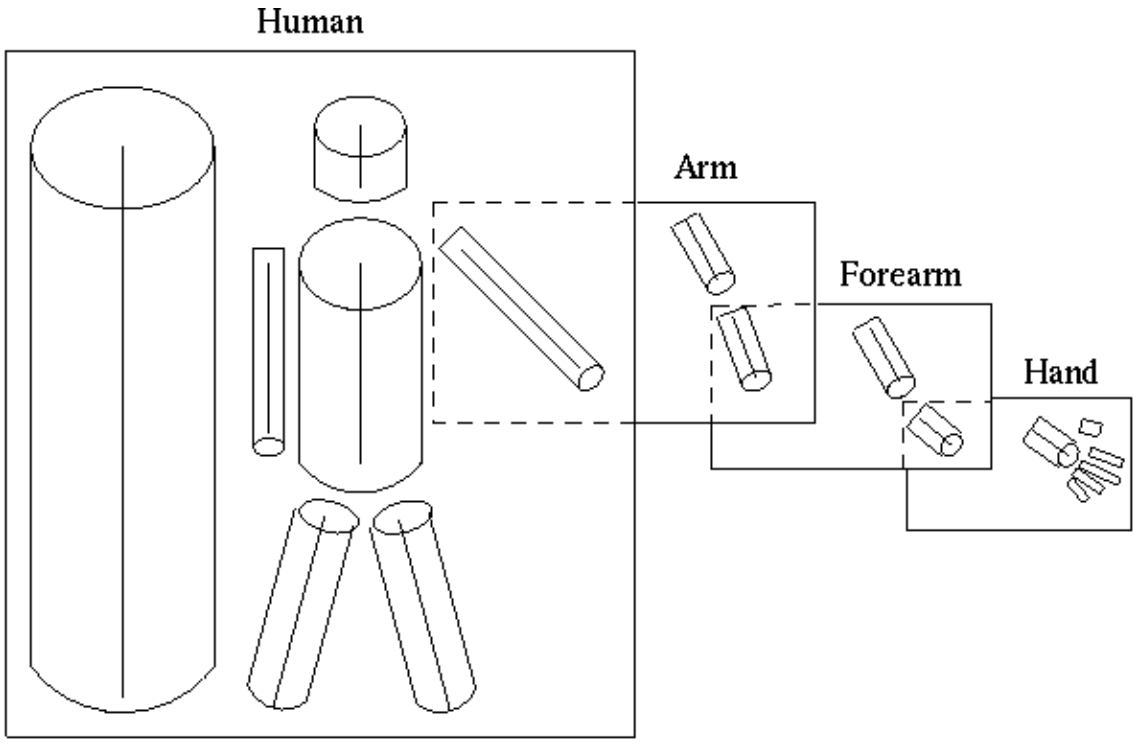


Figure 8 Marr's proposed hierarchical arrangement of the 3D model representation. From Marr (1982).

In summary, Marr's great contribution was in describing how different *levels* of computation/representation could help to 'break down' object recognition in to component processes. However, despite the huge theoretical influence of this model to the study of vision, there was only marginal empirical support in its favour.

1.4.1.4. Image based models

Another theoretical approach of shape representation comes from image-based models (Bülthoff et al., 1995; Bülthoff & Edelman, 1992; Edelman, 1995; Edelman & Weinshall, 1991; Poggio & Edelman, 1990). Some Image based accounts hypothesise that object representation is supported by multiple 2D views and transformed object images can be recognised by methods of interpolation or extrapolation between those known views. These accounts also support the notion of conjoint encoding of shape features and their spatial location (e.g., Edelman & Weinshall, 1991; Ullman & Basri,

1991). Other image based models propose that object classification is based on class-specific appearance or image-based feature hierarchies which are computed across multiple spatial scales (e.g., Ullman, 2006; Ullman & Bart, 2004; Ullman, Vidal-Naquet, & Sali, 2002).

More specifically, these models argue that for interpolation to occur, specific object views are expressed as sets of viewpoint dependent features, and each view is regarded as a specific point in a high dimensional space, thus capturing the appearance of all possible views (Tarr & Bülthoff, 1998). In order to generalise from unknown to known views it is necessary to establish the location of the unknown views within this high dimensional space and then calculate the similarity of its features relative to the features from the nearest known view.

Riesenhuber and Poggio (1999) proposed the HMAX model - a biologically inspired model of object shape representation, which intended to emulate the feedforward architecture of the stages of object recognition in the cortex. The model is based on the assumption that hierarchies occur naturally in the brain since specificity and invariance of position and scale need to be obtained in a biologically possible way. A key characteristic underlying this model concerns the fundamental organisation of the visual cortex, with parallel and gradual increase of feature complexity and receptive field size, thus initially requiring many cells to cover the necessary range of scales and positions representing a small set of simple features. On the other hand, in higher layers neurons are tuned to a larger number of complex features and neurons show greater invariance, thus requiring fewer cells tuned to the same feature at different positions and scales.

The original HMAX model is composed of four layers (S1, C1, S2, and C2, see Figure 9), consisting of simple (S), and complex (C) units. The model proposes that

the first input layer (S1) includes filters tuned to different orientations and areas of the visual field, which are similar to the cells found in the V1 receptive area of the brain. The C1 units accumulate responses by 'MAX' pooling operations over the S1 filters which are tuned to the same orientation but at different scales and positions. The complexity of the S input and the scale invariance in the C units increases going up the hierarchical layers. The S2 units combine the 4 bar orientations from the C1 units into 2 by 2 arrangements in order to create 256 intermediate feature detectors. C2 units spatially pool together the maximum output from the S2 units along with providing spatial invariance. The alternating architecture of S and C, combining simpler low level features into more complex features, gives an increased feature detector specificity and enhanced invariance of the model (Riesenhuber & Poggio, 1999).

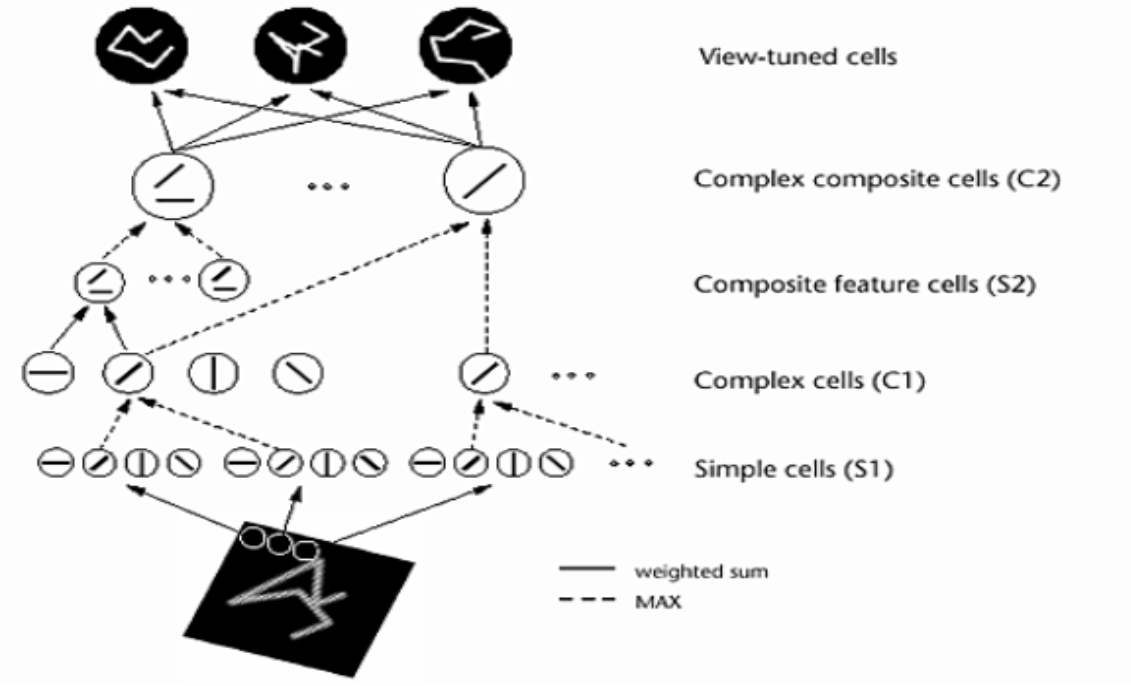


Figure 9 A diagram of the HMAX model. The first layer (S1) shows four different orientations (0^0, 45^0, 90^0, and 135^0) and consists of simple Gabor filters at several spatial scales. The second layer (C1) pools the filter outputs spatially and across nearby scales. The third layer (S2) is tuned to a combination of orientations, and the

fourth layer (C2) provides further spatial and scale invariance. The C2 outputs are directly fed to a classifier. From Riesenhuber and Poggio (1999).

Serre, Oliva and Poggio (2007) improved the original HMAX model by adding a new unsupervised stage of learning with the assumption that visual hierarchy in the cortex (from V1 to IT) builds a general dictionary of shape tuned units in order to provide task-specific representation for categorisation circuits. An example of a learning module can be seen in Figure 10a below. This module assumes that each unit measures the similarity between a given stored view and a given input image. The outputs of all units are then added and if their sum is above a threshold the output given is 1, if not it is 0. Throughout learning, weights and threshold adjustments optimise the correct classification of examples. As shown in Figure 10b the 3D model of the object is recognised by interpolation between small numbers of stored views (Riesenhuber & Poggio, 2000)

There are a number of different and more complex schemes developed in an attempt to solve the problem of object categorisation, and they tend to focus on categorising an image region for a specific viewpoint, followed by merging classifiers trained on different viewpoints. Riesenhuber and Poggio (2000) proposed that a key difference between these approaches is in the view-specific features with which the examples are presented, ranging from raw pixel values (Brunelli & Poggio, 1993) to over complete dictionaries of features allowing for more compact representations (Mohan, 1999).

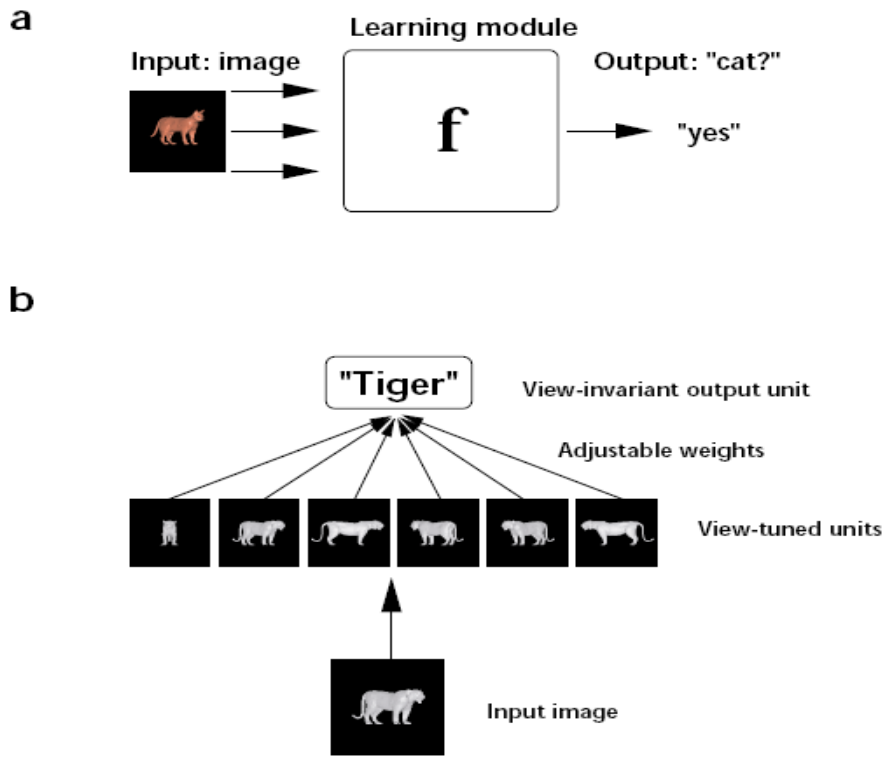


Figure 10 An example of learning module schematics. (a) A depiction of the general learning module. (b) A specific learning module: a classifier, trained to respond in a view-invariant manner to a given object. From Riesenhuber and Poggio (1999).

One of the issues concering image-based models is how an unfamiliar view of an object is matched to a familiar view of that same object and there were attempts to explain it using mental transformation and alignment methods. However, these processes need to establish the rotation direction before carrying out a rotation or an alignment and establishing this information implies that some form of recognition has already occurred (Tarr & Bultoff, 1999).

Moreover, image-based models appear to have difficulties similar to RBC with regard to object classification, as they suggest that the object representations are definitive to particular exemplars (given the assumption that objects are represented in a viewpoint-specific manner) not to object classes.

It should be noted that the two approaches (structural description and image based) described above are not mutually exclusive and both of them explain elements of human visual recognition; structural description models providing information about categorical (basic) level access, and image based for within-class level (subordinate) access. Moreover, recently some hybrid models propose that it is potentially achievable for both image-based and structural description approaches to be accommodated within the same framework (e.g., Foster & Gilson, 2002; Hummel & Stankiewitz, 1996; Hayward, 2003).

However, a great deal of the empirical debate between image based and structural description models is concerned with viewpoint dependency of object representations while the nature of 3D shape representation remains poorly understood (e.g., Arguin & Leek, 2003; Biederman & Gerhardstein, 1993; Edelman, 1999; Leek, 1998a, 1998b; Tarr & Bulthoff, 1998).

1.4.2. Local image features and shape representations

Another relevant source of information relating to fundamental elements mediating task performance during 3D shape representation are low-level image features (e.g., Non Accidental Image Features [NAPs], Biederman, 1985, 1987; Scale Invariant Feature Transform [SIFT], Lowe, 1999, 2004; corners, Harris & Stephens, 1988; codons, Hoffman & Richards, 1984), surface (e.g., Leek et al., 2005; Marr, 1982) and 3D primitives (e.g., geons, Biederman, 1987; generalised cylinders, Marr & Nishihara, 1978).

A number of shape representation theories suggest that local image features play a key role in object identification, mostly due to their advantages over global features, by providing invariance regarding noise, occlusion, and scale. Moreover, there is a lot of empirical evidence suggesting an important role of contour curvature

in the perception of part structure (e.g., Hoffman & Richards, 1984; Hoffman & Singh, 1997; Koenderink, 1984; Singh & Hoffman, 2001). More specifically, a psychologically relevant distinction has been found between negative and positive sign curvature (most commonly known as concave and convex curvature respectively), which suggests that concave minima curvature plays a vital role in part segmentation (e.g., Barenholtz et al., 2003; Hoffman & Richards, 1984; Marr, 1982). For illustration of concave and convex curvature see Figure 11 below. Although this seems to be a rather intuitive suggestion based on the fact that concave regions naturally appear when two parts of objects intersect (e.g., chair legs intersecting with chair seat), we must note that there is substantial amount of empirical evidence supporting this view.

Attneave (1954) was one of the first to propose that information along an objects' visual contour is not distributed evenly but is localised into regions of high magnitude of curvature. He also suggested that identifying the distribution of information along contours would provide key knowledge in understanding the properties of human mental shape representation. However, Attneave did not account for the fact that curvature (along with other shape descriptors) is scale dependent. Moreover, his main focus was on the curvatures' magnitude rather than the sign (positive or negative; Feldman & Singh, 2005). Feldman and Singh's (2005) theoretical account further specified that the total edge curvature must be larger than zero considering the object edges (principally convex) typically consist of closed curves (see Figure 11). Consequently, for any given contour region, the prior expectation would be that it is (slightly) convex, thus assigning more "surprising" properties to a given concave region, than to a convex one, suggesting that in turn means that its information content should be larger.

Hoffman and Richards (1984) proposed the *minima rule*, according to which, the human visual system facilitates visual recognition by segmenting shapes into parts at areas containing negative minima curvature (concave). Indeed, several experimental studies have demonstrated that the *minima rule* can be used to explain visual phenomena within shape perception, such as figure and ground assignment (Hoffman & Singh, 1997), perception of shape similarity (Hoffman, 1983), and visual search (Wolfe & Bennett, 1997). For example, Wolfe and Bennett (1997) found that a target with a shape containing sharp concavity 'pops out' amongst sets of distracter shapes that do not contain concavities, but not vice versa. The authors interpreted this finding as indication that negative minima curvature is computed early in the visual system. This finding was further investigated by Xu and Singh (2001), who confirmed that negative minima of curvature plays a role in segmenting object shapes into parts and that this happens in the early stages of the visual processing.

However, Singh and Hoffman (1999) argued that part boundaries defined by *the minima rule* alone are not sufficient to identify parts. They suggested that the *minima rule* needs to be coupled with other geometric factors, such as cut length, (Singh, Seyranian, & Hoffman, 1999), good continuation, and local symmetry (Singh, Seyranian, & Hoffman, 1996), and part-boundary strength (Hoffman & Singh, 1997), in order to establish correctly the part cuts.

Briefly, the above geometrical factors were named, *the short-cut rule*, and the significance of this approach is that it considers the distances between all pairs of points of a given silhouette outline. The authors suggested that the visual system prefers to divide shapes into parts by using the shortest part cuts and the minima curvature is the ideal cue for these cut locations.

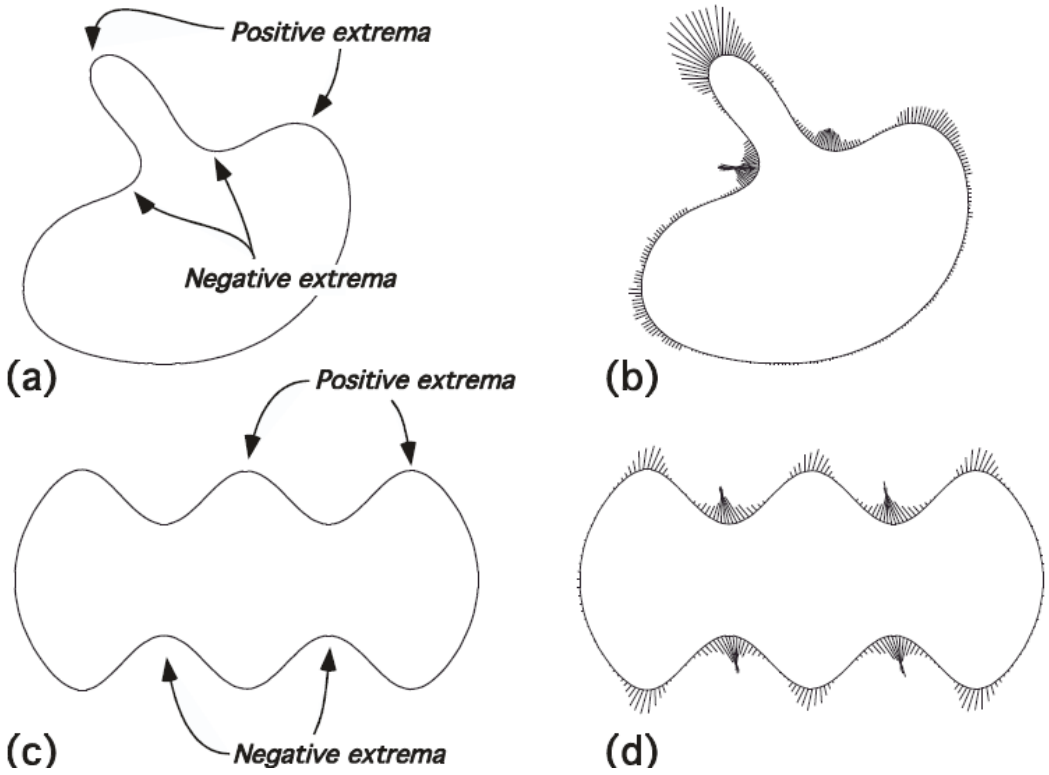


Figure 11 Depiction of the concentrated shape boundary information in regions of positive and negative curvature. The same shape used in Attneave (1954) in (b and d). Taken from Feldman and Singh (2005).

Although Singh and Hoffman's (1999) proposal has played an influential role in the vision research literature, there are opposing views that other curvature properties such as convex areas are more important for shape parsing. For example, Rosin (2000) argued that the short-cut rule incorporated only limited global shape information as it does not cross the local symmetry axis, thus potentially producing very different axes from the same shape, and consequently affecting the cuts. Rosin (2000) also suggested that using the length of cut as a sole determinant of salience is another weak point for the short cut rule, as it does not incorporate all the essential factors involved in shape perception such as the principles of symmetry and good continuation.

One often used method for investigating low level image features is the change detection paradigm. The rationale behind the method is based on the measure of change of differential sensitivity between two displays, which is considered to be an indication of differential representation. For example, it is assumed that if participants are frequently more sensitive to changes in one shape component of the visual display, but not to changes in another shape component within that display then it is this component that is more clearly represented in the visual system.

Barenholtz et al., (2003) used the change detection paradigm to investigate the representational differences between convex and concave curvature extrema. The authors used computer generated filled polygons consisting of a base shape and two modified versions of that shape (see Figure 12).

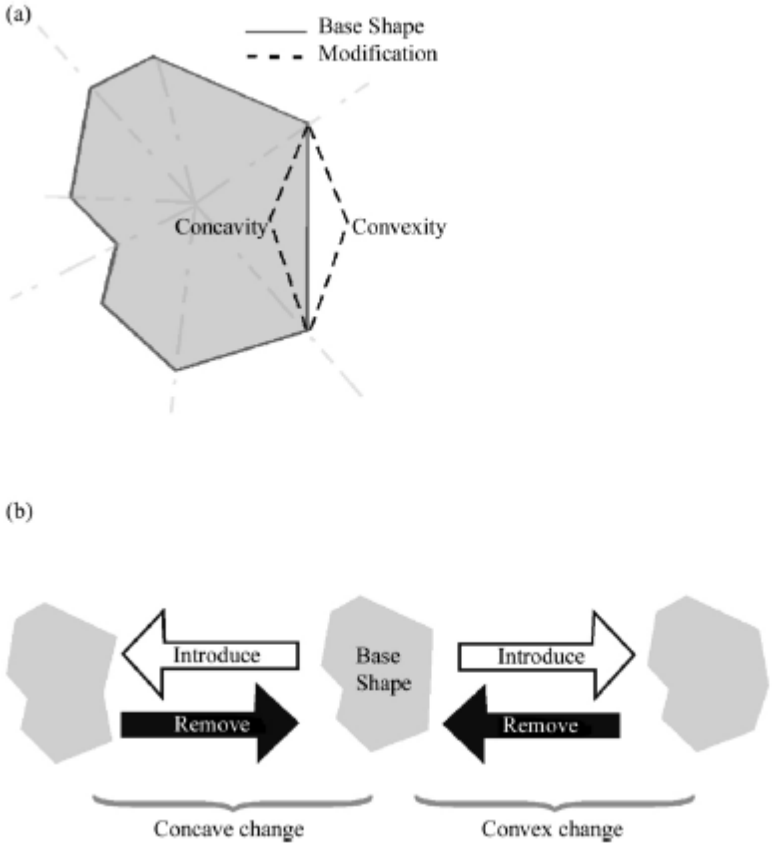


Figure 12 Example of the polygon pairs used in the experiment consisting of (a) Base shape and the modified versions of the Base shape either with added convexity or

concavity. (b) Example of the changes presented in the experiment with arrow directions showing the sequence of order. From Barenholtz et al., (2003).

The modified version of the shape was either with an additional concavity or convexity. The participants were presented with a two-way forced choice task: Change or No-Change. The Change trials included two independent variables, Change type (concave/convex) and Change direction (Introduce/Remove). In the No-Change trials, the base shape was displayed two times (see Figure 13). The results of this experiment showed a strong advantage for detecting concavities, regardless of change of direction (Introduced, or Removed) and the advantage was stable across all levels of magnitude change. The authors proposed that the apparent concave/convex asymmetry may account for 'local' vs. 'global' processing.

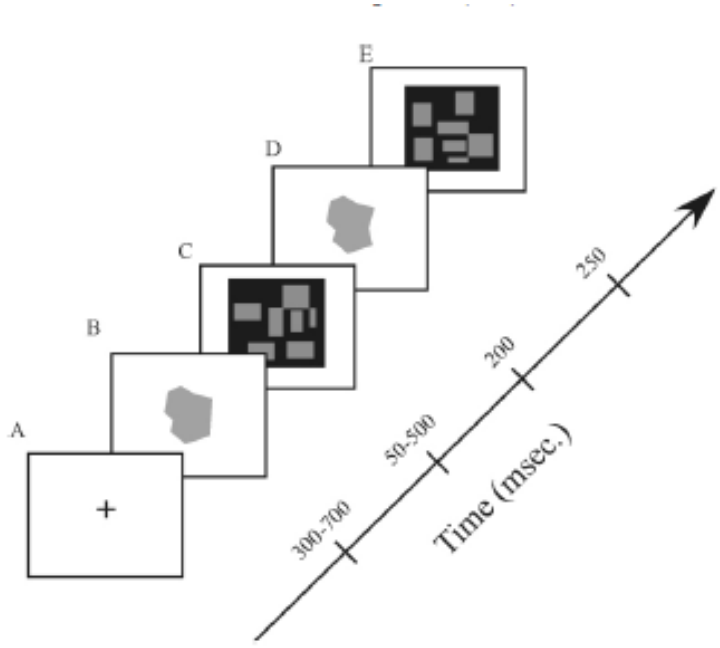


Figure 13 Example of the experimental sequence: (A) fixation point; (B) first shape stimulus; (C) mask; (D) second shape stimulus; (E) mask. From Barenholtz et al., (2003).

The localist hypothesis is that concave contours are intrinsically more salient and carry more information (Feldman & Singh, 2005); regardless of eventual role they may play in shape segmentation. The alternative explanation from the globalist

account (Palmer, 1977) is that concave regions are more detectable because they play a role in the global shape decomposition, in such a way that adding or removing a concavity could change the number of parts perceived in a given object. Conversely, adding or removing a convexity influences the shape of the existing part but does not change the number of parts perceived. It follows that observers may only be sensitive to global changes and not to changes in the local contour (Bertamini, 2001).

A rather different account (Barenholtz et. al., 2003) is that observers may instantly recognise the number of parts preattentively (van Oeffelen & Vos, 1982), hence they are more sensitive to changes in the number of parts. Barenholtz et al., (2003) concluded that despite the similar geometry, concave and convex contour regions have different perceptual representations, which in the current context indicate that concavities play an important role in visual shape representation. Furthermore, the authors pointed out that the methodology used could be a potentially powerful tool for separating representational differences and for testing a number of aspects in visual shape representation.

A different study by Bertamini and Farrant (2005) argued that the concave sensitivity found in a number of shape detection and visual search studies (e.g., Barenholtz, Chohen, Feldamn & Singh, 2003) is context dependent and thus based on how positive and negative curvature follow each other along the image contour with either type of curvature potentially showing greater sensitivity.

Another approach by Harris and Stephens (1988) suggested that information from local features, such as corners and edges, is necessary for connecting feature points which are essential for obtaining high-level descriptions such as surfaces or objects. However, their ‘corner detector’ was very sensitive to image scale changes so it did not provide a good framework for matching different size images.

Finally, a computational theory proffered by Lowe, (1999, 2004) investigated local image features in object recognition. Lowe proposed an algorithm for detection and extraction of local image features called the Scale Invariant Feature Transform (SIFT). This approach, which has been shown to be extremely useful in object recognition, transforms image data into scale invariant coordinates relative to local features and can be used to perform consistent matching between different views of an object or a scene. The SIFT approach also suggests that local image features are highly distinctive, relatively easy to extract, and help object identification with a low probability of mismatch.

1.4.3. Hybrid models

Recent hybrid models (Foster & Gilson, 2002; Hummel & Stankiewitz, 1996) provide evidence that image based and structural description theories can be accommodated within the same framework (Foster & Gilson, 2002; Hummel & Stankiewitz, 1996). One hypothesis (Hummel & Stankiewitz, 1996, 1998; Thoma, Davidoff & Hummel, 2007; Thoma, Hummel & Davidoff, 2004) proposes that object recognition is mediated by two types of representations, holistic (image-based) and analytic (structural descriptions). This model predicts that attended images are processed both analytically and holistically, whereas unattended images can only be processed holistically. The hybrid model proposed that attention plays an imperative role in the dynamic binding of information about object shape and spatial configuration during access to analytic structural descriptions (Hummel, 2001; Hummel & Biederman, 1992; Hummel & Stankiewicz, 1996, 1998). Conversely, during image-based holistic recognition binding is not important since representations do not separately encode feature dimensions such as shape and configuration.

One study supporting this proposal (Thoma et al., 2004) investigated the extent to which attended and unattended primes facilitate the subsequent identification of whole objects. In their study the participants were presented with prime displays consisting of either whole object or two split object parts. Participants named one of the two primes indicated via a cueing box. A second whole object or probe was presented at the end of the trial. This probe item could be related to the attended or unattended prime or unrelated to either. The results showed faster responses in probe identification for related trials in both attended conditions. In contrast, while unattended whole images elicited small but significant positive priming effects, unattended split part images showed no priming at all. It was suggested that the absence of priming for the unattended split part primes shows that attention is required to support analytic processing (alternatively, for conflicting findings see Conlan, Phillips & Leek, 2009).

Palermo and Rhodes (2002) proposed the opposing view that attention plays an important role in the generation of holistic representations in face discrimination. The authors demonstrated that participants performed better when recognising an isolated feature (e.g. nose, mouth) of a target face when that feature was presented in the context of a complete face rather than presented in isolation. This effect however, was only evident when the target face was attended. Hummel (2002) rationalised this finding as evidence that when image features are not bound to a specific context, such as location or relation, they have less structural information than holistic representations that have embedded implicit structural information. Hummel (2002) interpreted these findings as evidence that attentional demand is greater for analytic representations than holistic representations and holistic representations require greater attention than free floating features.

Another study supporting a hybrid form of representation (Foster & Gilson, 2002) proposed that visual object recognition benefits from two independent processes deriving both structural and image based information in an additive manner. In their study, participants were required to discriminate novel 3D objects (incorporating structural and image based information) in a simultaneous matching task. The objects differed from each other on the number of parts, part curvature, part length, and angle of join between parts. The authors suggested that the first property (number of parts) of the object provide information about object structure, whereas the rest of the parts' properties provide image based information which will change considerably across views. The results showed that detecting differences in objects parts was easier than detecting differences in their metric properties. Interestingly, these two types of discriminations showed the same view point dependence decline when the viewpoint orientation increased to 45°, and no interaction between the manipulation of structure and orientation. The authors interpreted these findings as an evidence for an additive relationship between parts based and image based processes in object recognition (Hayward, 2003).

Eye movements play a vital role in the human visual system allowing us to acquire visual sensory information. As such, the considerable interest from a substantial body of multidisciplinary research is not surprising. Due to the anatomical structure of the eyes the highest acuity of visual information acquired from the environment is restricted to a small region corresponding to the fovea, hence fixational eye movements are assumed to play an important role in extracting and the consequent processing of the perceptual input (e.g., Yarbus, 1967; Rayner, 2009; Henderson & Hollingworth, 1999, 2003).

The historical background of eye movement studies goes back to the 17th century when a detailed object analysis was reported by Porterfield (1737, 1738). Moreover, eye sampling rate is easily measurable and can potentially provide us with detailed information during object and scene representation and recognition. A large body of research provides evidence that measurement and analyses of eye movements gives us an opportunity to study the mechanisms of visual information processing in a variety of domains such as reading (e.g., Rayner, 1996), scene perception (e.g., Yarbus, 1967; Rayner, 2009; Henderson & Hollingworth, 1999, 2003), object shape representation (e.g., Renninger et al., 2005, 2007; Leek et al., 2012), movement (e.g., Li & Lisberger, 2011), and attention (e.g., Gilford, 1936; Hoffman, 1998). Several studies (Melcher & Kowler, 1999; Vishnwanath & Kowler, 2004; Wexler & Quarti, 2008) showed that eye movement patterns are influenced by cues of 3D structure, thus providing appealing background for employing eye fixation patterns to elucidate the representation of 3D shape.

Chapter 2

2.1. History and methods for investigating the human visual system (eye gaze tracking)

The preceding chapter provides an overview of the main object recognition theories and reasons for the decision of employing eye movement pattern analyses to investigate the mechanisms of object shape representation. In the current chapter I present a brief overview of the eye tracking history and methods used to investigate the human visual system.

Eye movements have been investigated from 1737 (Porterfield) in a variety of domains including reading, scene and face perception, object localisation and visual search (Land, Mennie & Rusted, 1999; Liversedge & Findlay, 2000; Henderson, Brockmole, Castelhano & Mack, 2007; Rayner, 1998; Renninger et al., 2007; Underwood, Foulsham, van Loon, Humphreys & Bloyce, 2006). In general, eye tracking methodologies have improved and changed considerably over the last centuries; from using a mirror and a telescope (e.g., Javal, 1879), through to more intrusive methods such as fitting a plaster cap over the cornea of the cocained eye and then connecting it mechanically to a kymograph in order to record the lateral and vertical movements of the eye (Delabarre, 1898). Later on, Dodge and Cline (1901) used more unobtrusive methods of light reflections from the eye, recording eye movements in the horizontal direction only. The first method, however, which provides the opportunity to process real time gaze data was designed by Jung (1939) and was called electrooculography (EOG). Jung applied electrodes on the skin close to the periphery of the eye which allowed him to measure the vertical and horizontal eye movements simultaneously.

In the past 70 years eye-tracking technology improved considerably, increasing accuracy and precision; moreover, psychological theories began linking the eye tracking data to specific cognitive processes (e.g., Monty & Senders, 1976; Senders, Fisher & Monty 1978), and changed the general attitude of looking at eye movements predominantly from a sensorimotor and physiology interest.

Stratton (1906) was first to highlight the importance of saccades during figure observation and he demonstrated that patterns of eye movements did not follow the shape of the observed figures. He interpreted this as evidence that the eye is searching for the best view of important features that need to be obtained. The lack of relationship between eye movement patterns, object shape and symmetry intrigued researchers for a long time. Thus, a great deal of research was devoted to studying the influence of image characteristics of eye movement patterns (Wade, 2009) and the assumed link between eye fixations and information acquisition.

The two commonly investigated components of voluntary eye movements are saccades and fixation locations. The main function of saccades is to bring a new area of the visual field into the fovea in order to gather high resolution information. Although information processing is assumed to be suppressed during saccadic eye movements itself (see Cambell & Wurtz, 1979, for report of context where some information can be acquired during a saccade), they reveal global aspects of visual perception such as scan patterns and areas of fixations while examining a variety of stimuli. During fixations however, the fovea (2 degrees of the centre of the visual field) is positioned at the part of the stimulus that needs to be seen clearly as this is the point where we extract maximally the visual information we need for further processing. The fovea region of the retina has the highest visual acuity, followed by

the parafoveal region (from 2 to 5 degrees) were acuity diminishes, followed by the peripheral region (beyond 5 degrees) were the acuity is poorer.

The length and duration of saccade and fixation patterns are found to differ as a function of the task in hand. For example, fixations during scene perception are found to be longer and saccade length to be larger than during reading (Rayner, 2009). Also, the scan path during scene analysis typically shows that not every part of the scene is covered, rather, a large number of the fixations tend to fall on the informative areas[2] of the scene. Empirical and computational research have suggested that some of the main factors influencing the eye movement patterns during scene perception are saliency and cognitive influences (e.g. conceptual knowledge) linked to the task in hand (e.g., Itti & Koch, 2000; Yarbus, 1967).

One interesting and robust phenomenon in eye movement research is the centre of gravity effect (COG), which refers to the tendency of initial eye fixations to fall at the centre of the target. This phenomenon seems to be present in scene and object perception, as well as in reading. For example, in reading research the first fixation to a given word falls at the optimal viewing position and this location tend to be at one character from the centre of the word (O'Regan & Levi-Schoen, 1987). Similarly, in object and scene perception viewers initially tend to fixate near the centre of the object, or the scene (Henderson, 1993).

A number of similarities can be found in eye movement behaviours that can be generalised across tasks such as reading, scene analyses, object recognition, and visual search. First, the difficulty of the stimulus seems to influence eye movement patterns. For example, in reading, eye fixations get longer and saccades shorter when the text becomes more difficult. Similarly, in scene analyses and visual search, fixations get

[2] A broad description of 'informative areas' involves two different kinds of information depending on the context and the task in hand. One includes local physical factors such as object curvature and discontinuities, and the other includes cognitive factors such as top down influences linked to the task in hand.

longer and saccades shorter when the observed stimulus is more complicated. The difficulty of the task is another variable that influences eye movement patterns (e.g. searching for an object in a scene vs. looking at a scene for a memory test: Rayner, 2009).

Predicting fixation patterns and saliency

Different theoretical and biological models have been proposed to account for eye fixations and information acquisition (e.g. Morrison, 1984; Henderson, 1993; Findlay & Walker, 1999). One influential computational model of saliency developed by Itti and Koch (2000, 2001; see also Itti, Koch & Niebur, 1998) was originally based on ideas suggested by Koch and Ullman (1985). The basic concept behind this model is that individual feature maps (e.g. based upon contrast, intensity and orientation) are combined into a salience map in a 'winner-take-it-all' fashion to direct visual attention, which is subsequently suppressed (inhibition of return, IOR) to allow attention processes to shift to the next salient point.

Itti and Koch (2000) justified their model from a neurobiological perspective by using centre surround computations to merge different spatial scales into a single feature map. This reflects the knowledge that visual receptive fields are organised in a centre surround fashion where preattentive processing for discontinuities begins. The saliency model also incorporates convolution of Gabor filtering resembling visual system functions such as the parallel fashion of early feature extraction which is assumed to be modulated locally, alongside more long range connections in primary visual cortex (V1). When these distant connections are extended outside the receptive field, this facilitates the response of orientation selective cells (Gilbert, Ito, Kapadia, & Westheimer, 2000) and stimulates the within feature competition which has been applied in the saliency model algorithm.

The saliency model has been empirically tested numerous times using stimuli with different complexities and the results provide evidence that this model can predict fixation locations better than chance (e.g. Itti & Koch, 2000; 2011; Foulsham & Underwood, 2008). Nevertheless, the predictive power of this model has been disputed and research suggests that only a relatively small number of human fixations can be explained solely by saliency models (see Parkhurst, Law, & Niebur, 2002; Tatler & Vincent, 2009; Betz, Kietzmann, Wilming & Konig, 2010; Schutz, Braun & Gegenfurtner, 2011). More recently researchers have questioned the changing magnitude of the observed correlations between salient features and fixation pattern (Tatler, Baddeley, & Gilchrist, 2005; Rothkopf, Ballard, & Hayhoe, 2007) in different tasks and concluded that these correlations do not automatically imply causation (Henderson, Brockmole, Castelhano, & Mack, 2007).

A growing body of research (Underwood, Foulsham & Humphrey, 2009; Matsukura, Brockmole & Henderson, 2009) has demonstrated that purely bottom up saliency models cannot adequately account for eye movement patterns during recognition tasks. Other studies investigating fixation patterns during complex behaviours found that observers tend to fixate predominantly on task relevant objects instead of visually salient objects (Land & Hayhoe, 2001; Land, Mennie, & Rusted, 1999). In general, the saliency model seems to account reasonably for circumstances when there is not a specific task in hand, but performs quite poorly during visual search in real world scenes (Henderson, Brockmole, Castelhano & Mack, 2007).

Predicting fixation patterns in 2D shape representation

Another approach to understanding fixation selection falls under the framework of the information theory. Renninger, Verghese, and Coughlan (2007)

examined fixation patterns during a 2D shape (see Figure 14) learning and matching task, and reported that observers fixated on the most informative object areas in order to reduce local uncertainty. In the learning phase of the study participants had to fixate on a cross presented either on the left or right side of the screen for 300ms, while a novel object shape was displayed at the periphery. Subsequently they were required to make an eye movement to the opposite side of the screen to the previously displayed shape belonging to a pair of shapes. Each shape was displayed for 1200ms and the participants had to memorise it. During the matching task, the learned shape was displayed at a new location alongside the highly similar partner shape and the participants had to answer (with no time constraints) which shape out of the two they have previously learned. Eye movements were recorded during both tasks.

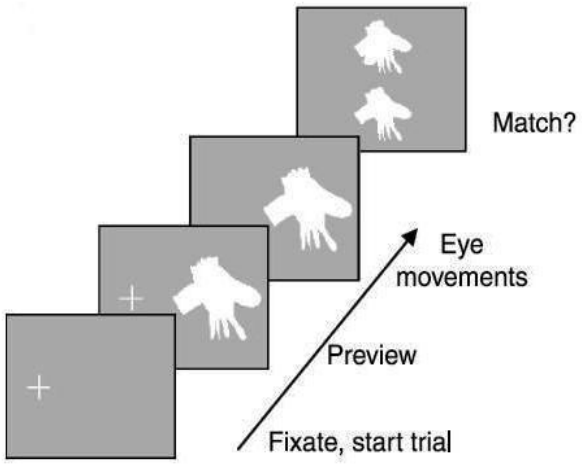


Figure 14 Depiction of the task. Participants fixate on the fixation cross for 300ms while the object shape appears on the opposite side of the screen. When the fixation cross is removed, the participants memorise the shape for 1200ms. This is followed by a matching phase where a pair of shapes is displayed until response. Taken from Renninger et al., (2007).

The authors' rationale was that the learning phase assists in building a correct representation of each shape in order for it to be distinguished from a very similar shape during the matching task. The authors' assumed that the information necessary for the task was edge orientations which derive from shape contour and constructed a probabilistic model of shape contour from the human visual data, demonstrating the steps needed for gathering the orientation information (see Figure 15).

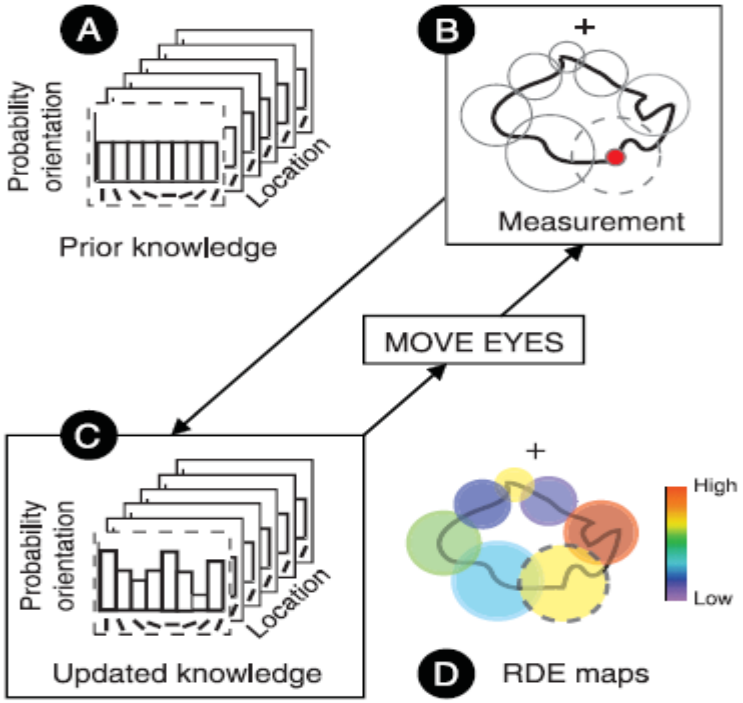


Figure 15 An illustration of the probabilistic model of shape contour information.
(A) Prior eye movements' exploration there is no knowledge or information about the orientation at every location in the stimulus space. Thus the probability distributions over orientation are flat at each location, and uncertainty is prominent everywhere.
(B) The sample fixation (+) position smaller pooling neighbourhoods near the top of the shape whereas larger neighbourhoods are places near the bottom. An appropriate pooling area (dashed circle) is used to compute the orientation distributions.
(C) In order for updated knowledge to be produced, the measurement distribution is multiplied with the prior distribution at that location and the updated knowledge turns into prior knowledge for the next fixation. Each time the eyes move, another measurement is taken.
(D) Schematic. The uncertainty (or information) at any point in space and time is computed from the updated knowledge and can be represented with a resolution dependent entropy (RDE) map. For example, during the first fixation, straight lines within a pooling neighbourhood result in lower entropy (blue) at a given location,

whereas curved or bumpy lines within a neighbourhood result in higher entropy (red). Taken from Renninger et al., (2007).

The model was compared to human observer data in order to investigate how information is used in the planning of eye movements to stimuli, and how other decision strategies for predicting fixation locations such as, maximising the amount of total information (global) and visual saliency may play a role. The findings were that the visual saliency model had a poor fit with the observers data. However, there was a good fit to the global strategy prediction. Renninger et al (2007) used global strategy that predicts the location of the next fixation based on updated knowledge and information gained from the previous fixation. The authors suggested that although human visual system may plan more than one fixation at any time, predicting the fixation sequence for more than one fixation is more computationally intense and have potential confounds. Their results showed that areas with local uncertainty dominated eye movement decisions, which the authors interpreted as evidence that observers tend to fixate at the most informative locations of outline contour, thus reducing local uncertainty in order choose where to look next.

Throughout the long history of studies of eye movements it is apparent that they are a fundamental feature in pattern analysis and provide a valuable method for investigating a variety of diverse factors mediating mental representations and cognitive processes.

Chapter 3

3.1. FROA Methodology

In the current chapter I outline step-by-step the methodology used to analyse the eye movement data in this thesis. There are relatively few ways of comparing the spatial distributions of fixation patterns to model data. Indeed, at the beginning of this thesis, there was no quantitative and statistically rigorous method for exploring image information content using empirically defined ROIs. A key advantage of FROA is that it provides a way to generate empirically defined ROIs and compare their spatial distributions to theoretical models. The FROA method does this using thresholding to generate ROIs and Monte Carlo's for determining statistical significance. Although there are other methods for comparing two data sets (e.g. Mannan, Ruddock & Wooding, 1995; Privitera & Stark, 2000; Fujita, Privitera & Stark, 2007), these methods do not take into account the distribution of fixation regions of interest (ROIs).

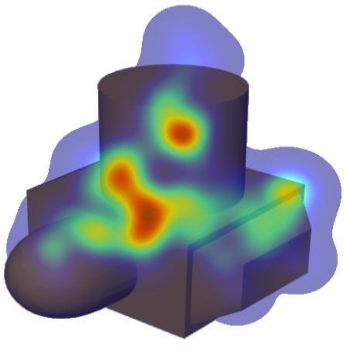
FROA was originally created by Johnston and Leek (2008), and later modified by Leek et al., (2012). This method has been applied in a number of experimental studies (e.g. Johnston & Leek, 2008, Leek et al., 2012) and has proven to be a reliable measure of fixation pattern performance.

FROA method defines fixations as eye movements that remain within the same circular region of diameter 60 pixels (2° visual angle given a viewing distance of 60 cm, a screen resolution of 1280 x 1024 pixels and a horizontal screen size of 34 cm) for at least 100 ms (e.g., Manor & Gordon, 2003). In addition, for each trial the first fixation following stimulus onset was discarded in order to eliminate early object localization fixations associated with COG effects which have been shown to be

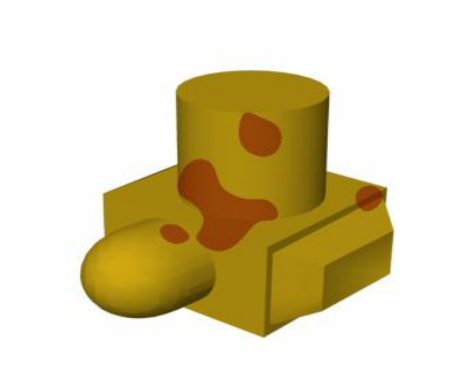
sensitive to global object shape (e.g., Denisova, Singh & Kowler, 2006; He & Kowler, 1989; Melcher & Kowler, 1999).

The empirical derivation of the area of interest (AOIs) or AOI region maps, and their quantitative comparison to the predicted distributions, was achieved using a modified version of FROA (see Johnston & Leek, 2009, for a full description and Matlab implementation of the FROA method). The AOI maps for each stimulus were created by summing the convolution of each fixation map (summed across subjects) with a 2D Gaussian kernel (SD = 4 deg) (see Figure 16). Since fixation frequency varies across subjects and conditions, the maps were normalised using z scores. The AOI region maps were derived by binary thesholding the fixation frequency distributions using a fixed parameter across all conditions. Here, the threshold was set to z = 1.2 in order to reduce the chances of obscuring potentially relevant (sub-threshold) peaks in the fixation frequency distributions.

a) Z scored Heatmap made with a Gaussian kernel of 4 degrees from the fixation frequencies overlaid on the object



(b) Thresholded map (z = 1.2) overlaid on the object



(c) Binary thresholded region map computed by FROA

Figure 16 Illustrative visualisation of the primary steps used to derive the binary region maps underlying FROA.

These binary AOI region maps formed the basis for the subsequent analysis of the pre-test and test phase fixation data. The primary dependent measure in FROA is spatial (i.e., area) overlap percentage (e.g., the amount of area overlap in the binary region maps for each stimulus and the predicted distribution of AOI regions for each theoretical model of shape information normalised by the size of the binary region maps for each stimulus; see Johnston & Leek, 2009). Overlap is determined by calculating the number of supra-threshold pixels that occur at the same spatial locations in the binary fixation region maps of each contrasted (observed versus modelled) image set. The statistical significance of the observed overlap percentage between data sets is then determined with reference to bootstrapped probability distributions derived from Monte-Carlo simulations. These are used to generate the expected random frequency distribution of area overlap percentage for a given observed, and modelled, fixation region. This technique provides a method for estimating the random distribution of overlap that would be expected for fixation

regions of the observed shape and size (area), and which is constrained to fall within locations bounded by the perimeter (occluding contour) of the original stimulus. It is important to note that this method thus controls for differences in the area of the respective region maps (and specific threshold parameters) in any set of contrasted images. Statistical analyses were conducted across objects (items) and across subjects. The statistical significance of the fit between the observed fixation data and each model prediction was calculated as follows:

Step 1: The '*Actual Overlap Percentage'* (AOP) between the binary images of the observed thresholded region maps and a given model is calculated for each stimulus (see Figure 17). This is computed as a percentage of the total region area in the observed thresholded region map (i.e. 0% if the model did not overlap at all with the observed fixation map or 100% if the model overlaps completely with the observed fixation map).

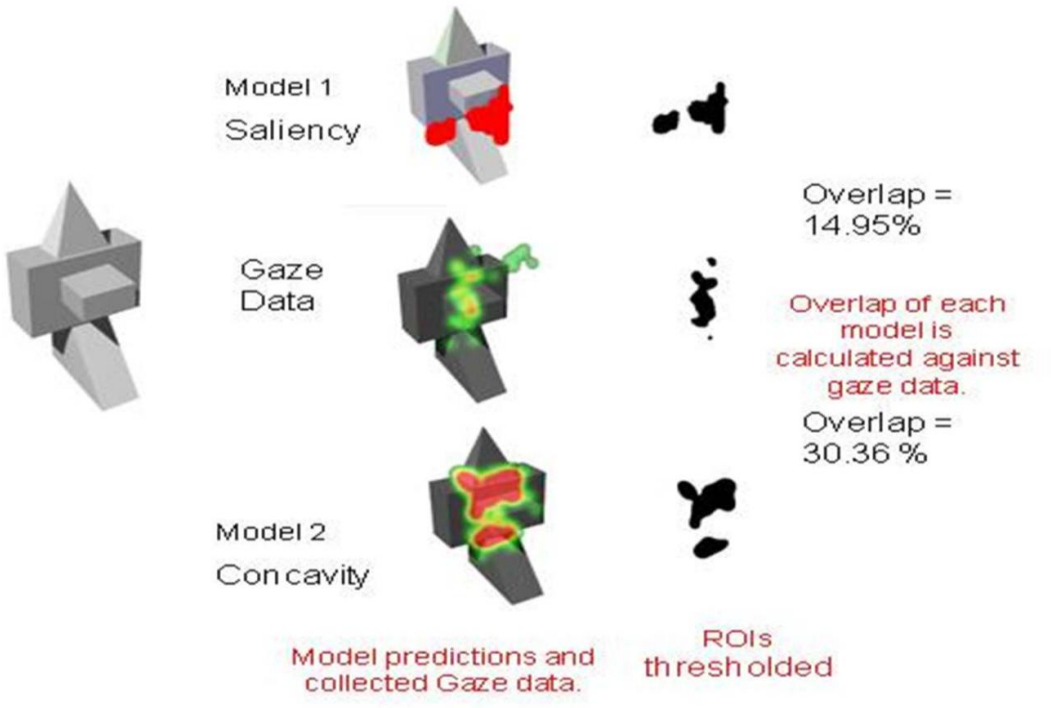


Figure 17'*Actual Overlap Percentage'* (AOP) between the binary images of the observed thresholded region maps and a given model**.**

Step 2: For each stimulus we calculate the '*Chance Overlap Percentage'* (COP) which corresponds to the percentage overlap we would expect at the 95% C.I. of the random distribution of observed fixation data-model overlap derived from a Monte Carlo procedure. In order to compute the random Monte Carlo distribution, the observed fixation data-model overlap (per item and model) is recomputed over 1000 iterations. During each iteration the thresholded region map is randomly relocated within the bounding contour of the stimulus, and data-model overlap recomputed. This allows us to compute a distribution of random region overlap for each stimulus, and a data-model contrast that takes into account region size and region threshold level (see Figure 18).

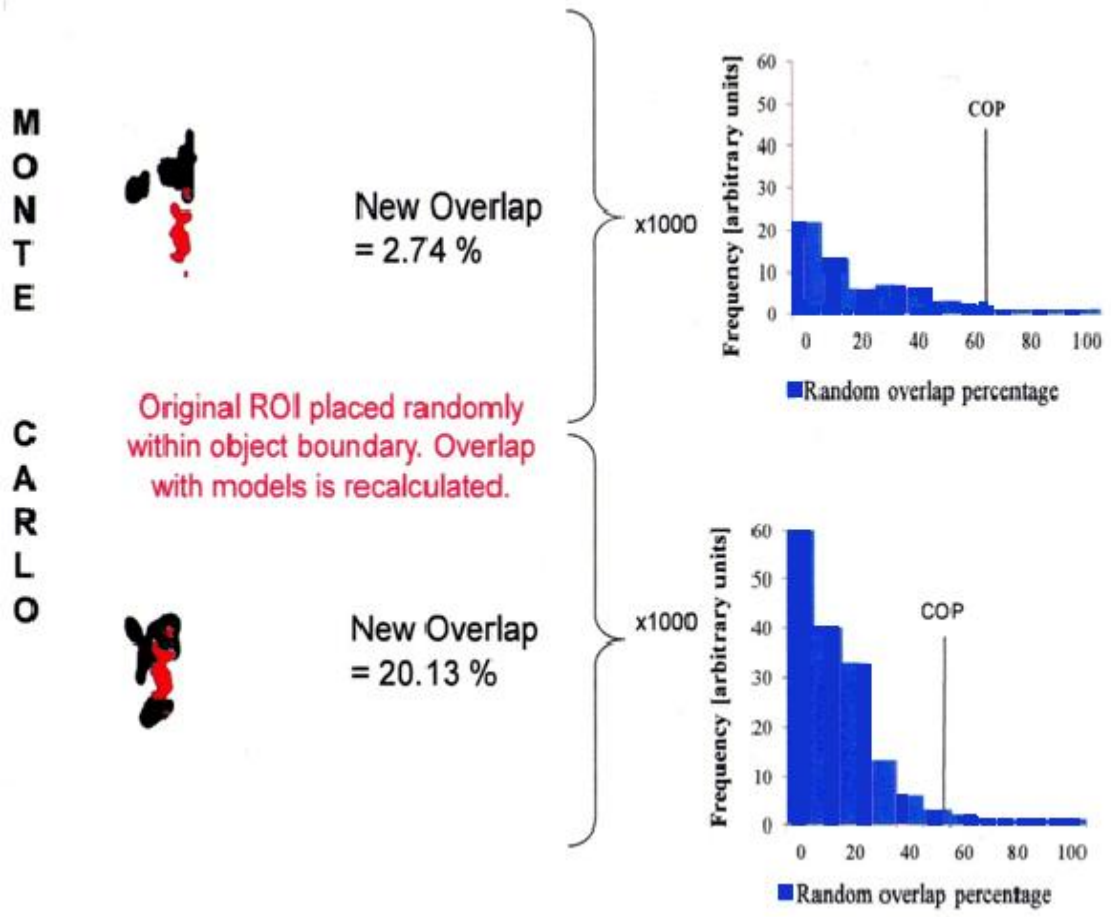


Figure 18 Visual description of Step 2 (calculation of the '*Chance Overlap Percentage'* [COP] and Step 3 (computing the measure called 'Model Matching Correspondence' [MMC] of FROA analyses.

Step 3: In order to then compare the degree of observed fixation data – model correspondence we compute a measure called 'Model Matching Correspondence' (MMC) where: MMC $_{Mx}$ = Actual Overlap Percentage (AOP) – Chance Overlap Percentage (COP) (where *Mx* is a given model). As COP and AOP are expressed in percentages of the total region area per item, the distance measure is normalised for variation in thresholded region size across items. Higher values of MMC $_{Mx}$ indicate better correspondence between the respective model and the observed fixation data (see Figure 17).

In order to factor out fixations potentially driven solely by low level statistics I computed a model based on a visual saliency paradigm which was employed as a

baseline contrast. This baseline model was computed from the visual saliency algorithm (Itti et. al., 1998), by using the Matlab Salency Toolbox implementation (Walter & Koch, 2006) with default model parameters. The default parameters were chosen as this gave optimal chance to account for the data, rather than modify it. The model generated saliency maps for each of the stimuli, giving a list of saliency values for each pixel, grouped into a saliency region map using a shape estimation function. The number of saliency regions generated was constrained to approximate the area and number of thresholded regions generated for other models (e.g., Convex, Concave, External contour). The saliency maps were thresholded and binarized using FROA (for detailed illustration see section 4.1.2), and represent the distributions of the fixation regions expected if eye movements were solely detetermined by low level image statistics such as intensity contrast, orientation, and colour.

Step 4: These saliency maps were used to compute the baseline measure MMC_{VS} where VS = visual saliency) for each item as follows: $MMC_{VS} = AOP_{VS} - COP_{VS}$. COP_{VS} is estimated using the same Monte Carlo procedure described above. This shows the extent to which observed overlap is greater or less than the 95% CI of the random distribution of the saliency algorithm.

Step 5: The final step was subtracting out the visual saliency baseline for each model as follows: $MMC_{Mx} - MMC_{VS}$. This shows the difference in overlap between the fixation data and a given model relative to the visual saliency baseline. A positive value here indicates a higher fixation data–model correspondence than that accounted for by visual saliency. In contrast, a negative value would indicate a lower fixation data–model correspondence than that accounted for by visual saliency.

The resulting MMC statistics were then subjected to analyses of variance (ANOVA) across models. In addition we also examined the generality and robustness

of the observed patterns across subjects. The subject analysis contrasted the mean normalised fixation frequency (*f*m) for thresholded and sub-thresholded object regions across subjects: here, thresholded regions corresponded to the AOIs defined by FROA. Sub-thresholded regions were defined by subtracting the thresholded AOIs from the remaining area of each stimulus image (within the bounding contour). We refer to these regions as sub-threshold AOIs. The fixation frequency distributions per subject were normalised for mean region area (across items) and converted to units of visual angle (30 pixels is equal to 1 degree of visual angle given a viewing distance of 60cm, a screen resolution of 1280 x 1024 pixels and screen size of 34 cm). Thus, this measure takes account of differences in pixel area between thresholded and non-thresholded regions. Subject analyses of mean fixation durations for thresholded versus sub-threshold AOIs using the same normalised measures are also reported.

Statistical significance is assessed relative to the two-tailed a priori alpha level ($p = .05$), unless otherwise stated. Exact probability values are reported ($p = x$) except where $p < .0001$.

Chapter 4

4.1. Experiment 1

In the preceding Chapter 3, I have outlined and justified the fundamental method used to analyse the eye movement data in this thesis. In the present Chapter I will clarify the rationale behind Experiment 1, report the results and discuss the findings.

As outlined in Chapter 2, eye movement patterns have been investigated in a variety of domains such as, reading, scene and face perception, object localisation and visual search (Land, Mennie & Rusted, 1999; Liversedge & Findlay, 2000; Henderson, Brockmole, Castelhano & Mack, 2007; Mannan, Ruddock & Wooding, 1997; Rayner, 1998; Renninger, Verghese & Coughlan, 2007; Underwood, Foulsham, van Loon, Humphreys & Bloyce, 2006). Surprisingly, to my knowledge, there have been no detailed analyses of eye movement patterns beyond two-dimensional (2D) pattern recognition (e.g., Renninger et al., 2007; Renninger, Coughlan & Verghese, 2005), that investigate three-dimensional (3D) visual object recognition. Although everyday object recognition can be accomplished quickly, and often within a single fixation for a distal stimulus, previous studies, using 2D stimuli, have shown that fixation patterns can be highly informative about shape processing during perception (e.g., Melcher & Kowler, 1999; Renninger et al., 2005; 2007; Vergilino-Perez & Findlay, 2004). For example, Melcher and Kowler (1999) have shown that the initial landing position during saccadic localisation is driven by a representation of target shape that determines Center-of-Gravity (COG) landing sites. Recent evidence also suggests that the perception of information about object presence and identity in a scene may be restricted to a relatively small region around the current fixation point

(Henderson, Williams, Castelhano & Falk, 2003), although the nature of the shape information processed during fixations, and the role of this information in object recognition, remains unclear.

In this context, a variety of different object recognition theories have been proposed which make different claims about how shape is represented. For example, some accounts suggest that shape classification is based on class-specific appearance or image-based feature hierarchies computed across multiple spatial scales (e.g., Ullman, 2007; Ullman & Bart, 2004; Ullman, Vidal-Naquet & Sali, 2002). Other image-based models have hypothesised the use of 2D views, or aspects, that conjointly encode information about shape and the spatial locations of image features (e.g., Edelman & Weinshall, 1991; Riesenhuber & Poggio, 2003; Ullman & Basri, 1991). In contrast, structural description theories propose that shape perception depends on the decomposition of object shape into generic primitives (e.g., generalised cylinders, geons or surfaces) and that recognition is mediated by representations that independently encode information about these primitives and their spatial configuration (Biederman, 1987; Hummel & Stankiewitz, 1996; Leek, Reppa & Arguin, 2005; Leek, Reppa, Rodriguez & Arguin, 2009; Marr & Nishihara, 1978). All these approaches are not mutually exclusive. For example, recent hybrid models have suggested that both image-based and structural description approaches can be accommodated within the same framework (Foster & Gilson, 2002; Hummel & Stankiewitz, 1996).

However, regardless of whether an image-based, structural description or some other form of representation is proposed, there remains a debate about the specific kinds of shape information, and shape analysis algorithms, that underlie object recognition. Theoretically, there are several different kinds of information from low-

level image contrasts (e.g., simple edges derived from luminance boundaries) to intermediate or higher-level features derived from combinations of lower-level image properties (e.g., vertices, curvature discontinuities, and volumetric parts) that could be used during shape perception. Moreover, the availability of specific kinds of shape information is dependent on the spatial scale of perceptual analysis. For example, some kinds of image features that may be useful in recognition are likely to be detected only at a relatively coarse spatial scale. These include edge co-linearity (parallelism), elongation, symmetry, aspect ratio and global outline (e.g., Biederman, 1987; Hayward, 1998; Hayward, Tarr & Corderoy, 1999). Determining elongation requires access to a relatively complete perceptual representation of object shape but it can be computed from relatively low spatial frequency information. Conversely, other potentially useful shape features may (and in some cases, must) be computed locally at a relatively finer spatial scale. These include the presence of edge boundaries, corners, vertices, surface depth and curvature. Other object properties including, for example, colour and texture, can also be computed locally. In some situations, relatively coarse global image features may be sufficient for shape classification in specific contexts – such as distinguishing between a banana (curved axis) and a cucumber (straight axis) on a kitchen table. However, real-world scenes are often cluttered, containing objects that partially occlude each other, making it difficult to reliably recover global shape descriptions all of the time. This is one reason why many current approaches to pattern classification in computer vision use algorithms based on the detection and matching (or indexing) of local image features, appearance-based feature hierarchies or interest point operators (e.g., Lowe, 2004; Mikolajczyk & Schmid, 2005; Ullman, 2007; Ullman et al., 2002).

The aim of Experiment 1 was use fixation patterns as an index of information processing during shape perception – where the assumption is that observers fixate locations of high information content (e.g., Renninger et al., 2005). In particular, our goal was to examine whether fixation patterns can be used to elucidate *local* shape analysis processes beyond those driven by low-level image statistics such as, simple contrasts in luminance, orientation and colour, in order to provide insights into the kinds of higher-level shape information that support shape perception. Some previous evidence using other tasks suggests that eye movement patterns are sensitive to 3D shape. For example, in one study, initial localisation saccades were compared when viewing 3D targets rendered with lighting and shadows or simple flat unicolor silhouettes (Vishwanath & Kowler, 2004). The results showed that saccades are sensitive to the 3D structure of an object: although the 2D projection of the target to the retina in both conditions was the same, participants showed a bias towards the 2D COG when viewing silhouettes, and the 3D COG when perceiving the target as a volume. In another study, Wexler and Ouarti (2008) have shown that saccadic eye movements during the spontaneous exploration of visual images follow surface depth gradients. A key finding was that surface orientation alone had a large effect on eye movements independent of the task when looking at stimuli in 3D.

In the present, study eye movement patterns were recorded whilst observers either actively memorised 3D novel objects or passively viewed them in a pre-test phase, and then performed a recognition memory task. An explicit memory task was employed to investigate visual object encoding and representation for a number of reasons: First, the task requires that a perceptual representation of shape is matched to long term memory allowing us to investigate local image features used during computations of shape representations. Second, it is widely used in the literature to

examine object recognition (e.g. Tarr & Bulthoff, 1995; 1998). Third, the Active vs. Passive manipulation allowed us to compare explicit vs. implicit shape encoding. Fourth, it was important to use a task in which stimulus exposure duration was relatively long in order to acquire sufficient eye movement data for analyses.

Following pre-test and test phase, the observed fixation patterns were compared to the predicted distributions derived from different models of shape information content. My interest was to examine whether fixation patterns can be driven by higher-level shape features, beyond low-level image statistics alone, and so I used visual saliency as a baseline contrast (Itti, Koch & Niebur, 1998; Koch & Ullman, 1985; Walther & Koch, 2006). The visual saliency model generates saliency maps based on weighted contrasts in luminance, orientation and colour. This model has been widely applied to eye movement studies of scene perception although its efficiency in predicting fixation patterns remains the subject of on-going debate (e.g., Baddeley & Tatler, 2006; Cristino & Baddeley, 2009; Henderson, Brockmole, Castelhano & Mack, 2007). The main question in this study was whether specific models of shape analysis could account for fixation patterns beyond that explicable by visual saliency.

I have evaluated three different models. *Model 1*: was based on external global shape features defined by bounding contour and was based on the hypothesis from recent work showing that outline shape influences object recognition (e.g., Hayward, 1998; Hayward Tarr, & Corderoy, 1999; Lloyd-Jones & Luckhurst, 2002). *Model 2* and *Model 3* were derived from the large body of work highlighting the importance of curvature in shape perception (e.g., Attneave, 1954; Barenholtz, Cohen, Feldman & Singh, 2003; Bertamini, 2008; Biederman, 1987; Cate & Behrmann, 2010; Cohen, Barenholtz, Singh & Feldman, 2005; Cohen & Singh, 2007; De Winter & Wagemans,

2006; Feldman & Singh, 2005; Hoffman & Richards, 1984; Hoffman & Singh, 1997; Lim & Leek, in press). This work has shown robust perceptual sensitivity to curvature extrema, where negative minima define concave image regions (relative to the figure), and positive maxima define convexities – a phenomenon that has also recently been demonstrated in infants as young as five months old (Bhatt, Hayden, Reed, Bertin & Joseph, 2006). Previous studies have examined curvature in the context of contour-defined 2D images such as polygons and line drawings (Cohen & Singh, 2007; Cohen, Barenholtz, Singh & Feldman, 2005; De Winter & Wagemans, 2006) in which curvature minima and maxima are defined along the occluding contour boundary. In comparison, there is relatively little data examining the role of curvature discontinuities defined by changes in the surface (rather than contour) curvature polarity of 3D objects. I examined two models of internal surface curvature defined by local internal convex curvature maxima (Model 2) and local internal concave minima (Model 3).

4.1.1. Method

Participants

60 students from Bangor University (36 female, mean age 20.83 years, *SD* = 4.33, 53 right handed) participated in the study for course credit. All participants had normal or corrected to normal visual acuity. Informed consent was obtained from each participant prior testing in line with local ethics committee and BPS guidelines.

Stimuli

There were 12 novel objects (see Figure 19) each consisting of a unique spatial configuration of four volumetric parts. The parts were uniquely defined by variation among non-accidental properties (NAPs) comprising: Edges (Straight vs. Curved),

symmetry of the cross section, tapering (co-linearity) and aspect ratio (Biederman, 1987).

Figure 19 The 12 novel objects used in the current study.

The object models were produced using Strata 3D CX software (Strata, USA) and rendered in Matlab using a single light source (top left) with anti-aliasing and scaled to fit within an 800 x 800 pixel frame (normalised in size across objects). All stimuli were uniformly coloured in mustard yellow: R = 227, G = 190, B = 43. Stimuli subtended 18 degrees of visual angle horizontally with participants seated 60 cm from the display. This scale was chosen to induce saccadic exploration over the stimuli. Each stimulus was rendered depicting the object from six different viewpoints at successive 60 degree rotations in depth around a vertical axis perpendicular to the line of sight. The zero degree viewpoint was a 'canonical' three-quarter view (see Figure 20). The 0, 120 and 240 degree versions served as familiar (pre-test) viewpoints, and the 60, 180 and 300 degree versions as novel viewpoints.

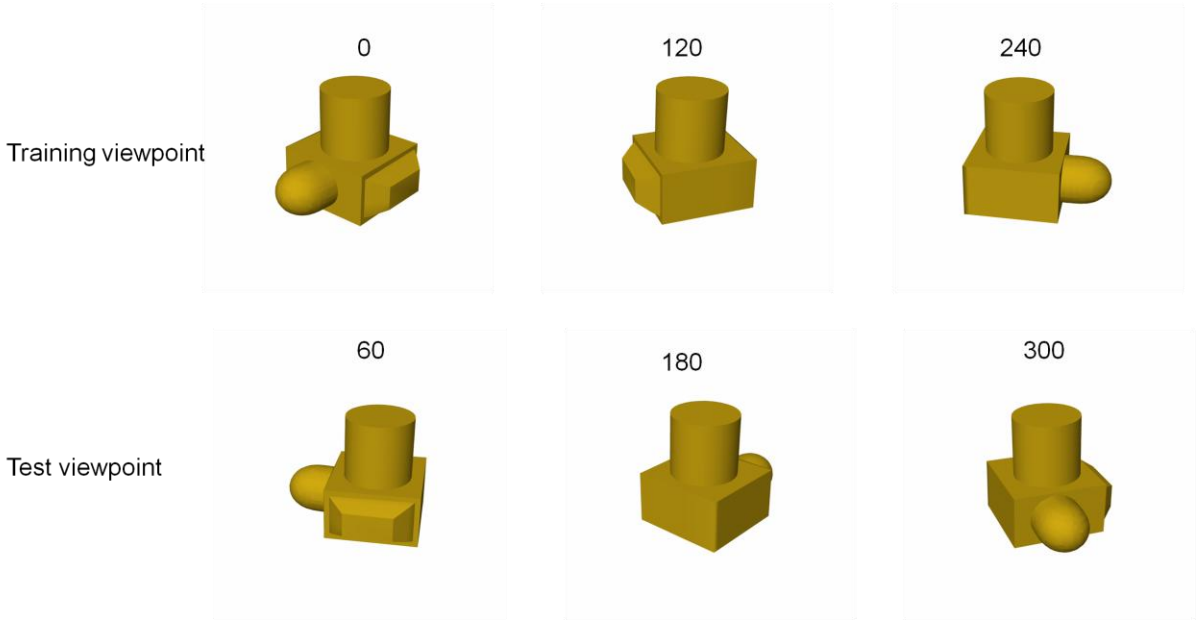


Figure 20 An illustration of the three trained and three novel viewpoints used in the experiment.

Apparatus

Eye movement data were recorded on a Tobii 1750 (Tobii Technology, AB, Sweden) binocular corneal reflection (CR)-based remote eye tracking system (< 0.5 degrees accuracy, 0.25 degrees spatial resolution, and drift < 1 degree). Stimuli were presented on a TFT monitor running at a resolution of 1280 x 1024 pixels and 60 Hz refresh rate. Mean surround luminance was 114.7 cd/m^2 (SD = 0.25 cd/m^2) measured with a Minolta CS-100 photometer. A chin rest was used to stabilise the participant's head at a 60 cm viewing distance and a standard USB keyboard was used for response collection.

Design and Procedure

Each participant initially completed a nine point eye tracking calibration procedure. This required the participants to view a static blue dot that appeared, randomly, in each of 9 possible screen locations. Noisy calibration points were resampled algorithmically to ensure accuracy and validity.

The study comprised two phases: pre-test and test phase. All subjects completed both phases. There were two versions of the pre-test: Active learning and Passive viewing, with participants assigned randomly to one of the two groups. Eye movement patterns have been found to differ depending on task requirements (e.g. Yarbus, 1967). Hence, employing two tasks (explicit vs. implicit) that potentially activate different representational structures, or the same structures, but in different ways, provides an opportunity to examine the robustness of data-model correspondences associated with local shape analyses during perception across different task requirements. For both pre-test groups the trial structure was the same comprising 18 trials (6 targets x 3 viewpoints). On each trial, participants initially fixated a square (1° x 1° visual angle) for 2000 ms presented in the centre of the display vertically and either 9 degrees to the left or right of the object. In the pre-test phase, following a 2000 ms blank ISI a single stimulus was presented in the centre of the monitor for 10 seconds.

In the active learning group, participants were instructed to study the shape of each stimulus and to try to memorise it for a subsequent recognition memory task. They were told that they would see six objects presented in a three different viewpoints. In the passive viewing group, participants were instructed only to visually inspect each stimulus. They were not told to memorise the objects, nor forewarned about the subsequent recognition memory task. For each pre-test group, half of the participants viewed objects 1-6, and half viewed objects 7-12. The objects viewed in the pre-test phase were assigned as targets. Thus, all 12 stimuli were used both as targets and distracters across groups. In the test phase (N trials = 72), targets (N = 6, depending on the set shown in pre-test) and distracters (N = 6) were presented in random order each at six viewpoints (3 familiar and 3 novel). Across groups there

were 12 targets and 12 distracters (each shown from six viewpoints). The trial structure was the same as in the pre-test phases, except that the stimuli were presented until the participants made a keyboard response. Both pre-test groups were given the same instructions in the test phase. They were asked to determine and respond via a key-press (k – 'yes'/ d – 'no') whether the presented stimulus was one of the objects viewed during the pre-test phase regardless of the viewpoint shown. Eye movement data, response time (RT) and accuracy were recorded as dependent measures. The experiment lasted approximately 30 minutes.

4.1.2. Algorithmically generated model predictions

Generating Model Predictions

The predicted distributions for each model of image information content were algorithmically computed from the 3D object models using Matlab. An illustration of the predicted thresholded fixation region maps for the tested models can be seen in Figure 21 below. All of these predicted distributions were generated algorithmically.

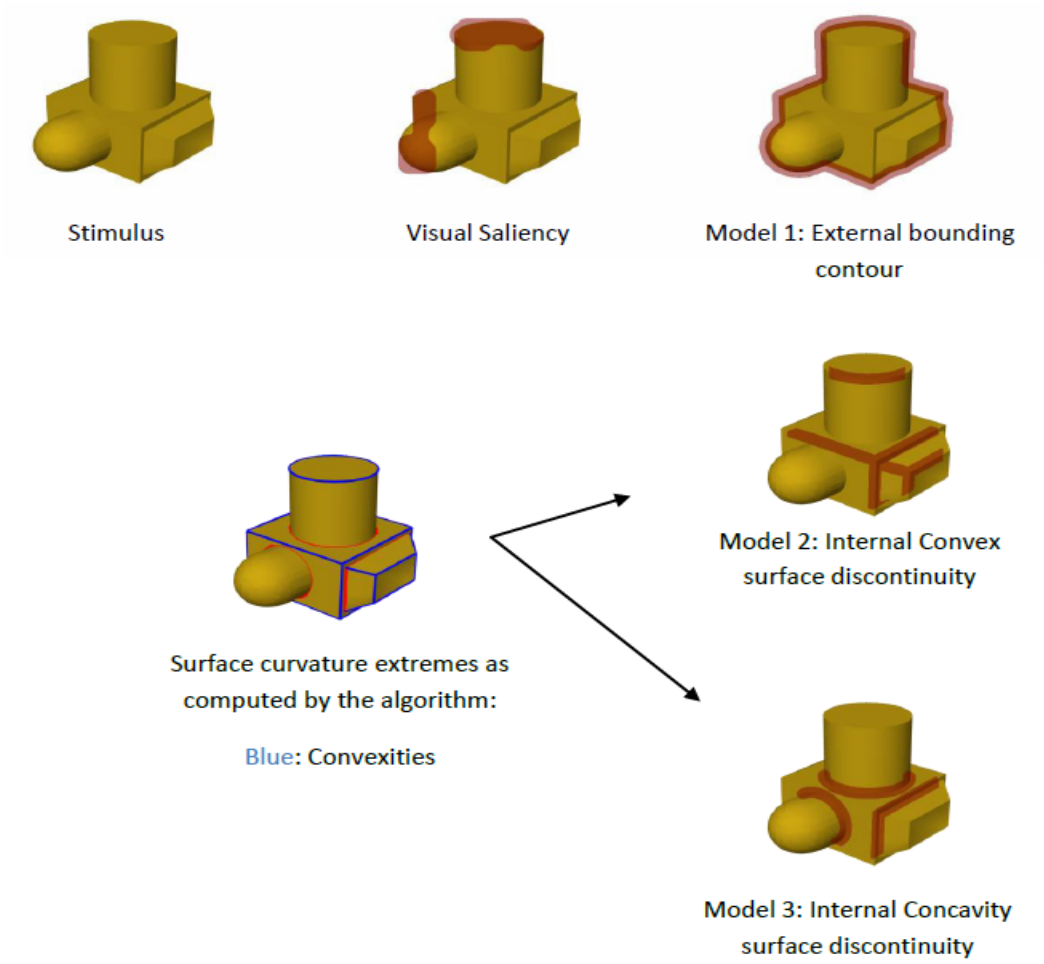


Figure 21 An illustration of the predicted thresholded fixation region maps for the three tested models (External bounding contour, Internal Convex surface discontinuity, and Internal Concave surface discontinuity). All of these predicted distributions were generated algorithmically.

Model 1: External (bounding) Contour

Model 1 examined the extent to which fixation patterns focus on external global shape features defined by bounding contour. This hypothesis derives from previous work showing that outline shape influences object recognition (e.g., Hayward, 1998; Hayward et al., 1999; Lloyd-Jones & Luckhurst, 2002). The bounding contour was computed using an edge detector on the image silhouette of the stimuli. It was then re-plotted using lines of 0.66 degrees width (see Figure 21). This value was used as it produced models of a similar size as the binarised eye movement data.

Model 2: Internal Convex Surface Discontinuity

Model 2 generated predicted fixation regions based on the locations of local features defined by *convex* surface curvature maxima. These were generated by applying a curvature estimation algorithm derived from Taubin (1995) to the object mesh models using the Peyre Matlab toolbox. From this we extracted edges along convex curvature maxima (see Figure 20). The convex features were re-plotted using lines of 0.66 degrees width. Edges on the exterior bounding contour were deleted. Due to the nature of the stimuli, convexities can occur both inside and on the bounding contour of an object but concavities are more likely to occur on the internal contour (see Figure 21). By keeping internal features only, we are able to compute a bias free measure of the preference for convex or concave image features.

Model 3: Internal Concave Surface Discontinuity

Model 3 generated predicted fixation regions based on the locations of local features defined by *concave* surface curvature minima (see Figure 21). The same curvature estimation method was used as for Model 2, except that here I extracted edges along concave curvature minima. As with Model 2 edges falling on the external bounding contour were removed.

Visual Saliency baseline

As outlined before, the visual saliency model generates saliency maps based on weighted contrasts in luminance, orientation and colour. Although this model has been widely applied to eye movement studies of scene perception its efficiency in predicting fixation patterns remains a subject of on-going debate (e.g., Baddeley & Tatler, 2006; Henderson, Brockmole, Castelhano & Mack, 2007; Cristino & Baddeley, 2009).

To create this model I used the Saliency Toolbox implementation in Matlab (Walther & Koch, 2006). The model was run on each of the 72 stimulus images (12 objects x 6 viewpoints) used in the recognition task and generated a saliency map for each stimulus. The output of the toolbox is a list of saliency values for each pixel which are grouped into a saliency region map using shape estimation function (see Walther & Koch, 2006). The number of saliency regions generated was constrained to approximate the area and number of thresholded regions generated for the other models. The saliency maps were thresholded and binarised using FROA in the same way as the empirically derived fixation data from the recognition task. These maps represent the thresholded distributions of fixation regions we would expect if eye movements were determined solely by low-level image statistics. As I mentioned earlier, this model was used as a baseline contrast as the question of interest here was whether specific models of shape analyses could account for fixation patterns beyond the explicable by visual saliency.

4.1.3. Behavioural data analyses

Analyses of behavioural data (Test Phase)

Analyses were conducted on the mean median test phase RTs and accuracy data. Only RTs for correct responses were included. Mean median RTs and accuracy rates are shown in Table 1 (targets only) for both the active learning and passive viewing pre-test groups.

Table 1 The mean median RTs and accuracy rates (targets) for familiar and novel viewpoints in the Test phase. Standard error of the mean is shown in parentheses.

	Pre-Test Group			
	Active Learning		**Passive Viewing**	
	RTs (ms)	% Correct	RTs (ms)	% Correct
Familiar Views	1323.76 (51.34)	93 (1.40)	1499.08 (63.66)	74 (0.30)
Novel Views	1501.22 (102.78)	87 (1.80)	1695.28 (112.98)	69 (0.30)

Reaction time data

A 2 (Pre-test task: Active learning vs. Passive view) x 2 (Viewpoint: Familiar vs. Novel) x 2 (Stimulus type: Target vs. non-target) mixed factor ANOVA showed a significant main effect of pre-test task, F (1, 35) = 13.65, p = .001, η_p^2 = .281, and a significant interaction between viewpoint and stimulus type, F (1, 35) = 6.86, p = .013, η_p^2 = .164. There were no interactions involving the factor of pre-test group. As seen in Table 1, these results indicate that RTs were faster overall in the test phase for the Active learning than the Passive viewing pre-test group. Post-hoc planned comparisons showed that target RTs were faster for familiar than novel viewpoints in both the Active learning, t (71) = -2.48, p = .018, R^2 = 0.28, and Passive viewing pre-test groups, t (71) = -3.12, p = .003, R^2 = 0.35. In contrast, there was no difference in mean median RTs between familiar M = 1653.56 ms; SE = 66.47 ms and novel views M = 1570.56 ms; SE = 60.15 ms for non-targets t (71) = -1.37, p = .17, ns.

Accuracy rates

Accuracy data were analysed using non-parametric significance test for related groups unless otherwise stated. For the Active learning pre-test group there was no

significant difference between accuracy rates for targets (*M* = 90%; *SE* = 0.012) versus non-targets (*M* = 90%; *SE* = 0.013), *z* =.226, *p* = .821, ns. For test phase target trials, accuracy was significantly higher for trained views (*M* = 93%; *SE* = 0.014) than for novel views (*M* = 87%; *SE* = 0.018), *z* = -2.93, *p* = .003. There was no significant difference in accuracy for non-targets across viewpoints *z* = -1.47, *p* = .141, ns. For the Passive viewing pre-test group overall response accuracy was higher for non-targets (*M* = 81.5%; *SE* = 0.015) than targets (*M* = 72%; *SE* = 0.021), *z* = 3.21, *p* = .001.

In the test phase there was no difference in accuracy for targets between familiar (*M* = 74%; *SE* = 0.031) and novel viewpoints (*M* = 69%; *SE* = 0.029), *z* = -1.77, *p* = 0.77, ns, or for non-targets: Familiar, *M* = 84%; *SE* = 0.016; novel, *M* = 87%; *SE* = 0.014, *z* = -1.75, *p* = 0.79, ns. Overall accuracy rates for the Active Learning group (*M* = 90%, *SE* = 0.01) and Passive Viewing group (*M* = 82%, *SE* = 0.01) were significantly different Mann Whitney: *z* = - 4.37, *p* <.001.

4.1.4. Analyses of eye movement data

Pre-Test Phase: Active Learning Group

A subject analysis was first performed to test the generality and reliability of the thresholded fixation region distributions across participants. This was done by contrasting the frequency of fixations between thresholded (fixated) and sub-threshold (insufficient number of fixations) AOIs (see Methods). Separate subject analyses were performed on the pre-test (targets) and test phase (targets and non-targets) data. Table 2 shows the mean normalised frequencies for the thresholded and sub-thresholded AOIs across participants. These data show that the mean normalised fixation frequency for thresholded AOIs is higher than for sub-threshold AOIs in both the active learning and test phases for targets, and in the test phase, for targets and non-targets

Table 2. The mean normalized fixation frequencies (mean fixation per degree of visual angle) for thresholded and sub-threshold AOIs for the Active Learning group in both the pre-test and test phases. Standard error of the mean is shown in parentheses.

	Pre-test phase	**Test phase**	
	Targets	**Targets**	**Non-targets**
Thresholded AOIs	0.52 (0.03)	0.38 (0.04)	0.39 (0.03)
Sub-threshold AOIs	0.001 (0.0001)	0.03 (0.002)	0.03 (0.003)

For the active learning phase, there was a significant difference between the mean normalized fixation frequencies across participants for the thresholded vs. sub-threshold AOIs, t (29) = 16.95, $p < .001$, $R^2 = 0.95$. For the test phase, a 2 (AOI: Thresholded vs. Sub-threshold) x 2 (Stimulus: Target vs. non-target) repeated measures ANOVA showed a significant main effect of AOI, F (1, 29) = 109.22, $p < .001$, $\eta_p^2 = .790$, but no other main effects or interactions. These analyses show that the fixation regions identified using FROA were robust across subjects for the active learning group.

Analyses of the local shape feature analysis patterns

(Active learning task: Pre-test Phase)

The remaining analyses of the fixation data for the active learning group were computed across items. For the pre-test phase, the distributions of fixation regions to targets presented at trained viewpoints (N=36) were initially analysed across 3 epochs allowing us to compare the spatial distributions of fixations occurring at different time periods following stimulus onset. To do this, fixations were divided subject-by-subject

and trial-by-trial into bins containing the first third, middle third and final third (e.g., for a particular subject making 9 fixations on a given item, fixations 1-3 would be allocated to the first bin, 4-6 to the second bin and 7-9 to the third bin). The respective bins were then pooled across subjects for each stimulus. A 3 (Epoch) x 4 (Models 1-3, plus the Baseline saliency model) repeated measures ANOVA on the MMC distance measure across targets showed a significant main effect of Model, F (3, 105) = 13.33, $p < .0001$, $\eta_p^2 = .276$, but no main effect of Epoch F = (6, 210) = .737, p = .621, and no significant interaction. In the absence of an interaction, the MMC distance statistics were collapsed across epoch.

A one-way ANOVA across models on the MMC measure was significant, F (3, 140) = 10.08, $p < .001$. Subsequent post-hoc analyses using the Bonferroni test showed that the pairwise contrasts between models were significantly different for Internal features Concave vs. Visual saliency, $p < .0001$; Internal features convex vs. Visual saliency, $p < .0001$; External features vs. Visual saliency, p = .029. There were no other significant differences. These analyses show that the fixation data-model correspondence is greater for all three models of shape analysis than the baseline saliency model. However, there were no differences in the degree of the data-model correspondence between the models of shape analysis in the pre-test phase.

Passive Viewing Group: Pre-Test Phase

An initial analysis by subjects was undertaken contrasting the frequency of fixations between thresholded and sub-threshold AOIs using normalized frequency statistics. Table 3 shows the mean normalised frequencies for the thresholded and sub-threshold AOIs across participants. These data show that the mean normalised fixation frequency for thresholded AOIs is higher than for sub-threshold AOIs in both the

passive viewing and test phases for targets, and for targets and non-targets in the test phase.

Table 3. The mean normalized fixation frequencies (mean fixation per degree of visual angle) for thresholded and sub-threshold AOIs for the Passive Viewing group in both the pre-test and test phases. Standard error of the mean is shown in parentheses.

	Pre-test phase	**Test phase**	
	Targets	**Targets**	**Non-targets**
Thresholded AOIs	0.49 (0.03)	0.42 (0.05)	0.43 (0.04)
Sub-threshold AOIs	0.014 (0.01)	0.003 (0.02)	0.003 (0.02)

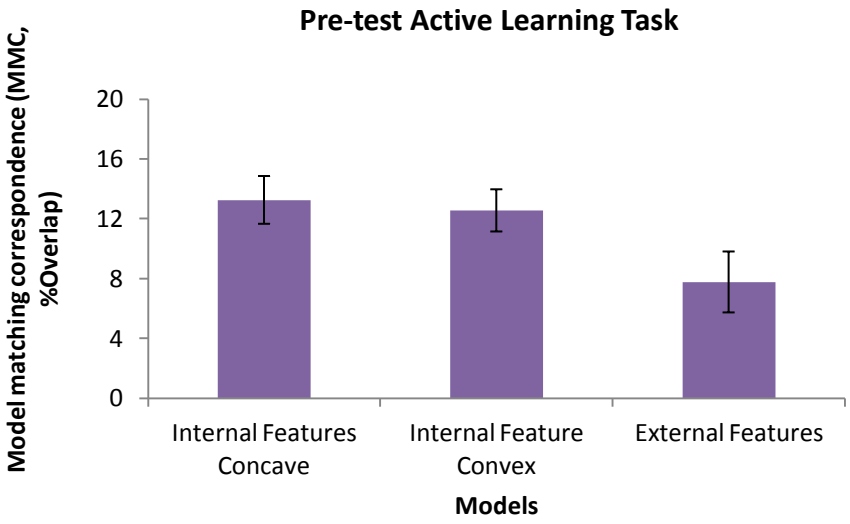
There was a significant difference between the mean normalised fixation frequencies across participants for the thresholded versus sub-threshold AOIs, t (29) = 15.63, $p < .0001$, $R^2 = 0.94$. For the test phase, a 2 (AOI: Thresholded vs. Sub-threshold) x 2 (Stimulus: Target vs. non-target) repeated measures ANOVA showed a significant main effect of AOI, F (1, 29) = 103.89, $p < .001$, $\eta_p^2 = .782$, but no other main effects or interactions.

Analyses of the local shape feature patterns (Pre-test, Passive viewing task)

As previously, the distributions of fixation regions to targets presented at trained viewpoints (N=36) were analysed across 3 epochs following stimulus onset. At each epoch, Fixation Region Overlap Analyses (FROA) was used to compute the observed overlap between the gaze data and each model prediction relative to the random Monte Carlo distribution. A 3 (Epoch) x 4 (Models 1-3, plus the Baseline saliency model) repeated measures ANOVA on the MMC distance measure across targets showed a significant main effect of Model, F (3, 105) = 25.66, $p < .0001$, $\eta_p^2 =$

.423, but no effect of Epoch F = (2, 70) = 1.47, p = .235ns., and no interaction. In the absence of any interaction, normalised distance was collapsed across epoch. A one-way ANOVA on mean MMC values across models (Visual saliency, Internal features convex, Internal features concave, External features) was significant, F (3, 140) = 15.72, p < .0001. Post-hoc analyses showed that the pairwise contrasts between models were significantly different for Internal features concave vs. Visual saliency, p < .0001; Internal features convex vs. Visual saliency, p < .0001; External features vs. Visual saliency, p = .007. In addition, unlike the Active learning group there was also a significant difference between Internal features concave vs. External features, p = .020. There were no other significant contrasts (see Figure 22).

(a)



(b)

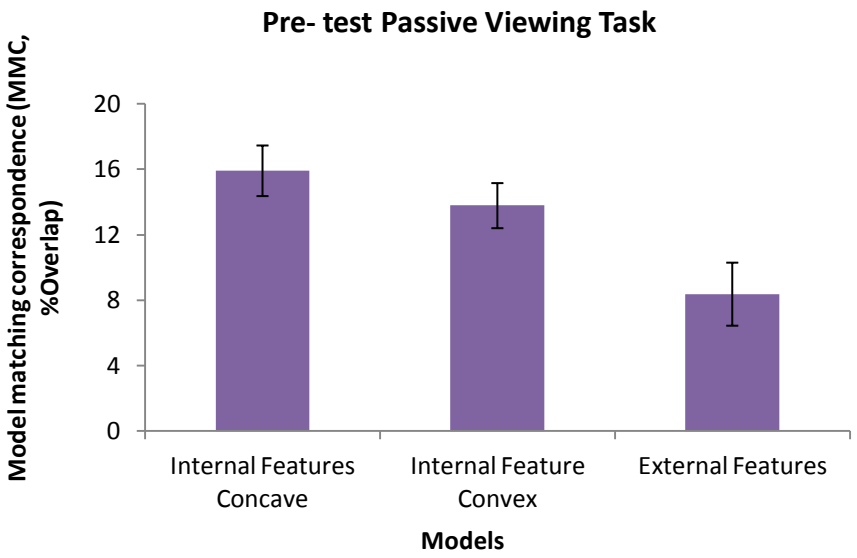


Figure 22 Mean MMC $_{Mx}$ – MMC $_{vs}$ measure of data-model correspondences between models (relative to visual saliency) for (a) pre-test Active Learning group (b) pre-test Passive Viewing group. Bars show standard error of the mean (% overlap).

Pre-test phase: Active Learning and Passive Viewing Groups

These analyses were run on the MMC data from the pre-test phase. A 2 (Active Learning, Passive viewing, *BS*) x 3 (Model: Internal features concave, Internal features convex, External features, Visual saliency, *WS*) mixed ANOVA across target mean MMC for the pre-test phase data showed a significant main effect of Model, F (3, 210) = 22.95, p < .0001, ηp^2 = .247, but no effect of Group, F (1, 70) = 1.11, p = .294 ns, and no interaction. Given the lack of interaction I have collapsed the pre-test phase data across group and run post hoc analyses across model means. Post-hoc comparison (t-test) was significantly different for Internal concave vs. Visual saliency, p <.0001, R^2= 0.63, Internal convex vs. Visual saliency, p <.0001, R^2= 0.60, External features vs. Visual saliency, p =.0001, R^2= 0.42, External features vs. Internal concave, p =.003, R^2= 0.34, External features vs. Internal convex, p =.006, R^2= 0.32, but not Internal Concave vs. Internal Convex, p =.232ns.

Test Phase: Active Learning and Passive Viewing Groups

These analyses were run on the MMC data from the test phase. Figure 23 shows the mean $MMC_{Mx} - MMC_{Vs}$ vs. contrasts across models. A 2 (Group: Active learning vs. Passive viewing) x 2 (Stimulus type: Target vs. Non target) x 4 (Model: Models 1-3, plus the Baseline saliency model) mixed design ANOVA showed a significant main effect of Model, F (3, 138) = 39.08, p <.001, η_p^2 = .459. There were no other significant main effects or interactions. Post-hoc analyses showed that the pairwise contrasts (Bonferroni) between models were significantly different for all model-visual saliency baseline contrasts: External features vs. Visual saliency, p < .0001; Internal features concave vs. Visual saliency, p < .001; Internal features convex vs. Visual saliency, p < .001. In addition, mean fixation data-model correspondence was higher for the Internal Concave vs. Internal convex contrast, p = .024; Internal features concave vs. External features, p = .001; and Internal features convex vs. External features, p = .030.

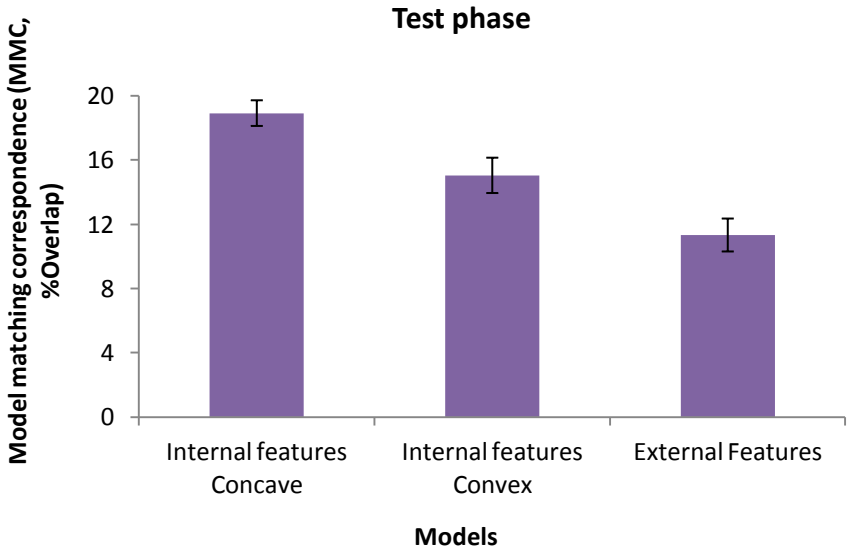


Figure 23 Mean $MMC_{Mx} - MMC_{Vs}$ measure of data-model correspondences between models (relative to visual saliency) for the recognition memory test phase (collapsed across pre-test groups). Bars show standard error of the mean (% overlap).

Analysis of fixation duration

We also conducted analyses of fixation duration, contrasting mean durations for fixations falling within the FROA-defined thresholded regions versus duration for fixations falling outside of the thresholded regions (sub-threshold AOI). Separate analyses were conducted for the Active Learning and Passive Viewing groups, and for the pre-test and test phases.

Active Learning Group

For the pre-test, mean fixation durations were longer for fixations within thresholded AOIs (*M* = 295.59 ms, *SE* = 19.76 ms) than for those within sub-threshold AOIs (*M* = 266.58 ms, *SE* = 16.16 ms). This difference was significant, *t* (29) = 3.35, *p* = .002. R^2 = 0.53. For the test phase target mean fixation duration was also longer for thresholded (*M* = 259.19 ms, *SE* = 15.88 ms) than sub-threshold AOI fixations (*M* = 223.25 ms, *SE* = 10.77ms. The same pattern was found for non-targets: thresholded AOI fixations (*M* = 252.95 ms, *SE* = 14.30 ms), sub-threshold AOI fixations (*M* = 226.26 ms, *SE* = 10.86 ms). A 2 (AOI: Threshold vs. Sub-threshold) x 2 (Stimulus: Target vs. non-target) repeated measures ANOVA on the test phase mean fixation duration data showed a significant main effect of AOI, *F* (1, 29) = 22.52, *p* < .001, η_p^2 = .437, but no other main effects or interactions.

Passive Viewing Group

Analyses of the pre-test phase fixation duration data showed longer mean durations for the thresholded AOI fixations (*M* = 313.76 ms, *SE* = 31.67 ms) than for sub-threshold AOI fixations (*M* = 275.87 ms, *SE* = 23.03 ms). This difference was significant, *t* (29) = 3.17, *p* = .004, R^2 = 0.51. For the test phase target mean fixation duration was also longer for thresholded AOI fixations (*M* = 266.39 ms, *SE* = 23.30 ms) than for sub-threshold AOI fixations (*M* = 239.75 ms, *SE* = 16.48 ms). The same

pattern was found for non-targets: thresholded (*M* = 265.01 ms, *SE* = 22.97 ms), sub-threshold fixations (*M* = 236.03 ms, *SE* = 16.54 ms). A 2 (AOI: Threshold vs. Sub-threshold) x 2 (Stimulus: Target vs. non-target) repeated measures ANOVA on the test phase duration data showed a significant main effect of AOI, F (1, 29) = 10.77, p = .003, η_p^2 = .271, but no other main effects or interactions. These analyses show that for both the Active Learning and Passive Viewing groups mean fixation durations were longer for fixations falling within thresholded regions than for those outside of those regions in both the pre-test and test phases of the study.

Active Viewing vs. Passive Learning

Finally, we also contrasted mean fixation durations for thresholded and sub-threshold fixations across task groups on the pre-test phase data. A 2 (Task: Active vs. Passive) x 2 (AOI: Threshold vs. sub-threshold) mixed ANOVA showed a significant main effect of AOI, F (1, 58) = 20.54, p < .0001, η_p^2 = .262, but no other main effects or interactions. This suggests that mean durations were not significantly different between pre-test task groups.

Summary of the Results

Behavioural data

- The overall RTs and accuracy data in the test phase showed faster and more accurate responses for the Active learning pre-test group than the Passive viewing pre-test group.
- Both Active and Passive pre-test groups showed significantly faster RTs for targets at familiar than novel viewpoints and no differences in RTs for non-targets.

Eye movement data

- Eye movement analyses showed a higher fixation frequency for thresholded AOIs than for sub-thresholded AOIs for active learning, passive viewing phase for targets, and in the test phase for targets and non-targets.
- Analyses of local shape feature patterns for both groups (Active learning, Passive viewing) in the pre-test phase showed that data-model correspondences were greater for Internal Concave, Internal Convex, and External contour models than the Visual saliency model.
- The Passive viewing pre-test group showed greater data-model correspondence for Internal Concave model relative to the External contour model.
- Analyses of local shape feature patterns in the test phase for both groups showed greater data-model correspondences for Internal Concave, Internal Convex, and External contour models relative to the Visual saliency model.
- Data-model correspondence was higher for the Internal Concave than the Internal Convex models.
- Fixation duration for both Active and Passive learning groups in both pre-test and test phases of the study were longer for fixations falling within thresholded regions, than sub-tresholded regions. This pattern was evident for both targets and non-targets.

4.1.5. General discussion

In this study we examined the fixation patterns during the perception and recognition of 3D objects. Observers either actively memorised or passively viewed sets of visually similar novel objects prior to performing a recognition memory test. The main empirical findings were as follows: First, the analyses of the RT and accuracy data showed that while observers performed the recognition memory task more accurately following the active learning than passive viewing pre-test, the

patterns of test phase RTs for both groups showed faster responses for targets presented at familiar (pre-test) viewpoints than at novel viewpoints. This finding suggests that the participants in the active learning and passive viewing pre-test groups performed the recognition memory task in a similar way, and that recognition in both groups was viewpoint-dependent. This finding is consistent with other reports in the literature that recognition is mediated by viewpoint-dependent representations of object shape (e.g., Bülthoff & Edelman, 1992; Edelman & Weinshall, 1991; Reisenhuber & Poggio, 1999; Tarr & Bülthoff, 1998). The Active learning pre-test group showed no differences between targets and non-target accuracy rates, whereas the Passive viewing pre-test group showed significantly higher accuracy for non-targets than targets. One possible explanation of these results could be speed-accuracy trade off as Passive viewing pre-test group had faster RTs for targets than non-targets. Moreover, the results are consistent with previous theories of perceptual matching (Krueger, 1978; Ratcliff, 1981) proposing that accuracy and speed are influenced by the number of features 'matches' or 'non-matches' necessary to match perceptual representations to memory. Under the current context the participants were more accurate responding to non-targets which could be a result of having more feature non-matches between study and test phase, than feature matches. Given that the Passive view pre-test group were not memorising the objects during pre-test phase, the so called targets had less matches between study and test phase than non-targets.

The analyses of the fixation data showed a consistent pattern of data-model correspondences across tasks. More specifically, during both active learning and passive viewing pre-test phases, and during the recognition memory task, we found evidence that fixation patterns are driven by regions containing higher-level shape information defined either by the external bounding contour or by internal regions of

convex or concave surface discontinuity. Moreover, despite the similarity of the patterns of data-model correspondences between the active learning and passive viewing groups, the distributions of fixations across object shape features differed between the study and test phases: notably, during the recognition task we found a preference for fixation at internal regions of surface concavity.

These findings are consistent with previous studies demonstrating the importance of curvature singularities in the visual perception of shape (e.g., Attneave, 1954; Barenholtz et al., 2003; De Winter & Wagemans, 2006; Feldman & Singh, 2005; Hoffman & Richards, 1984; Hoffman & Singh, 1997) – moreover this is, to our knowledge, the first empirical confirmation from 3D object perception and recognition showing a preference for fixation at these regions in both active and passive viewing tasks. In addition, the finding of a preference for fixation at regions of concave surface discontinuity during the recognition task provides evidence for a direct link between the encoding of information about surface concavity and object recognition. The current results raise two essential issues; (1) the apparent preference for fixation at regions of surface concavity during recognition and (2) the observation of similar fixation distributions, and similar perceptual strategies for the acquisition of shape information, across active and passive viewing tasks. I discuss both of these issues in turn.

Eye movements, surface curvature and recognition

In other domains, such as scene perception, there is ongoing debate about the relative influence of bottom-up, stimulus-driven factors and top-down, conceptually driven factors in determining eye movement behavior (e.g., Foulsham & Underwood, 2007; Henderson et al., 2007; Itti et al., 1998). The data here show that fixation patterns during the perception and recognition of object shapes cannot be solely

accounted for by low-level visual saliency. Moreover, the data further showed a fixation preference for concave regions over convex regions during recognition. This concave preference did not interact with pre-test group. That is, regardless of whether observers actively memorized or passively viewed objects in the pre-test, they showed a preference for fixation at regions of internal concave minima in the recognition task. How can this pattern of results be accounted for?

One possibility is that observers specifically fixate those particular internal regions because they are the optimal locations for extracting global (e.g., outline) shape properties rather than because of their status as regions containing perceptually relevant shape curvature. However, such an account would not provide an obvious explanation for the apparent preference for fixation at regions of concave surface discontinuity in the recognition task but not in the pre-test phase. Additionally, it is more likely that the optimum location for extracting global shape attributes (e.g., elongation, orientation, or symmetry) would be close to the center of mass but this is clearly not the case as early COG fixations were removed from the data. Rather, the preference for fixation at regions of concavity during the recognition task is consistent with hypotheses that outline a special functional status for concave minima in shape recognition (e.g., Feldman & Singh, 2005; Hoffman & Richards, 1984; Lim & Leek, 2012).

One influential hypothesis is that concave regions play an important role as segmentation points allowing for the computation of parts-based structural descriptions (e.g., Hoffman & Richards, 1984; Marr & Nishihara, 1978). In this context, one interesting aspect of the data stems from the concurrent observation of a fixation preference for concave surface minima along with viewpoint- dependent performance in the recognition task. The former finding is consistent with the claim

that negative curvature minima play a functional role in part segmentation during the derivation of a structural description representation (e.g., Biederman, 1987; Hoffman & Richards, 1984; Marr & Nishihara, 1978), while the latter finding, according to some interpretations of viewpoint-dependent effects, is consistent with image-based view interpolation models (e.g., Bulthoff & Edelman, 1992; Edelman & Weinshall, 1991; Riesenhuber & Poggio, 1999; Tarr & Bulthoff, 1998; Ullman, 1998). How might these two findings be reconciled? One possibility is that they reflect different stages of object processing within the context of more recent hybrid models of object recognition, which propose the use of both structural description and image-based representations (e.g., Foster & Gilson, 2002; Hummel & Stankiewicz, 1996).

Alternatively, within an exclusively image-based approach, one could suppose that the apparent preference for fixation at regions of surface curvature concavity reflects the encoding of local depth information in image based object representations. Some supporting evidence comes from the recent demonstration by Wexler and Ouarti (2008) showing that saccadic eye movements during the spontaneous exploration of visual images follow surface depth gradients. Thus, these findings present a challenge to image-based models that are based solely on the use of 2D image properties (e.g., Bulthoff & Edelman, 1992) and appear to necessitate, within this theoretical framework, the encoding and use of image features that specify local surface depth information.

Task generality of shape analysis patterns

A further aspect of the results that is of theoretical interest is the consistency of the patterns of data–model correspondences across the active learning and passive viewing tasks. This is perhaps surprising given that one might expect task requirements to affect the perceptual analysis of shape. Here, despite the fact that one

group of observers were explicitly told to memorize shape for a subsequent recognition task, the perceptual analysis strategies of the two groups, as evidenced by the patterns of data–model correspondences, were similar. One implication of this finding is that local shape analysis strategies during perception are "hard-wired" in the sense of being invariant to task requirements at least across the range of tasks tested here. This hypothesis is intuitively appealing in that during everyday recognition observers cannot entirely predict when unfamiliar objects might become relevant to their immediate or future goals and intentions. However, it remains to be determined whether the observed patterns of shape analyses found here will generalize across other tasks, including, for example, those related to the computation of shape representations for reaching and grasping (e.g., Land et al., 1999).

Chapter 5

5.1. Experiment 2

Experiment 2 builds on the previous experiment and aims to examine the robustness of fixational eye movement patterns across different sets of stimuli and methods for collecting the comparable model data. Here the generated model data were collected from trained observers in order to incorporate the same amount of error as in the recognition test task, thus allowing us to compare task acquisition and performance while taking into account human visual system characteristics. Similar to Experiment 1, observers actively memorised or passively viewed sets of novel 3D objects each comprising of four components or volumetric parts varying in local part structure and spatial configuration. Participants then performed a recognition memory test in which they discriminated the previously viewed targets from visually similar distracters at both trained and novel orientations. In both phases, we recorded the eye movements, response times (RTs) and accuracy measures. To examine in more detail the robustness of Experiment 1 findings we asked four additional questions: First, is any specific local image region fixated during initial viewing and subsequent recognition? Second, are the same image regions consistently fixated across changes in object viewpoint? Third, what kinds of local shape information do these regions contain? Fourth, are the patterns of fixations associated with local shape analyses robust across task demands (i.e., active learning versus passive viewing).

The goal of this study was not only to determine where observers fixate during shape perception and recognition, but also to examine what they fixate by undertaking detailed analyses of shape information content at fixated image regions. To do this, we contrasted the observed fixation patterns against the predicted distributions derived

from three theoretical hypotheses about local image information content. Model 1 was based on visual saliency which assumes a processing (or attentional) bias determined by local low-level image statistics at salient regions defined by contrasts in luminance intensity, orientation and colour (Itti, Koch & Niebur, 1998; Koch & Ullman, 1985; Walther & Koch, 2006). This model has been widely used in eye movement studies of scene perception although its efficiency in predicting fixation patterns remains the subject of on-going debate (e.g., Baddeley & Tatler, 2006; Henderson, Brockmole, Castelhano & Mack, 2007). Similarly to Experiment visual saliency was used as a baseline as I wanted to examine fixation patterns beyond low-level image statistics. Model 2 and Model 3 (explained in more detail below) derived from the large body of work highlighting the importance of contour curvature magnitude and the sign of curvature in visual perception (e.g., Attneave, 1954; Bertamini, 2008; Biederman, 1987; Cate & Behrmann, 2010; Cohen & Singh, 2007; De Winter & Wagemans, 2006; Feldman & Singh, 2005; Hoffman & Richards, 1984; Hoffman & Singh, 1997).

In the current study the focus of interest was to examine the extent to which local shape analyses, as shown by fixation patterns, can be predicted by the presence of local curvature (convex maxima and concave minima) in 3D object recognition. Previous empirical work demonstrated that concave and convex curvature plays a key role in shape perception, but to my knowledge no one yet have examined these two types of curvature with eye movement analyses and/or 3D images. Hence the main focus of interest is to investigate concave and convex models rather than visual saliency.

5.1.1. Method

Participants

60 students from Bangor University (51 female, mean age: 22.26 years, *SD* = 6.58, 54 right handed) participated in the study for course credit. All participants had normal or corrected to normal visual acuity. Informed consent was obtained from each participant prior to testing, in line with local ethics committee and BPS guidelines.

Stimuli

Each of the ten novel objects (see Figure 24) consisted of a unique spatial configuration of four volumetric parts. The parts were uniquely defined by variation among non-accidental properties (NAPs) comprising: Edges (Straight vs. Curved), symmetry of the cross section, tapering (collinearity) and aspect ratio (Biederman, 1987).

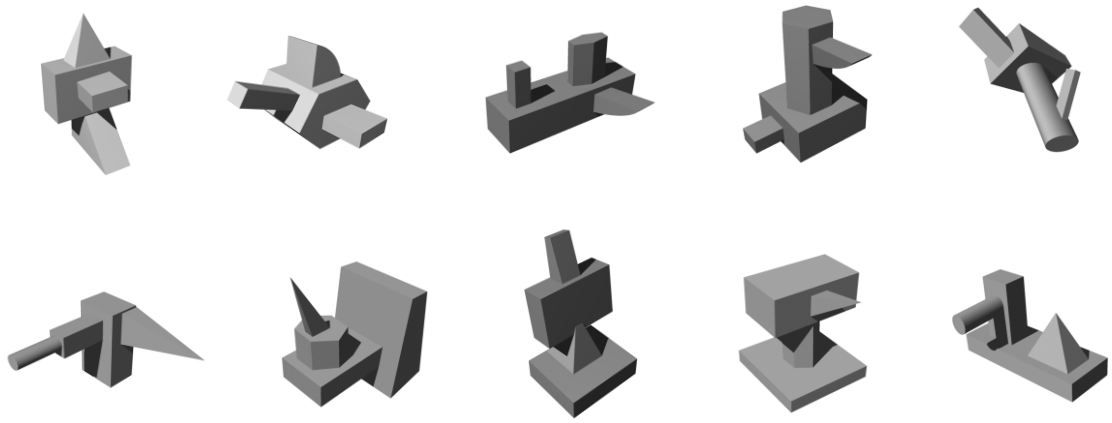


Figure 24 The 10 surface rendered novel object stimuli used in the Experiment.

These object models were produced using Strata 3D CX software (Strata, USA) and rendered using a single light source (top left) model in greyscale at 72 dpi

with anti-aliasing and scaled to fit within an 800 x 800 pixel frame. Ground shadow was removed from the images. Stimuli subtended 18 degrees of visual angle horizontally from a viewing distance of 60 cm. This scale was chosen to induce saccadic exploration over the stimuli. Versions of each model were created depicting the object from each of six different viewpoints at successive 60 degree rotations in depth around a vertical axis perpendicular to the line of sight. The zero degree viewpoint was a 'canonical' three-quarter view (see Figure 25). The 0, 120 and 240 degree versions served as training viewpoints, and the 60, 180 and 300 degree versions as novel test viewpoints.

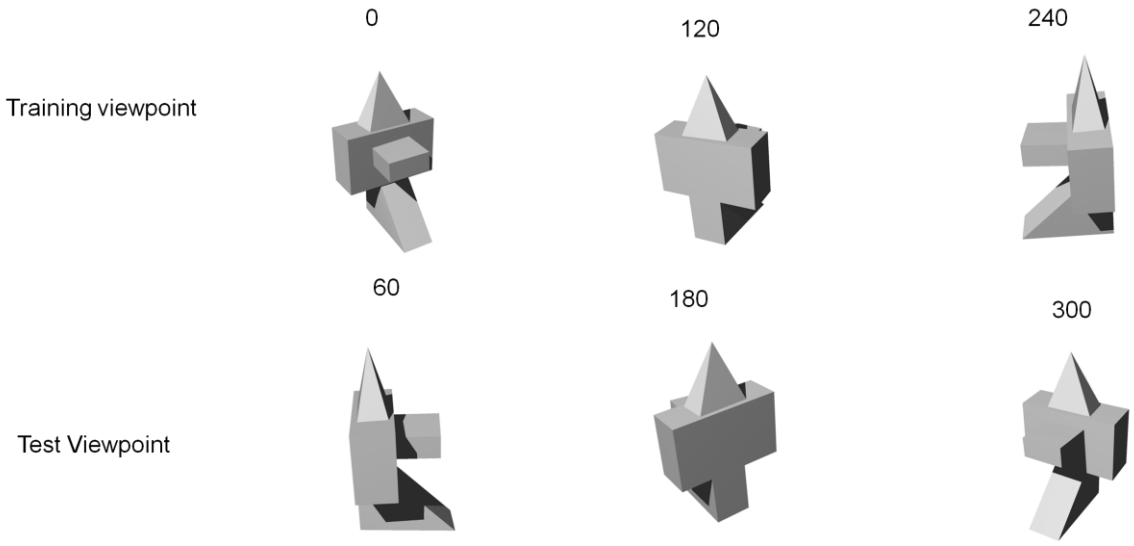


Figure 25 An illustration of the three trained and three novel viewpoints used. In the learning phase each stimulus was shown at each of the three trained viewpoints. In the test phase, targets and non-targets were each shown at all six viewpoints.

Apparatus

A Tobii 1750 eye tracking system was used to record eye-movement data. This apparatus allows for high precision binocular tracking with 0.5 degrees accuracy, 0.25 degrees spatial resolution, and drift < 1 degree. Stimuli were presented on a TFT

monitor running at a resolution of 1280 x 1024 pixels and 60 Hz refresh rate. A chin rest was used to stabilize the participant's head at a 60cm viewing distance and a standard USB keyboard was used for response collection.

Design and Procedure

The design and the procedure were exactly the same as in Experiment 1.

5.1.2. Generated model predictions

Empirically-defined AOIs generated using FROA were compared to predicted distributions of fixation regions. The predicted distributions were computed as follows:

Model 1: Visual Saliency baseline

The first model tested the visual saliency hypothesis (Itti, Koch & Niebur, 1998) using the Saliency Toolbox implementation in Matlab (Walther & Koch, 2006). The model was run on each of the 60 stimulus images (10 objects x 6 viewpoints) used in the recognition task to generate a saliency map for each stimulus. The output of the toolbox is a list of saliency values for each pixel which are grouped into a saliency region map using shape estimation function (see Walther & Koch, 2006). The number of saliency regions generated was constrained to approximate the area and number of thresholded regions generated for the other models: Mean pixel area across saliency maps per object was 21090 (*SE* = 622.70). The saliency maps were thresholded and binarised using FROA in the same way as the empirically derived fixation data from the recognition task. These maps represent the thresholded distributions of fixation regions we would expect if eye movements were determined solely by low-level

image statistics, that is, by the most visually salient image regions defined by colour, intensity contrast, and orientation.

Similarly to Experiment 1, this model served as a baseline contrast given the question of interest here was whether specific models of shape analyses could account for fixation patterns beyond the explicable by visual saliency.

Model 2: Convex Surface Curvature Maxima

The second model generated predicted fixation regions based on the locations of local features defined by convex surface curvature maxima. In order to generate predicted region maps that incorporate the same error measures as the recognition task data (that is, variation in fixation patterns due to both within and between-subject variability, as well as error arising from eye tracker accuracy, drift and resolution) we used a trained observer technique (Johnston & Leek, 2009). Thirteen participants (11 right handed, *M* = 22.54 years, *SD* = 8.00; range = 19-41 years) were trained to fixate only at convex areas of the 10 experimental stimuli, each from the same six viewpoints used in the recognition task. Stimulus exposure duration was 10 seconds as in the learning phase of the recognition task. Fixation region maps were generated using FROA by applying the same filtering, Gaussian smoothing and thresholding criteria as used for the recognition task data (see above). Mean pixel area across thresholded convexity maps per object was 20457.53 (*SE* = 447.39).

Model 3: Concave Surface Curvature Minima

The third model generated predicted fixation regions based on the locations of local features defined by concave surface curvature minima. Thirteen participants (13 right handed, *M* = 26.54 years, *SD* = 9.40; range = 18-40 years) were trained to fixate only at convex areas of the 10 experimental stimuli, each from the same six viewpoints used in the recognition task. Stimulus exposure duration was 10 seconds as

in the learning phase of the recognition task. Fixation region maps were generated using FROA by applying the same filtering, Gaussian smoothing and thresholding criteria as used for the recognition task data (see above). Mean pixel area across thresholded convexity maps per object was 16695.17 (*SE* = 526.63).

Independence of predicted fixation patterns

It is important to verify that the predicted region distributions of the three models are sufficiently different (in order that they may be statistically distinguished when compared to the gaze data) (see Figure 26). In order to do this FROA was used to compare region overlap across models.

Mean pixel region overlap across items for the Visual saliency and Convexity models was 12.98% (*SD* = 9.32%) of the total pixel area for the convexity model. Analyses of these data using FROA showed that MMC score for the observed region overlap between the two models was -0.298.

Mean pixel region overlap across items for the Visual saliency and Concavity models was 7.31% (*SD* = 8.83%) of the total pixel area for the Concavity model. Analyses of these data using FROA showed that MMC score for the observed region overlap between the two models was -0.940.

Mean pixel overlap between the Convexity and Concavity models was 21.90 % (*SD* = 21.07%) of the total pixel area for the Convexity model. Analyses of these data using FROA showed that the MMC score for the observed region overlap between the two models was 6.99.

A one way ANOVA (target vs. between models overlap: Visual saliency vs. Concave, Visual saliency vs. Convex, Convex vs. Concave) on the distance measure showed a significant main effect of Model, F (2, 179) = 17.90, $p < .0001$. Post-hoc analyses showed that the pairwise contrasts between models were significantly

different for Visual saliency vs. Concave, $p < .0001$; Visual saliency vs. Convex, $p < .0001$, and Concave vs. Convex, $p = .018$.

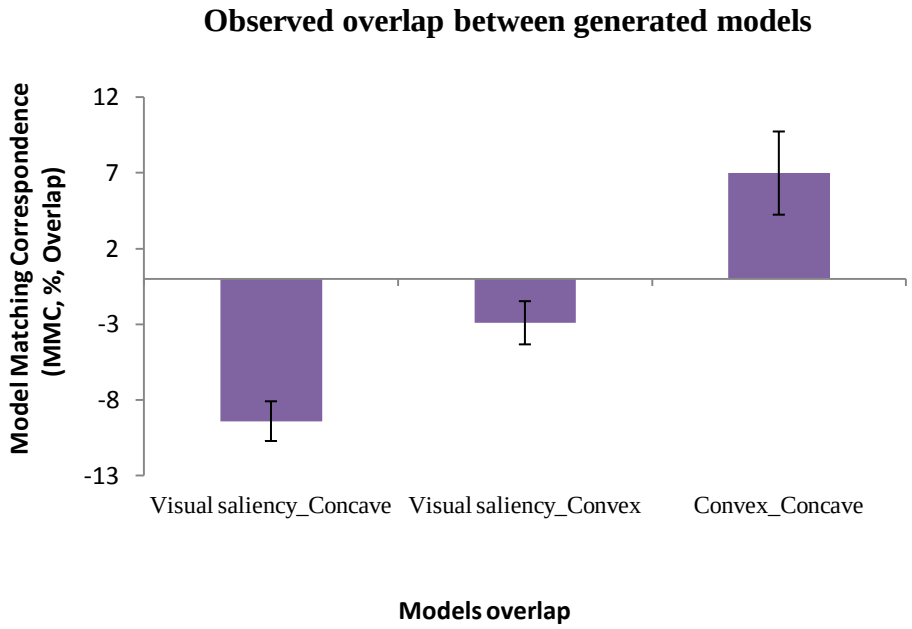


Figure 26 The mean frequency of observed overlap (expressed in MMC) between the generated model data. Bars show standard error of the men (% overlap).

This shows that the pattern overlap between all of the generated models (Visual saliency, Convexity and Concavity) is significantly different; thus, we have significantly distinct empirical fixation models distributions to compare with the experimental data.

5.1.3. Behavioural data analyses

Analyses were conducted on the mean median test phase RTs and accuracy data. Only RTs for correct responses were included.

Table 4. The mean median RTs and accuracy rates (targets) for familiar and novel viewpoints in the Test phase. Standard error of the mean is shown in parentheses.

	Pre-Test Group			
	Active Learning		**Passive Viewing**	
	RTs (ms)	% Correct	RTs (ms)	% Correct
Familiar Views	1538.29 (66.29)	83 (2.4)	1589.60 (76.83)	74 (2.7)
Novel Views	1847.34 (128.54)	74 (3.2)	1623.34 (86.00)	67 (4.2)

Reaction time and Accuracy data

A 2 (Pre-test task: Active learning vs. Passive viewing) x 2 (Viewpoint: Familiar vs. Novel) x 2 (Stimulus type: Target vs. non-target) mixed factor ANOVA for the reaction time data showed no significant main effects or interactions. Overall accuracy rates for the Active Learning group (M = 78.5%, SE = 2.80) and Passive Viewing group *(*M = 70.5%, SE = 3.40) were significantly different Mann Whitney: z = - 4.20, p <.0001. Given the lack of significant main effects or interaction for the RT data, between both groups, the rest of the analyses are reported per group.

Analyses of behavioural data for Active learning group

Overall response accuracy in the test phase was high for both targets (M = 78%; SE = 2%) and non-targets (M = 91%; SE = 1.10%). This difference was statistically significant, (Wilcoxon), z = .422, p <.0001. For test phase target trials, accuracy was higher for trained views (M = 83%, SE = 2.4%) than for novel views (M = 74%; SE = 3.2%), but not significantly different (Wilcoxon), z = -1.877, p =.060, ns. There was no significant difference in accuracy for non-targets across the two groups of viewpoints. Mean Median RTs (correct responses only) were calculated per condition for targets and non-targets in the test phase. A 2 (Trained view: 0˚, 120˚,

240˚ vs. Novel view 60˚, 180˚, 300˚) x 2 (Stimulus: Target vs. Non-target) repeated measures ANOVA showed a significant interaction between View and Stimulus, F (1, 27) = 6.190, p = .019, η_p^2 = .186. Mean Median RTs were significantly faster for targets at familiar (trained) viewpoints (M = 1538.29 ms; SE = 66.29 ms) than at novel viewpoints (M = 1847.34 ms; SE = 128.54 ms), t (27) = -2.05, p = .050, R^2 = 0.37. For non-targets there was no significant difference between mean RTs for the 0˚, 120˚ and 240˚ views (M = 1774.55 ms; SE = 154.20 ms) and the 60˚, 180˚ and 300˚ views (M = 1621.52 ms; SE = 108.44 ms*); t* (27) = 1.22, p = .230, ns.

Analyses of behavioural data for Passive viewing group

Overall response accuracy in the test phase was high for both targets (M = 70%; SE = 2.5%) and non-targets (M = 81%; SE = 2%). This difference was statistically significant Wilcoxon, z = 3.021, p = .003. For test phase target trials, accuracy was higher for familiar views (M = 74%, SE = 2.7%) than for novel views (M = 67%; SE = 4.2%), but not significantly different (Wilcoxon) z = -1.463, p = .144, ns. There was no significant difference in accuracy for non-targets across the two groups of viewpoints. Mean median RTs (correct responses only) were calculated per condition for targets and non-targets in the test phase. A 2 (Familiar view: 0˚, 120˚, 240˚ vs. Novel view 60˚, 180˚, 300˚) x 2 (Stimulus: Target vs. Non-target) repeated measures ANOVA showed no significant main effects or interaction. Mean median RTs were faster for targets at familiar viewpoints (M = 1589.60 ms; SE = 76.83 ms) than at novel viewpoints (M = 1623.34 ms; SE = 86.00 ms), but not significantly different, t (27) = -.304, p = .763ns. For non-targets there was no significant difference between mean RTs for the 0˚, 120˚ and 240˚ views (M = 1650.40 ms; SE = 78.59 ms) and the 60˚, 180˚ and 300˚ views (M = 1570.10 ms; SE = 81.96 ms); t (9) = 1.327, p = .196, ns.

5.1.4. Fixation data analyses

Active Learning group

Using FROA, distributions of fixation regions for each target and non-target were empirically derived from the fixation data (see Methods). A subject analysis was first performed to test the generality and reliability of the thresholded fixation region distributions across participants. This was done by contrasting the frequency of fixations between thresholded and sub-thresholded regions (see Methods). The frequency statistics for thresholded (fixated) and sub-thresholded regions were normalized for mean region size and express frequency in units of visual angle. Separate subject analyses were performed on the learning (targets) and test phase (targets and non-targets) data. Table 5 shows the mean normalized frequencies for the thresholded and sub-thresholded regions across participants. These data show that the mean normalized fixation frequency for thresholded regions is higher than for sub-thresholded regions in both the learning and test phases for targets, and in the test phase, for targets and non-targets.

Table 5. The mean normalized fixation frequencies (mean fixation per degree of visual angle) for thresholded and sub-thresholded regions derived using FROA. Standard error of the mean is shown in parentheses.

	Pre-test Phase	**Test Phase**	
	Targets	Targets	Non-targets
Thresholded regions	0.38 (0.02)	0.31 (0.03)	0.30 (0.036)
Sub-thresholded regions	0.0001 (0.0001)	0.002 (0.002)	0.002 (0.002)

For the Active learning phase, there was a significant difference between the mean normalized fixation frequencies across participants for the thresholded vs. sub-

thresholded regions, t (29) = 12.94, $p < .0001$, $R^2 = 0.92$. For the test phase, a 2 (Region: Thresholded vs. sub-thresholded) x 2 (Stimulus: Target vs. non-target) repeated measures ANOVA showed a significant main effect of Region, F (1, 29) = 85.45, $p < .0001$, $\eta_p^2 = .747$, but no other effects or an interaction.

Passive viewing group

Table 6. The mean normalized fixation frequencies (mean fixation per degree of visual angle) for thresholded and sub-thresholded regions derived using FROA. Standard error of the mean is shown in parentheses.

	Pre-test Phase	**Test Phase**	
	Targets	Targets	Non-targets
Thresholded regions	0.30 (0.02)	0.25 (0.03)	0.25 (0.029)
Sub-thresholded regions	0.0001 (0.0001)	0.002 (0.002)	0.002 (0.001)

For the pre-test task, there was a significant difference between the mean normalized fixation frequencies across participants for the thresholded vs. sub-thresholded regions, t (29) = 12.32, $p < .0001$, $R^2 = 0.92$ (see Table 6). For the test phase, a 2 (Region: Thresholded vs. sub-thresholded) x 2 (Stimulus: Target vs. non-target) repeated measures ANOVA showed a significant main effect of Region, F (1, 29) = 60.05, $p < .0001$, $\eta_p^2 = .674$, but no other effects or an interaction.

The subject analyses show that the distribution of fixation regions identified using FROA is robust across participants in both pre-test and test phases of the study, and across targets and non-targets. All remaining analyses of the fixation frequency data were computed by items.

Analyses of the local shape feature analysis patterns for Active Learning group

In order to elucidate the information content at fixated image regions FROA was used to determine the degree of overlap in the observed spatial distributions of fixation regions and those predicted by each tested model of local shape analysis. Separate analyses are presented for the learning and test phases. Three of the stimuli used were excluded from the analyses, as naming errors were found which potentially could have confounded the analyses.

Active Learning task

For the learning phase, the distributions of fixation regions to targets presented at trained viewpoints (N=29) were analysed across 3 epochs following stimulus onset (see Figure 27).

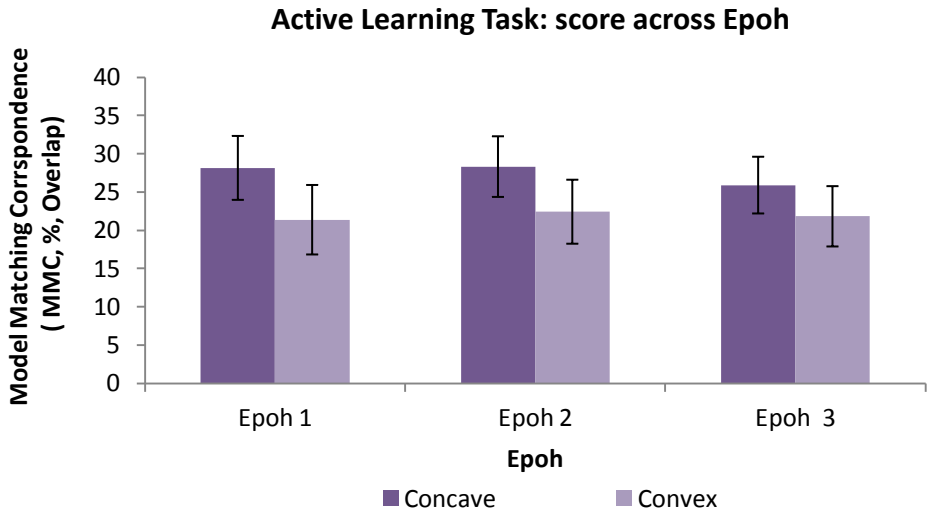


Figure 27 Mean MMC $_{Mx}$ measure during Active Learning Task: The frequency of significant contrasts per epoch for concave and convex models relative to visual saliency. Bars show standard error of the mean (% overlap).

A 3 (Epoch) x 3 (Model) repeated measures ANOVA on the MMC distance measure across targets showed a significant main effect of Model, F (2, 56) = 18.73, p < .0001, η_p^2 = .401, but no main effect of Epoch, F = (2, 56) = 2.07, p = .136ns, and no significant interaction. In the absence of an interaction, the MMC distance statistics were collapsed across epoch (see Figure 28). A one-way ANOVA across models on the MMC measure was significant, F (2, 86) = 15.44, p < .0001. Post-hoc analyses showed that the pairwise contrasts between models were significantly different for Concave vs. Visual saliency, p < .0001; Convex vs. Visual saliency, p < .0001; but not for Concave vs. Convex, p = .295ns.

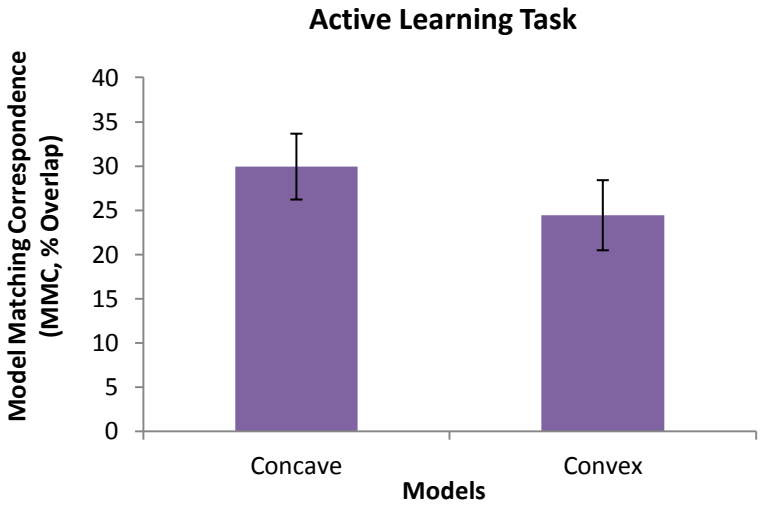


Figure 28 Mean MMC $_{Mx}$ measure of data-model correspondences for pre-test Active learning task collapsed across epoch for concave and convex models relative to visual saliency. Bars show standard error of the mean (% overlap).

Test phase

Analyses of the local shape feature patterns (Active learning group vs. Generated models)

A 2 (Stimulus: Target, Non-target) x 3 (Model) repeated measures ANOVA on the distance measure showed a significant main effect of Model, F (2, 112) = 47.50, p < .0001, η_p^2 = .459, but no other significant main effect or interaction (see Figure 29). In the absence of an interaction, the MMC distance statistics were collapsed across targets and non-targets (see Figure 30). A one-way ANOVA across models on the MMC measure was significant, F (2, 341) = 36.13, p < .0001. Post-hoc analyses showed that all of the pairwise contrasts between models were significantly different, Concave vs. Visual saliency, p < .0001; Convex vs. Visual saliency, p < .0001; Concave vs. Convex, p = .044. R^2 = 0.88.

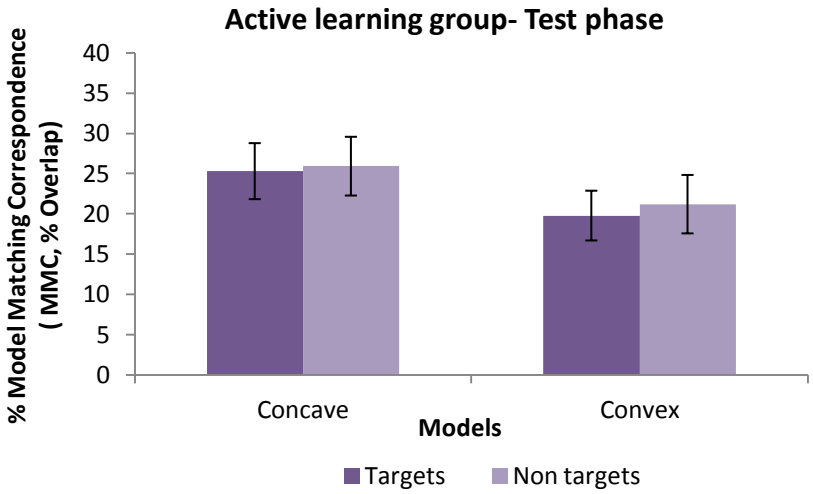


Figure 29 Mean MMC $_{Mx}$ measure of data-model correspondences for concave and convex models (relative to visual saliency) for the recognition memory test phase Bars show standard error of the mean (% overlap).

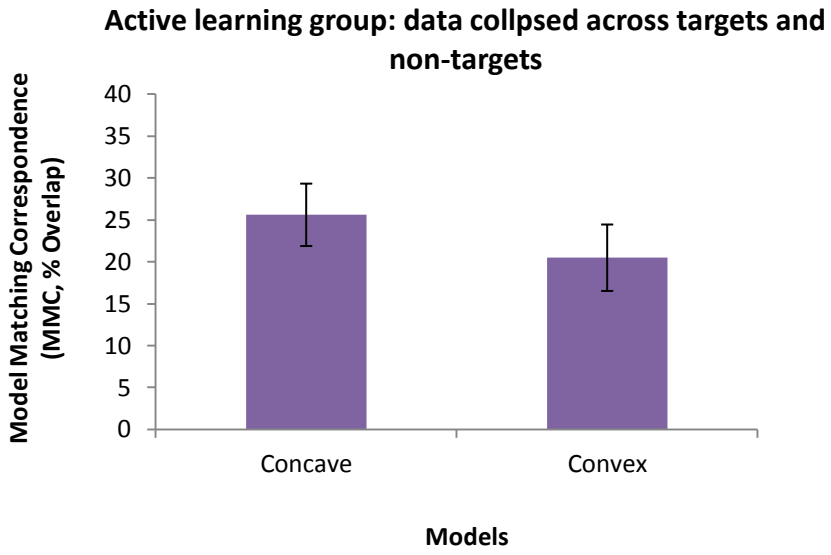


Figure 30 Mean MMC $_{Mx}$ measure of data-model correspondences for concave and convex models (relative to visual saliency) for the recognition memory test phase collapsed across targets and non-targets. Bars show standard error of the mean (% overlap).

Analyses of the local shape feature analysis patterns for Passive viewing task

For the learning phase, the distributions of fixation regions to targets presented at trained viewpoints (N=29) were analysed across 3 epochs following stimulus onset (see Figure 31).

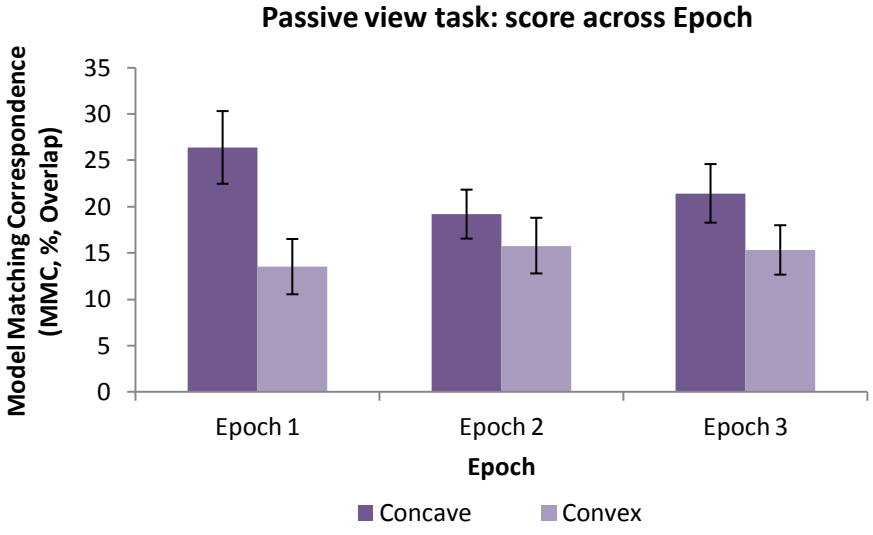


Figure 31 Mean MMC $_{Mx}$ measure of data-model correspondences for concave and convex models (relative to visual saliency) in pre-test Passive viewing phase across Epoch. Bars show standard error of the mean (% overlap).

A 3 (Epoch) x 3 (Model) repeated measures ANOVA on MMC distance measure across targets showed a significant main effect of Model, F (2, 56) = 18.42, p < .0001, η_p^2 = .397, and an interaction between Epoch and Model , F (4, 112) = 3.19, p = .016, η_p^2 = .102, but no main effect of Epoch, F = (2, 56) = .037, p = .963ns. Post-hoc analyses, of pairwise contrasts between models (within Epoch) were all significantly different for Epoch 1: Visual saliency vs. Convex, p = .002; Visual saliency vs. Concave, p < .0001; Convex vs. Concave, p =.007. For Epoch 2 and 3 the pairwise contrasts between models were significantly different for Visual saliency vs. Convex, p <.0001, and Visual saliency vs. Concave, p < .0001, but not for Convex vs. Concave, p =.343ns, and, p = .158ns, respectively.

Analyses of the local shape feature patterns (Passive view group vs. Generated models)

In a further analysis using FROA we contrasted passive view phase with the model region overlap (see Figure 32). A one way ANOVA (Target vs. Models: Visual saliency, Convex, Concave) on the distance measure showed a significant main effect of Model, F (2, 86) = 18.71, p < .0001. Post-hoc analyses showed that the pairwise contrasts between models were significantly different for Concave vs. Visual saliency, p < .0001; Convex vs. Visual saliency, p < .0001; but not for Concave vs. Convex, p = .123ns.

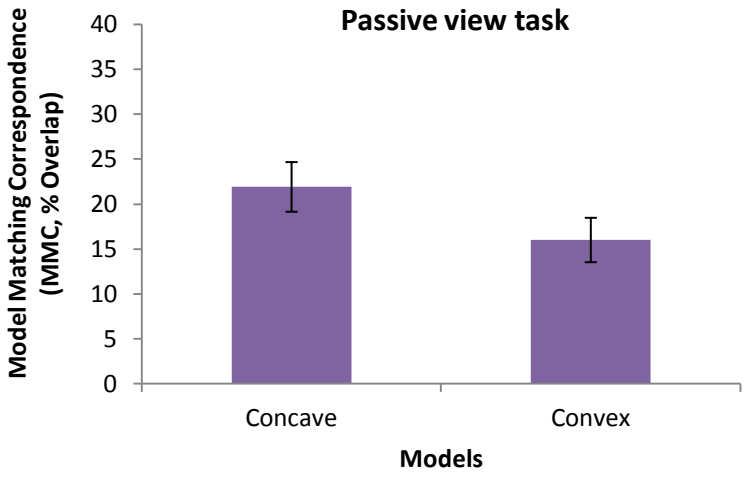


Figure 32 Mean MMC $_{Mx}$ measure of data-model correspondences for concave and concave (relative to visual saliency) models in pre-test Passive viewing phase collapsed across target and non-target. Bars show standard error of the mean (% overlap).

Test phase

Analyses of the local shape feature patterns (Test vs. Generated models)

A 2 (Stimulus: Target, Non-target) x 3 (Model) repeated measures ANOVA on the distance measure showed a significant main effect of Model, F (2, 112) = 27.54, p < .0001, η_p^2 = .330, but no other significant main effect or interaction (see Figure 33).

In the absence of an interaction, the MMC distance statistics were collapsed across targets and non-targets (see Figure 34). A one-way ANOVA across models on the MMC measure was significant, F (2, 341) = 20.60, p < .0001. Post-hoc analyses showed that pairwise contrasts between models were significantly different for, Concave vs. Visual saliency, p < .0001; Convex vs. Visual saliency, p < .0001; but not for Concave vs. Convex, p = .131ns.

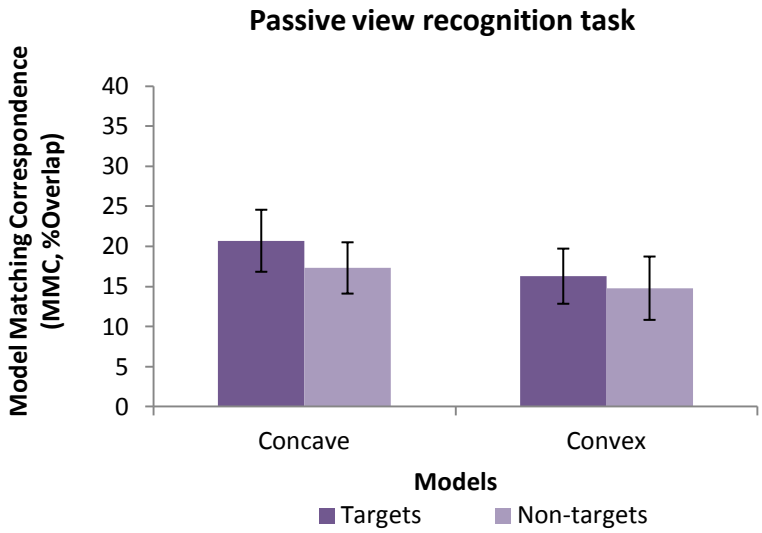


Figure 33 Mean MMC $_{Mx}$ measure of data-model correspondences for concave and convex models (relative to visual saliency) in pre-test Passive view recognition task for targets and non-targets. Bars show standard error of the mean (% overlap).

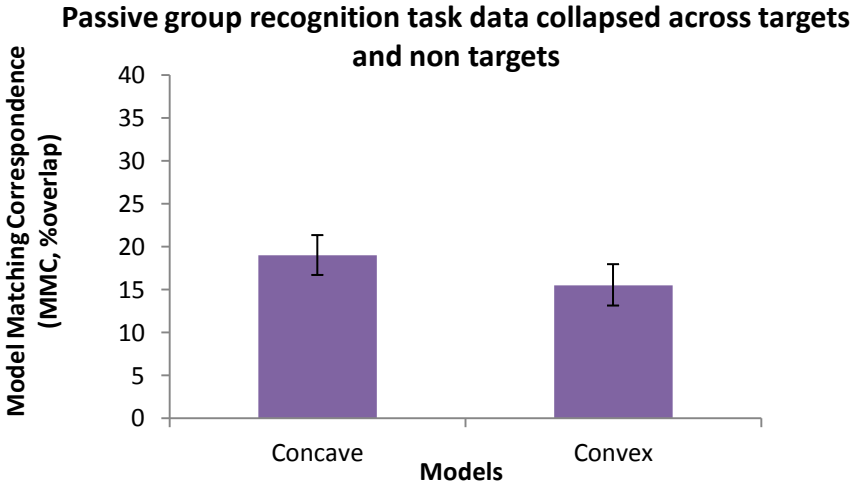


Figure 34 Mean MMC $_{Mx}$ measure of data-model correspondences for concave and convex models (relative to visual saliency) for Passive recognition task collapsed over targets and non-targets. Bars show standard error of the mean (% overlap).

Pre-test phase: Active Learning and Passive Viewing Groups

These analyses were run on the MMC data from the pre-test phase. A 2 (Active Learning, Passive viewing, *BS*) x 3 (Model: Concave, Convex, Visual saliency, *WS*) mixed ANOVA across target mean MMC for the pre-test phase data showed a

significant main effect of Model, F (2, 112) = 36.42, p < .0001, ηp^2 = .394, but no effect of Group, F (1, 56) = 1.10, p = .297 ns, and no interaction. Given the lack of interaction I have collapsed the pre-test phase data across group and run post hoc analyses across model means. Post-hoc comparison (t-test) was significantly different for Concave vs. Visual saliency, p <.0001, R^2 = 0.72 Convex vs. Visual saliency, p <.0001, R^2 = 0.68, but not Concave vs. Convex, p =.075ns.

Test phase: Active Learning and Passive Viewing Groups

These analyses were run on the MMC data from the test phase. A 2 (Active Learning, Passive viewing, *BS*) x 2 (Phase: Trained vs. Novel, *WS*) x 3 (Model: Concave, Convex, Visual saliency, *WS*) mixed ANOVA across target mean RTs for the test phase data. This showed a significant main effect of Model, F (2, 112) = 36.43, p < .0001, η_p^2 = .394, but no effect of Task, F (1, 56) = 1.20, p = .277 ns, and no interaction. Given the lack of interaction I have collapsed the test phase data across targets and non-targets and run a one way ANOVA across target mean RTs (see Figure 35). A one way ANOVA (Target vs. Model: Visual saliency, Convex, Concave) on the distance measure showed a significant main effect of Model, F (2,170) = 35.05, p <.0001. Post-hoc comparison (t-test) was significantly different Convex vs. Concave, p =.033, R^2 = 0.72, Concave vs. Visual Saliency, p <.0001, Convex vs. Visual Saliency, p <.0001.

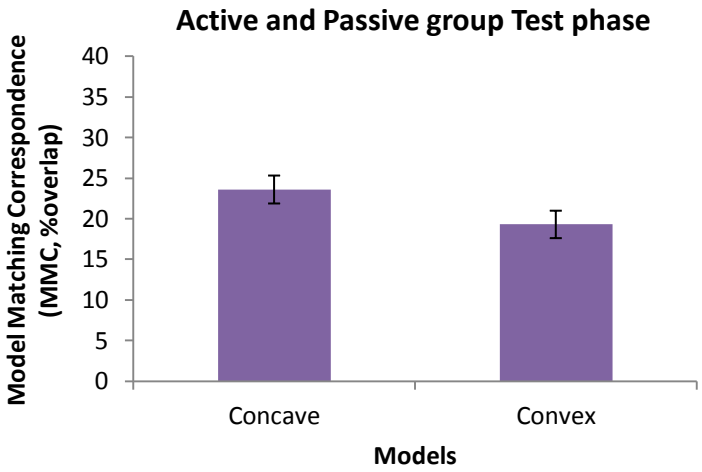


Figure 35 Mean MMC $_{Mx}$ measure of data-model correspondences for concave and convex models (relative to visual saliency) for the recognition memory test phase (collapsed across Active and Passive groups and across targets and non-targets). Bars show standard error of the mean (% overlap).

Summary of the Results

Behavioural data

- The overall RTs and Accuracy data in the test phase showed faster and more accurate responses for the Active learning group, than the Passive viewing pre-test group.
- The Active pre-test group showed faster RTs for targets at familiar than novel viewpoints.
- Both Active learning and Passive viewing pre-test groups showed significantly higher accuracy for non-targets than targets.

Eye movement data

- Eye movement analyses showed higher fixation frequency for thresholded AOIs than for sub-thresholded AOIs for active learning, passive viewing phase for targets and non-targets.

- Analyses of local shape feature patterns for both groups in the pre-test phase showed main effect of model. Post-hoc analyses for Data-model correspondences were greater for Concave and Convex models relative to Visual saliency.
- Analyses of local shape feature patterns for both groups in the test phase showed that data-model correspondences were larger for the Concave model, followed by the Convex model, and the Visual saliency model.

5.1.5. Conclusions

In this experiment observers memorised a sub-set of 3D novel objects, and then performed a recognition memory test phase in which targets were discriminated from visually similar distracters across trained and novel viewpoints. Similarly to Experiment 1, the analyses of RT and accuracy data showed that observers were more accurate following the active learning than the passive viewing pre-test, and in the test phase RTs for both groups were faster for targets at familiar (pre-test) viewpoint than at novel viewpoint. This finding suggests that the participants in the active learning and passive viewing pre-tests performed the recognition memory task in a similar way and that the recognition in both groups was viewpoint dependant. This finding supports previous research suggesting that recognition is mediated by a viewpoint-dependent represenatations of shape (e.g., Bulthoff & Edelman, 1992; Reisenhuber & Poggio, 1999). The analyses of eye movement data were best accounted for by models of shape analysis based on local regions of curvature extrema and there was no evidence that fixation distributions are determined by low-level visual saliency. Instead, the observers showed a strong preference for fixation at regions of concave curvature minima relative to convex curvature maxima, which interacted in the pre-test phase. More specifically, the observers showed no preference for fixating concave over convex regions during the passive viewing task, whereas in the active learning

task the observers predominantly fixated concave regions. Although this pattern of results seems to be somehow intuitive in respect to the imposed task differences, it provides a dissimilar pattern of results with Experiment 1 where concave regions were preferentially fixated during both pre-tasks, which we interpreted as evidence for 'hard-wired' mechanisms of 3D object shape representation. In this context, there may be a number of possible interpretations of these results. For example, local convex and concave curvature may both be processed to some extent during passive viewing (e.g. no task in hand), whereas concave curvature regions are processed more during an active learning phase when an object shape representation is computed and stored in memory. The apparent difference to the Experiment 1 results could very likely be a product of the models we have used for data-model comparison (e.g. algorithmically based, vs. trained observer data incorporating the same error measures as the recognition task).

The main empirical findings were as follows: (1) In the recognition memory task we found a strong viewpoint-dependent pattern of identification latencies consistent with other previously reported studies supporting the use of viewpoint-dependent object representations (e.g., Bülthoff & Edelman, 1992; Edelman & Weinshall, 1991; Tarr & Bülthoff, 1998). (2) The fixation patterns in both pre-test and test phases of the study were not well accounted for by low-level visual saliency as implemented in the Itti, Koch and Neibur (1998) model. As a matter of fact, visual saliency performed no better than a random model of fixation region distribution. (3) The fixation distributions were best modelled in terms of local shape analyses at regions of curvature extrema corresponding to concave or convex surface discontinuities. (4) While in the passive view phase there was no significant difference in the spatial distributions of data-model correspondences between the convex and

concave surface discontinuity models, observers in the active learning phase and recognition memory test phase showed a fixation bias for regions of concave surface discontinuity.

These findings support a large body of work in the psychophysics literature concerning the importance of surface curvature extrema in visual object recognition (e.g., Cohen & Singh, 2007; Cohen et al., 2005; De Winter & Wagemans, 2006; Feldman & Singh, 2005; Hoffman & Richards, 1984). Of particular interest here is that, unlike many previous studies that have reported perceptual biases for convex and/or concave contour curvature in 2D outline forms, we report a processing bias revealed through fixation patterns determined by surface curvature extrema in 3D forms. However, the data did reveal a statistically reliable fixation bias for concave surface discontinuities in the active learning phase and the test phase of the study. This supports the hypothesis that local regions of surface concavity play an important role in the indexing, encoding and /or matching of perceptual input to stored object shape representations.

Furthermore, our analyses showed that similar local image regions were fixated during the active learning and test phases, and that observers tend to fixate regions of concave surface discontinuities across changes in stimulus viewpoint – which both support the hypothesis that these local image regions are inherently linked to object recognition.

Chapter 6

6.1. Experiment 3

One fundamental question in human vision is how expertise with objects influences shape representation. Previous studies have shown that experience with objects can result in qualitative behavioural changes (e.g., Wong, Palmery & Gauthier, 2009) as well as potential changes in cortical representations (e.g., Gauthier & Tarr, 1997; Downing, Jiang, Shuman & Kanwisher, 2001; Wong, Palmeri, Rogers, Gore & Gauthier, 2009).

A number of studies suggest that object categorisation is a hierarchically organised process (e.g., Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976; Bulthoff, Edelman & Tarr, 1995) with links between the superordinate level (e.g., Animal), with basic level (e.g., Cat), to the subordinate level (e.g., Siamese). However, these levels of categorisation are seen as distinct processes relying on different perceptual information. For example, the superordinate level of object categorisation is suggested to incorporate more functional and abstract information (Tversky & Hemenway, 1984; Tversky, 1989), whereas basic and subordinate levels rely more on perceptual information.

More specifically, basic level categorisation is suggested to rely more on configurable information between objects parts and their components (Biederman, 1987; Tversky & Hemenway, 1984), and is referred to as the preferred 'entry access' of object classification. For example, various behavioural studies have shown that naming performance is fastest at a basic level of categorisation (e.g., Rosch et al., 1976; Jolicoeur, Gluck, & Kosslyn, 1984), although Tanaka and Taylor (1991) provide evidence that category expertise improves the speed of access at subordinate level,

equating the response time to that of basic level access. In their study, dog and bird experts participated in a category-verification task, where after hearing a category label, their subjects (dog and bird experts) first heard a category label (superordinate, basic, or subordinate) and were asked to specify whether the subsequently presented picture was one of the previously labelled category. The dog and bird experts were equally fast at subordinate and basic categorisation.

In one hypothesis; basic level of categorisation and object detection have been presumed to be supported by the same perceptual mechanisms as there was no significant difference in response time and accuracy between object detection and object categorisation processes (e.g., as soon as participants could detect an object, they already knew the category of that object) (Grill-Spector & Kanwisher, 2005). This suggestion was also supported by a functional magnetic resonance imaging (fMRI) study, which showed that the same cortical regions are activated during detection and identification of stimuli of a certain category (Grill-Spector, 2003).

However, more recent research (Mack & Palmeri, 2010) demonstrated that effective object detection is possible without categorising that object at basic-level, as well as that object detection is faster than a basic level of categorisation. In contrast to basic level categorisation, subordinate level categorisation is suggested to rely on finer distinctions in order to discriminate between two individual exemplars of one object class (e.g. recognising individual faces) (Bulthoff et al., 1995). Learning to classify visually similar objects at a subordinate level typically involves identifying small perceptual changes, thus making this task appropriate for exploring the role of local features in object shape representation. A wide range of studies (Bukach, Gauthier & Tarr, 2006; Gauthier & Tarr, 2002) have presented evidence that subordinate expertise with objects influences their perceptual representations from feature-based in novices

to holistic (i.e., integral, attending to all parts of an object) in experts. Moreover, expertise with objects seems to change processing strategies by redirecting attention from feature analyses to more configurable processing and computation of spatial relations between features.

Commonly, studies investigating object shape categorisation have compared different stimulus classes such as faces and non-face objects (e.g., Kanwisher, Chun, McDermott & Ledden, 1996; Kanwisher, McDermott & Chun, 1997). The rationale behind using these stimuli is that faces evoke automatic subordinate processing, whereas objects are typically processed at the basic level. However, faces and non-face objects are qualitative different in a number of dimensions including shared part configuration, social importance, number of parts, number of familiar exemplars, along with the participants' expertise. Therefore, comparing these sets of stimuli on any single dimension cannot verify that the sets do not differ on some other dimension (Gautier, Anderson, Tarr, Skudlarski & Gore, 1997).

Downing et al. (2001) have provided evidence that specific object categories activate different areas of the cortex. For example, the authors highlighted a region in human lateral occipitotemporal cortex that responds predominantly to images of human bodies and body parts, but not faces, or object parts, and this was interpreted as evidence that there is an expert system for processing the visual appearance of different object classes such as the human body. However, these differences could be the result of a number of confounding factors, such as shape dissimilarity, name, and history of experience with the object.

A more recent neuroimaging study (Wong, Palmeri, Rogers, Gore & Gauthier, 2009) found category selective patterns of activation in the cortex while using shape controlled novel object stimuli. More specifically, the authors demonstrated that

learning to individuate novel objects at a subordinate level increased activity for these objects in the right fusiform region. In contrast, learning to categorise the same objects at the basic level resulted in increased activation in the medial ventral occipito temporal cortex (VOT) relative to lateral parts. In another behavioural study Wong, Palmeri and Gautier (2009) showed that expertise in both basic and subordinate level of novel objects resulted in a selective improvement as a function of training. The authors examined whether manipulating learning history for the objects, whilst holding the object shape constant, would result in qualitatively different behavioural changes. For example, the 'expertise hypothesis' (Gauthier & Tarr, 1997) suggests that expertise at subordinate level of classification within a visually similar category is the main cause of participants relying more on configural information and developing an additional holistic processing strategy. Moreover, according to this theory object expertise that does not involve a subordinate level of classification should not produce holistic processing.

In experiment by Wong et al. (2009), all of the participants completed a sequential matching task in pre-test and post-test where they had to judge if two sequentially presented novel objects called 'Ziggerins' were the same or different individuals or whether they belong to the same or different family. In addition, in post-test the participants completed a composite task which is traditionally used to measure holistic processing and its dependence on configuration (For more details, see Cheung, Richler, Palmeri & Gauthier, 2008; Gauthier, Curran, Curby & Collins 2003) and triplet recognition task in order to maximize the difference between the training groups. The control group took part in the composite task in order to obtain a base - line measure. The participants were randomly assigned to three groups, two of the groups were trained with the same set of novel objects in two different ways, and a

third (control) group received no training. One of the training groups learned to categorise the objects at the subordinate level, whereas the other training group learned to categorise the same set of objects at basic level. More specifically, one of the groups learned the individual names of 18 (out of 36) objects called 'Ziggerins', while the other 18 objects were used as distracters. The other group learned to categorise the set of 36 'Ziggerins' into six families. The family and individual names assigned for objects were two-syllable nonsense words (e.g., xedo, kimo). The 'Ziggerins' were introduced progressively from session 1 through session 3; and all the 36 'Ziggerins' were present from session 4 to session 10. The results showed a selective increase in holistic processing (defined by the level of sensitivity to part configuration and the congruency effect to an aligned configuration of parts) for the subordinate level training group with speeded the response times as a function of training. In contrast, the basic level training group was faster than the subordinate training group in basic level recognition after training. The results of the above two studies provide evidence that learning to categorise objects not only affects perceptual strategies and leads to behavioural changes (Wong et al., 2009) but also results in qualitative differences in neural activity of the visual cortex (Wong, Palmeri, Rogers, Gore, & Gauthier, 2009).

Although a good deal of research has investigated the object shape representations during basic and subordinate level of categorisation, none to our knowledge have yet employed eye movement pattern analysis to explore the perceptual basis of these two types of classification. Latest research findings showed that analyses of eye movement patterns can elucidate shape perception in human vision (Leek et al., 2012). In this study two groups of participants either actively memorised or passively viewed sets of visually similar novel objects prior to

performing a recognition memory test. The results showed that the distribution of eye movement patterns across object shape features was different between the study and test phases. More specifically, both groups demonstrated a preference for fixation at regions of surface concavities during the test phase (recognition task). The striking consistency of this preference raises a further theoretical question of whether local shape analysis strategies during perception are 'hard-wired' in the sense of being invariant to task requirements, and how robust this preference is across different levels of object classification, and level of training/expertise.

In this context, traditional theories in visual object recognition often posit controversial views about the organisation and structure of shape representations mediating object recognition such as those stemming from 'structural description' and those of 'image based models'. In general, the structural description models approach suggest that objects are represented as an arrangement of elementary viewpoint invariant 3D parts, called geons (Biederman, 1987; Biederman & Cooper, 1991; Hummel & Biederman, 1992), which are cylinders, bricks, wedges, or cones, with specified interrelations and spatial configurations which are viewpoint invariant. In contrast, image based models propose that object representation is supported by multiple 2D views and conjointly encode information about shape and the spatial locations of image features (Bülthoff et al., 1995; Bülthoff & Edelman, 1992; Edelman & Weinshall, 1991; Riesenhuber & Poggio, 2006). In this case one should expect to observe different shape analysis strategies in eye movement patterns depending on the level of classification. More specifically, we would expect a concave preference for part based classification at the basic level but not at the subordinate level. This suggestion is based on the premise that the structural description mediating the basic

level of classification involves part segmentation which occurs at point with concave local discontinuities.

However, both approaches appear to have difficulties with explaining the mechanisms of object classification. For example, image based models suggest that object representations are definitive to particular exemplars (given the assumption that objects are represented in a viewpoint-specific manner) not to object classes, thus these models do not provide information about recognising novel views of familiar object categories. Conversely, structural description models (e.g., Biederman, 1987) present an explanation about typical object recognition tasks regarding general classes of objects (e.g., basic level), but do not suggest how we process objects at a subordinate level.

Nevertheless, it should be noted that the two approaches (structural description and image based) are not mutually exclusive and both of them explain elements of human visual recognition; structural description providing information about categorical (basic) level access, and image based for within class level (subordinate) access. Moreover, as previously noted, some hybrid models propose that it is potentially achievable for both image based and structural description approaches to be accommodated within the same framework (e.g., Foster & Gilson, 2002; Hummel & Stankiewitz, 1996).

However, Murray (1998) demonstrated that basic level recognition is not exclusively mediated by structural description models and that viewpoint dependent mechanisms are evident during basic level discrimination involving visually dissimilar objects. Moreover, the author also shown that viewpoint invariant mechanisms are present during basic level of discrimination amongst visually similar objects. More specifically, the author proposed that whether an image based or structural description

based approach is employed, depends on the task in hand, the level of familiarity, and the similarity amongst the stimuli.

In the current study we employed the general design of the Wong et al., (2009) paradigm and in addition recorded eye movement patterns in order to investigate whether visual object shape representation changes with experience and if so, how. For example, when participants learn to attach semantic/conceptual information (i.e. family names vs. individual names) to novel objects, this should improve their subsequent recognition in the task they have been trained to do. However, but would this put different demands on the visual system and result in a change of the eye movement patterns as a function of training?

Nevertheless, regardless of whether holistic or analytic object shape representations are activated there is still a remaining question as to whether the visual system uses the same shape information (e.g., local image features) during object shape representation and if this same information is used when categorizing objects in to either the basic and/or subordinate level.

6.1.1. Method

Participants

Participants were 36 undergraduate and postgraduate students from Bangor University, participating in exchange for course and printer credits. Twelve of the participants were assigned to the Subordinate training group (8 females age M = 20.08, SD = 2.5), twelve to the Basic group (9 Females age M = 19.75, SD = 1.05) and twelve to a No-training control group (9 females, age M =20.33, SD = 3.23). All the participants reported normal or corrected to normal vision and seven were left handed.

Stimuli

Thirty-six novel objects called Ziggerins were kindly supplied by Alan Wong of The Chinese University of Hong Kong (see Figure 36). There were six different classes of Ziggerins, each defined by a distinctive part structure. Each class consisted of 6 Ziggerins, each defined by a part variation of size, aspect ratio and cross-sectional shape. The same style variations were applied to each of the six classes. This combination was suggested to be analogous to 6 different letters shown in 6 different fonts (Wong et al., 2009). The models were rendered in yellow at 72 dpi and scaled to fit within an 800 x 800 pixel frame. Stimuli subtended 18 degrees of visual angle horizontally from the viewing distance of 60 cm. This scale was specifically chosen to instigate saccadic movements over the stimuli.

Apparatus

Eye movement data were recorded on a Tobii ET-17 binocular eye-tracker as used in Experiment 1. Stimuli were presented on a TFT monitor running at a resolution of 1280 x 1024 pixels and 60 Hz refresh rate.

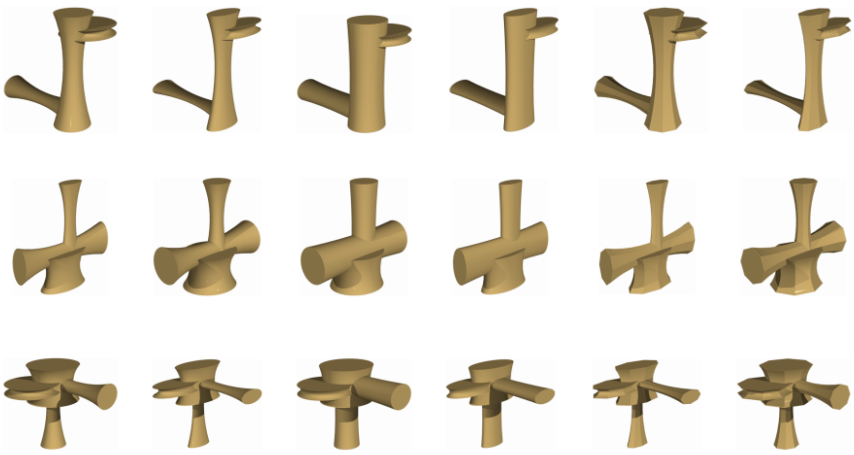


Figure 36 An illustration of the stimuli used in the experiment (Ziggerins).

Design

The design involved a pre-test, treatment, post-test configuration, with both pre- and post-tests requiring the participant to complete two sequential matching tasks (216 trials each) whilst eye movement fixation patterns were recorded. The 'treatment' (936 trials) in this case was the type of training group that each participant was assigned to, Basic training, or Subordinate training or a no-training (Control) group. The experiment used a between-subject design with three levels (Group: subordinate, basic, and control), and within-subjects design with the following factors: phase, with two levels (Pre-test; post-test), test-type, with two levels (subordinate; basic), and model, with two levels (Concave, Convex). The pre- and post-training tests were counterbalanced across participants. Half of the participants did the basic test first; the other half did the subordinate test first.

Procedure

Initially each participant completed calibration on the eye-tracker, where a static blue dot appeared randomly in each of 9 possible screen locations. Eye position and known screen position was recorded which allowed a transformation matrix to be constructed (via a linear interpolation method, which was used to determine gaze position from eye position). Prior to continuing beyond the calibration stage, the calibration results were visually inspected to ensure that a sufficiently good calibration have been performed.

Post-calibration, the participants were trained with 18 of the 36 Ziggerins while the remaining 18 Ziggerins were used for the pre-post tests. Following the sequential matching pre-test, the Subordinate training was completed in three one-hour sessions (on different days), after which the post-test was administered. The Subordinate group

learned the individual *names* of 12 Ziggerins, with 6 other Ziggerins left unnamed to be used as distracters. The Basic group learned to categorise 12 Ziggerins into 4 *different classes*, with 2 unnamed classes of Ziggerins used as distracters. The Ziggerin names (two-syllable nonsense words, e.g., Kimo, Vico) were randomly assigned and represented either names of classes or individual objects. All of the 'to-be-learned' Ziggerins were introduced at the first training session, which made the training sessions more intensive than those of the Wong et al. (2009) study, where the introduction was gradual. The participants received instructions stressing the importance of speed and accuracy in their performance.

Training

The participants were trained with 18 of the 36 Ziggerins while the remaining 18 Ziggerins were used for the pre-and post-training tests. Following the sequential matching pre-test, the Subordinate training was completed in three one-hour sessions (different days), after which, the post-test was administered. The Subordinate group learned the individual *names* of 12 Ziggerins, with 6 other Ziggerins left unnamed to be used as distracters. The Basic group learned to categorise 12 Ziggerins into 4 *different classes*, with 2 unnamed classes of Ziggerins used as distracters. The Ziggerin names (two-syllable none words, e.g., Malo, Divo) were randomly assigned and represented either names of classes or individual objects. All of the 'to-be-learned' Ziggerins were introduced at the first training session which made the training sessions more intensive than those of the Wong et al. (2009) study, where the introduction was gradual. The participants received instructions stressing the importance of speed and accuracy in their performance.

Each session of Basic and Subordinate-level training included a sequence of five tasks: *inspection, feedback, naming, verification, and matching* (See Table 7). The inspection task began with a 500 ms fixation cross; (fixation visual angel subtended 23.9º x 17.3º) followed by presentation of a Ziggerin alongside with a name for 1000ms, followed by blank screen for 200ms, and so on, until all of the 12 objects with names were presented. At this point no responce was required from the participants. During the feedback task, participants saw 500ms fixation cross, followed by a single Ziggerin and were required to press the correct key that denoted the first letter of the Ziggerin's name, and correct/incorrect feedback was given. The correct response for the distracters was pressing the space bar. In the naming task, following a 500ms fixation cross, a single Ziggerin was presented and participants were required to press the correct key associated with the first letter of the Ziggerin name. The correct response for distracters was pressing the space bar. No feedback was provided.

During the *verification task* a single name (either individual or family depending on the type of training) was shown for 1000ms, followed by a blank ISI of 200ms, followed by a Ziggerin which remained on the screen until a response (match or non-match) was made. Finally in the *matching task a* single name was presented for 1000ms, followed by a blank ISI of 200ms, and two Ziggerins simultaneously presented side by side. The participants has to respond by pressing the key L (left) or R (right) to indicate whether the name belonged to the Ziggering shown on the left or the right part of the screen. The participant had to decide whether the two objects either came from the same family (Basic-level training group) or had the same individual name (Subordinate-level training group). On 25% of the trials the Ziggerins came from the unused stimulus set.

Table 7 The five tasks in sequence, and a number of trials for Subordinate and Basic group training procedure.

		Subordinate Training	Basic Training
Task	Block	Number of trials	Number of trials
Inspection	Block 1	24	24
Feedback	Block 2	24	24
Naming	Block 3	300	300
Verification	Block 4	300	300
Matching	Block 5	287	287

Pre- and post-tests

The pre- and post-training tests were sequential matching tasks randomly counterbalanced across participants. Half of the participants did the basic test first; the other half did the subordinate test first. An initial fixation was shown for 800ms in one of four randomly selected corner locations (fixation visual angel subtended 20.0º x 19.2º) and followed by a Ziggerin for 800ms (S1), a mask for 800ms and a second Ziggerin stimulus (S2). S2 was displayed until response. There was a blank inter-trial interval of 1000ms between each trial. In half of the trials (216) the participants had to judge whether the two Ziggerins were the same or different individual; on the remaining trials (216), participants had to judge whether the two Ziggerins were from the same or different family. Prior to this task, in order to demonstrate the meaning of a family, a sheet with images of all Ziggerins was shown and participants told that objects within a particular row formed a family. The participants had to respond by pressing key ’z’ for same and ‘m’ for different (see Figure 37).

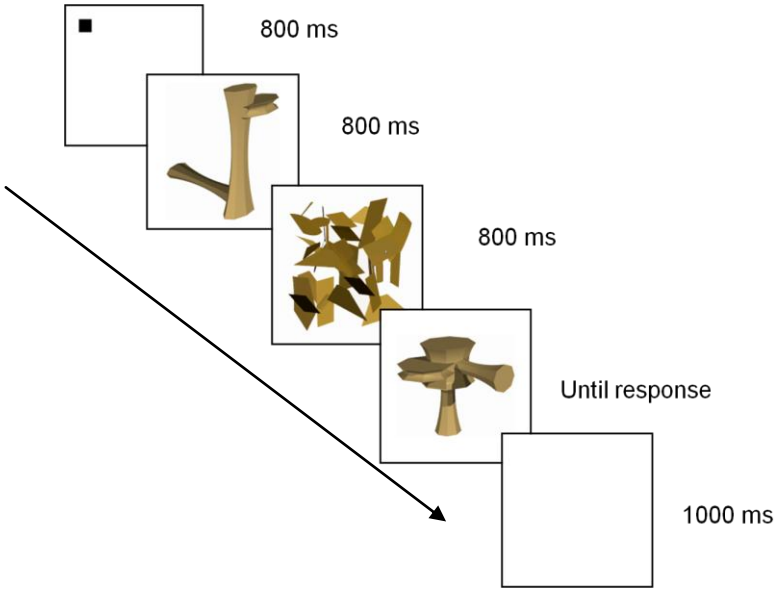


Figure 37 Example of the sequential matching task for pre and post tests, depicting a no match trial.

Analyses of Eye Movement Data

We have analysed the eye movement data using FROA (see Chapter 3).

6.1.2. Generating model predictions

In order to elucidate the shape information content at fixated image regions in recognition memory, predicted fixation region maps for three contrasting theoretical hypotheses were generated and compared with the recognition data from the pre-and post-tests. As described earlier, previous empirical work demonstrated that concave and convex curvature plays a key role in shape perception, but to my knowledge no one has yet specifically investigated these two types of curvature with eye movement analyses in the context of 3D object recognition. Hence the main focus of interest here was to investigate concave and convex curvature models as predictors of fixation patterns during recognition.

Model 1: Concave Surface Curvature Minima

The first model generated predicted fixation region locations based on the locations of local features defined by *concave* surface curvature maxima (e.g., Cohen & Singh, 2007; Hoffman & Richards, 1984; Hoffman & Singh, 1997). In order to generate predicted region maps that incorporate the same error measures as the recognition task data we used a trained observer technique. Twelve observers (9 females, mean age 28.00 *SD* = 10.18; 10 = right handed) were trained to fixate only locations containing an intersection between two surfaces that forms a concave (negative) minima of curvature. These observers had not taken part in the recognition task. Participants were first given extensive training in locating concavities in six novel objects otherwise not used in the recognition task. Following this, participants were shown the 18 stimuli for exposure duration of 10 seconds. A key advantage of this trained observer method is that it necessarily includes measurement error into the calculation of predicted fixation regions; that is, noise due to within and between-subject variation, as well as tracking accuracy and resolution.

Model 2 Convex Surface Curvature Maxima

The second model generated predicted fixation locations based on the locations of local features defined by *convex* surface curvature maxima (De Winter & Wagemans, 2006; Feldman & Singh, 2005). Twelve observers (11 females, mean age 29.75 *SD* = 9.55; 12 = right handed) were trained to fixate only image regions containing an intersection between two surfaces forming a convex (positive) curvature maxima. The observers had not taken part in the recognition or concavity detection tasks. The training, the stimuli and test phase procedures were identical to those used to generate the concave fixation data. Using identical FROA criteria as those used for the recognition task data we computed the fixation region maps. Data from this

condition also provided a control measure for the reliability of the trained observer patterns derived for Model 1 by demonstrating that observers can be trained to fixate only specific image features (that is, that their fixation patterns are not solely stimulus driven).

Visual saliency baseline

Similarly to the previous two studies, the visual saliency model was created using Saliency Toolbox implementation in Matlab (Walther & Koch, 2006). The model was run on each of the 18 stimulus images used in the recognition task to generate a saliency map for each stimulus. The output of the toolbox in terms of saliency values for each pixel which was grouped into a saliency region map using shape estimation function (see Walther & Koch, 2006). The number of saliency regions generated was constrained to approximate the area and number of thresholded regions generated for the other models: The saliency maps were thresholded and binarised using FROA in the same way as the empirically derived fixation data from the recognition task. These maps represent the thresholded distributions of fixation regions we would expect if eye movements were determined solely by low-level image statistics, that is, by the most visually salient image regions defined by colour, intensity contrast, and orientation.

This model was used as a baseline contrast as the question of interest here was whether specific models of shape analyses could account for fixation patterns beyond the explicable by visual saliency.

Independence of predicted fixation patterns

It is important to verify that the predicted region distributions of the two models are sufficiently different in order that they may be statistically distinguished

when compared to the gaze data. In order to do this FROA was used to compare region overlap across models.

Mean pixel region overlap across items for the Visual saliency and Convexity models was 23.84% (*SD* = 21.98%) of the total pixel area for the convexity model. Analyses of these data using FROA showed that the MMC distance for the observed region overlap between the two models was -3.13.

Mean pixel region overlap across items for the Visual saliency and Concavity models was 3.71% (*SD* = 6.92%) of total pixel area for the Concavity model. Analyses of these data using FROA showed that the MMC distance for the observed region overlap between the two models was -25.59.

Mean pixel overlap between the Convexity and Concavity models was 4.50 % (*SD* = 5.83%) of the total pixel area for the Convexity model. Analyses of these data using FROA showed that the MMC distance for the observed region overlap between the two models was -22.43.

A one way ANOVA (target vs. between models overlap: Visual saliency vs. Concavity, Visual saliency vs. Convex, Convex vs. Concave) on the MMC distance showed a significant main effect of Model, F (2, 53) = 14.08, $p < .0001$. Post-hoc analyses showed that the pairwise contrasts between models were significantly different for Visual Saliency_Concave vs. Visual Saliency_Convex, $p = .002$; Visual Saliency_Concave vs. Concave_Convex, $p = .004$, but not for Concave_Convex vs. Visual Saliency_Concave, $p = .145$ns.

6.1.3. Behavioural data analyses

Reaction time and accuracy analysis

Training data - Both Subordinate and Basic groups demonstrated learning and improved performance over the three training sessions. For the Subordinate group (see

Figure 38), Session 2 produced a mean improvement of 123.5 ms (*SD* = 153.12), this performance gain increased less sharply to 114.8 ms (*SD* = 46.48) by Session 3; a total improvement of 238.3 ms (*SD* = 199.59) between Session 1 and Session 3. Paired t-tests showed this trend of improved performance to be significant between Sessions 1 and 2, p = .034, and Sessions 1 and 3, p =.004. Response accuracy also improved commensurately from 84%, to 98%, for sessions 1 to 3 respectively. Wilcoxon tests showed this trend of improved response performance to be significant between Sessions 1 and 2, Z = -2.118, p =.034, Sessions 2 and 3, Z = -2.121, p = .034, and between Sessions 1 and 3, Z =-3.059, p = .004.

A similar pattern of results was obtained for the Basic group (see Figure 38) with mean response performance gain increasing from 675.4 ms (*SD* = 567.59), to 588.4 (*SD* = 401.62), and 567.9 ms (*SD* = 357.45) over the three respective Sessions showing an improvement of 107.6 ms between first and last sessions. Paired t-tests showed this trend of improved performance to be significant between Sessions 1 and 2, p < .0001, and Sessions 1 and 3, p <.0001.

Response accuracy improved from 97% to 98% for sessions 1 and 3 respectively. Wilcoxon tests showed this trend of improved response performance to be non-significant between all of the Sessions.

The above data patterns of significantly faster RT and accuracy well over 90 % post training demonstrates that both groups (basic and subordinate) considerably improved their performance as a function of training.

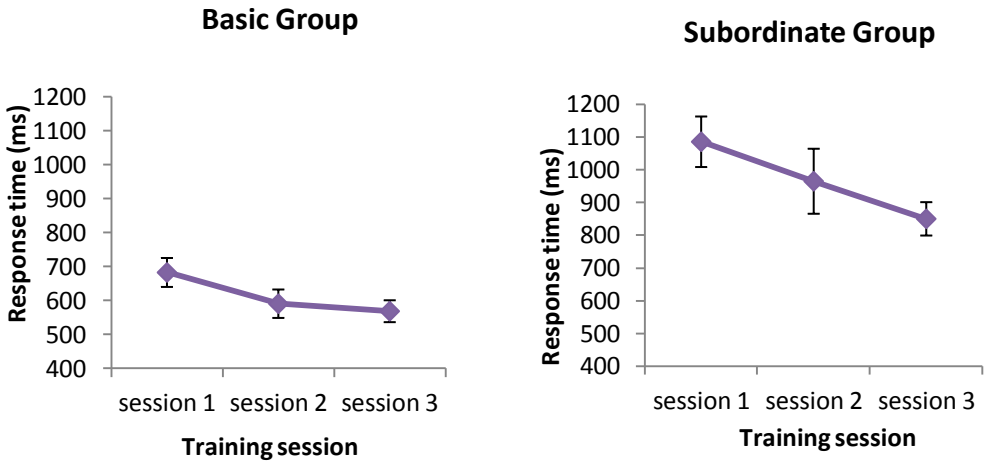


Figure 38 Mean RTs across sessions for the Basic and Subordinate training groups. Bars show standard error of the mean.

Sequential matching task: pre and post tests

Accuracy

The data from one participant in the Subordinate group was excluded because of a low accuracy rate (< 50%). The overall accuracy for the Subordinate, Basic and Control groups can be seen in Table 8 below. All three groups performed more accurately in the Subordinate test (94.6 %; *SE* =.009) than in the Basic test (91.0%; *SE* = .031) for both pre- and post-tests.

Table 8 The accuracy rates for Subordinate, Basic, and Control group for each phase (pre and post tests), and Test-type (subordinate and basic). Standard error of the mean is shown in parentheses.

	Pre-test phase		**Post-test phase**	
	Subordinate	**Basic**	**Subordinate**	**Basic**
	Mean accuracy (%)		Mean accuracy (%)	
Subordinate group	.93 (.011)	.90 (.028)	.96 (.007)	.91 (.028)
Basic group	.94 (.014)	.93 (.026)	.96 (.012)	.95 (.018)
Control group	.94 (.010)	.89 (.040)	.95 (.007)	.88 (.049)

Reaction times

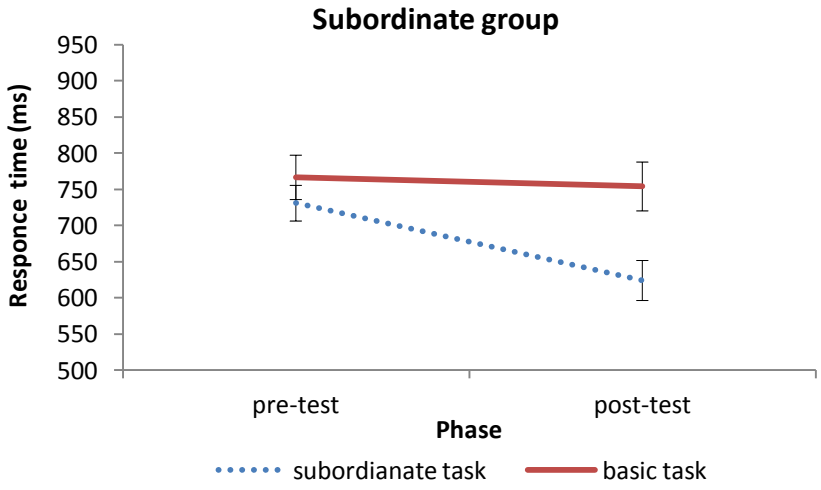
The data from four participants from the Subordinate group was excluded from the analyses. The reason for this was low accuracy in one of the cases (< 50%), and lack of training effects for the other three cases. Since the aim of the study was to examine changes in eye movement patterns resulting from training, the subjects who did not show reliable training effects were excluded from the analyses. The training effect was calculated as follows: we computed the difference between mean RT scores for pre and post tests for each Test-type (subordinate vs. basic), which gave us a single value for each Test-type. Then we computed the difference between these two values (subordinate – basic) for each participant and if the final value was negative, this was taken as evidence that the participant did not improve RT performance as a function of training, thus was excluded from further analyses. The overall Mean RT for the Subordinate, Basic and Control groups can be seen in Table 9 below.

Table 9 The Mean RT for Subordinate, Basic, and Control group for each phase (pre and post tests), and Test-type (subordinate and basic). Standard error of the mean is shown in parentheses.

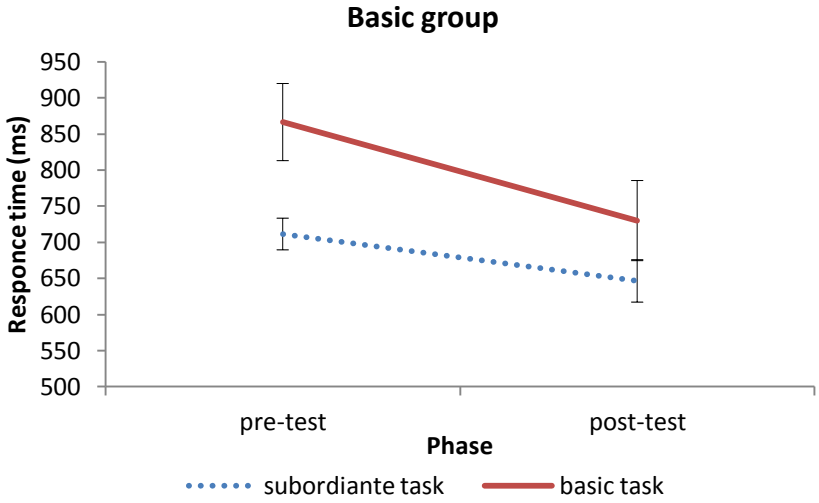
	Pre-test phase		**Post-test phase**	
	Subordinate	**Basic**	**Subordinate**	**Basic**
	Mean RT (SE)	Mean RT (SE)	Mean RT (SE)	Mean RT (SE)
Subordinate group	731.02 (24.63)	766.74 (30.70)	624.03(27.68)	754.10 (33.85)
Basic group	711.43 (21.96)	866.41(53.39)	646.52(29.31)	730.00 (55.47)
Control group	636.99 (20.20)	799.23(25.93)	598.03(17.94)	626.95 (13.91)

Each group showed improvement as a function of training; Subordinate group difference for pre-post subordinate tests was 107.00 ms, and 12.64 ms difference for basic tests. The Basic group difference for pre-post subordinate tests was 64.91 ms, and 136.41 ms for basic tests (see Figure 39).

(a)



(b)



(c)

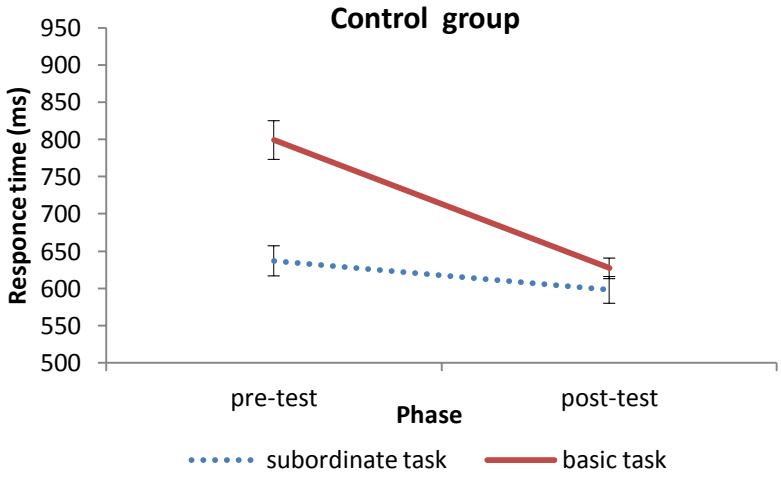


Figure 39 Response performance (mean RT) for Subordinate, Basic, and Control groups. Each figure shows mean RT for each phase (pre-test and post-test) and each Test-type (subordinate and basic). Bars show standard error of the mean.

A within group 2 (Phase) x 2 (Test-type) ANOVA for the Subordinate Group showed no significant main effects of Phase, F (1, 7) = 3.48, p =.104, η_p^2 = .332ns, and Test-type, F (1, 7) = 4.19, p =.080, η_p^2 = .375ns. There was a two way interaction between Phase and Test-type, F (1, 7) = 5.73, p =.048, η_p^2 = .450. A paired t- test showed a significant difference between the subordinate level pre-test and post-test, p =.025, but not between the basic level pre-test and post-tests, providing evidence that the Subordinate group improved their performance selectively as a function of training.

A within group 2 (Phase) x 2 (Test-type) ANOVA for the Basic Group showed a significant main effect of Phase, F (1, 11) = 7.94, p =.017, η_p^2 = .419, and Test-type, F (1, 11) = 4.85, p =.050, η_p^2 = .306. There were no other main effects and interactions.

A within group 2 (Phase) x 2 (Test-type) ANOVA for the Control Group showed a significant main effect of Phase, F (1, 11) = 38.61, p <.0001, η_p^2 = .778, and Test-type, F (1, 11) = 19.54, p =.001, η_p^2 = .640. There was a two way interaction between Phase and Test-type, F (1, 11) = 14.79, p =.003, η_p^2 = .573. A paired t-test showed a significant difference between the basic level pre-test and post-test, p <.0001, but not between the subordinate level pre-test and post-test.

A mixed 3 x 2 x 2 analysis of variance (ANOVA) with a *Between Subject* factor of treatment Group (Subordinate, Basic, Control) and *Within Subject* of Phase (pre-test or post-test) x Test-type (Basic or Subordinate) showed a significant main effect of Group, F (2, 29) = 7.21, p =.003, η_p^2 = .33, Phase, F (1, 29) = 26.66, p <.0001, η_p^2 = .479, and Test-type, F (1, 29) = 106.32, p <.0001, η_p^2 = .360. A three way interaction failed to reveal significance, F (2, 29) = 3.23, p =.054, η_p^2 = .182ns.

These results show that both training groups (Subordinate and Basic) improved their post-test mean RT performance, as a function of training, but only the Subordinate group showed a specific training effect on the subordinate task. The Basic group showed a large but not significant improvement in the basic post-tests and a small and not significant improvement in the subordinate post-test. Two potential interpretations for the latter improvement could be, first this may be result of a greater basic level experience that proceeds and translates in to the next level of subordinate processing (e.g. Jolicoeur et al., 1984) and second, this could be interpreted as a priming effect in (pre-vs.-post test) subordinate level of categorisation.

The Control group showed a significant improvement in the basic post-tests and a small but not significant improvement in the subordinate post-test. Some possible explanations for this pattern are: First, the Control group showed a basic level priming effect which was not evident in the Subordinate group performance as it was interrupted by training. Second, the improvement in the basic post-test may be due to the fact that there were 6 stimulus-response mappings in the basic sequential matching test, compared to 18 stimulus-response mappings in the subordinate sequential matching test. Third, this result appears to be consistent with previous findings (e.g., Rosch et al., 1976) showing that basic level categorisation is the preferred entry access of object classification in novices.

6.1.4. Analyses of eye movement data

Analyses between groups

A 3 (Group: Basic, Subordinate, Control) x 2 (Phase: pre, post) x 2 (Test-type: basic, subordinate) x 3 (Model: Concave, Convex, Visual saliency) ANOVA showed a significant main effect of Model, F (2, 102) = 36.99, $p < .0001$, $\eta_p^2 = .420$, and Phase,

F (1, 51) = 15.28, p < .0001, η_p^2 = .230, along with a two way interaction between Phase and Group, F (2, 51) = 4.69, p = .013, η_p^2 = .155, and a three way interaction between Model, Phase, and Test-type, F (2, 102) = 4.00, p = .021, η_p^2 = .073. There were no other main effects or interactions. Post -hoc pairwise comparisons (Bonferroni) between each model were all significant for Concave vs. Convex, p < .0001, Concave vs. Visual saliency, p <.0001, and Convex vs. Visual saliency, p =.001. Further pairwise comparisons between Concave and Convex models were significantly different for each phase and test-type; pre-subordinate, p =.004, pre-basic p < .0001, post-subordinate p <.0001, post-basic = .007. These analyses show that the fixation pattern differences can be attributed to the differential group treatment; hence, further within group analyses were performed.

Analyses within group

Basic group

A 2 (Model: Concave, Convex) x 2 (Phase: pre, post) x 2 (Test-type: Subordinate, Basic) within subjects ANOVA showed a significant main effect of Model, F (1, 17) = 7.34, p = .015, η_p^2 = .301, along with two-way interaction between Model and Test-type, F (1, 17) = 4.67, p = .045, η_p^2 = .216. There were no other main effects or interactions (see Figure 40). Post-hoc pairwise comparison between model and Test-type was significant for pre-basic test (Concave vs. Convex), p = .004, as well as both post-tests basic and subordinate (Concave vs. Convex), p =.009, p = .048 respectively. Post-hoc pairwise comparison (Bonferroni) between both Models was significant for Concave vs. Convex, p = .015.

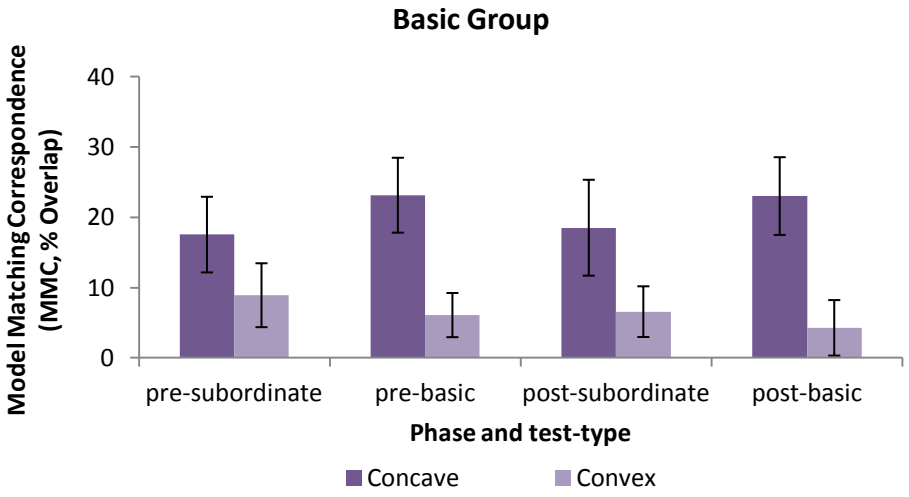


Figure 40 Mean MMC $_{Mx}$ – MMC $_{vs}$ measure of data-model correspondences between models (relative to visual saliency) for for Basic group for each phase (pre-test and post-test) and each Test-type (subordinate and basic). Bars show standard error of the mean (% overlap).

Subordinate Group

A 2 (Model: Concave, Convex) x 2 (Phase: pre, post) x 2 (Test-type: subordinate, basic) within subjects ANOVA showed a significant main effect of Model, F (1, 17) = 9.02, p = .008, η_p^2 = .347. There were no other main effects or interactions (see Figure 41). Post-hoc pairwise comparison (Bonferroni) between both Models was significant for Concave vs. Convex, p = .008.

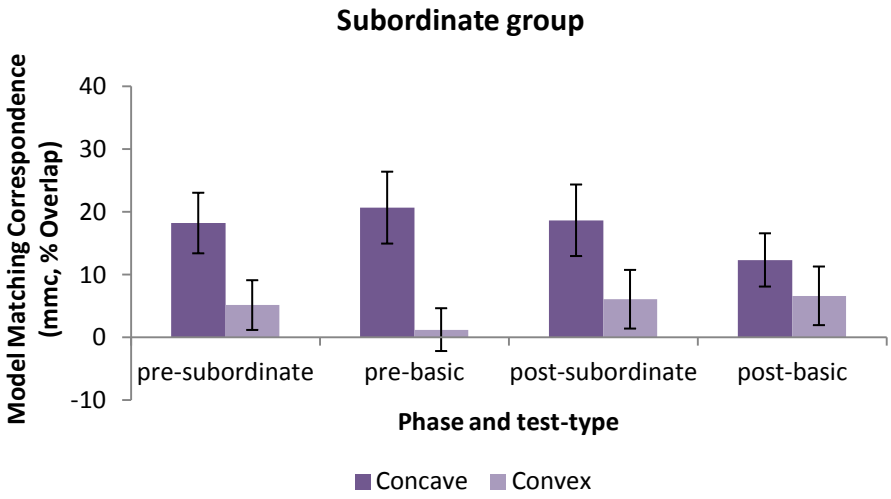


Figure 41 Mean MMC $_{Mx}$ – MMC $_{vs}$ measure of data-model correspondences between models (relative to visual saliency) for Subordinate group for each phase (pre-test and post-test) and each Test-type (subordinate and basic). Bars show standard error of the mean (% overlap).

Control Group

A 2 (Model: Concave, Convex) x 2 (Phase: pre, post) x 2 (Test-type: subordinate, basic) within subjects ANOVA showed a significant main effect of Model, F (1, 17) = 7.42, p = .014, η_p^2 = .304. There was a significant two-way interaction between Phase and Test-type, F (1, 17) = 6.23, p = .023, η_p^2 = .268, along with a three-way interaction between Model, Phase, and Test-type F (1, 17) = 8.63, p = .009, η_p^2 = .337. Post-hoc pairwise comparison (Bonferroni) between both Models was significant for Concave vs. Convex, p = .014 (see Figure 42).

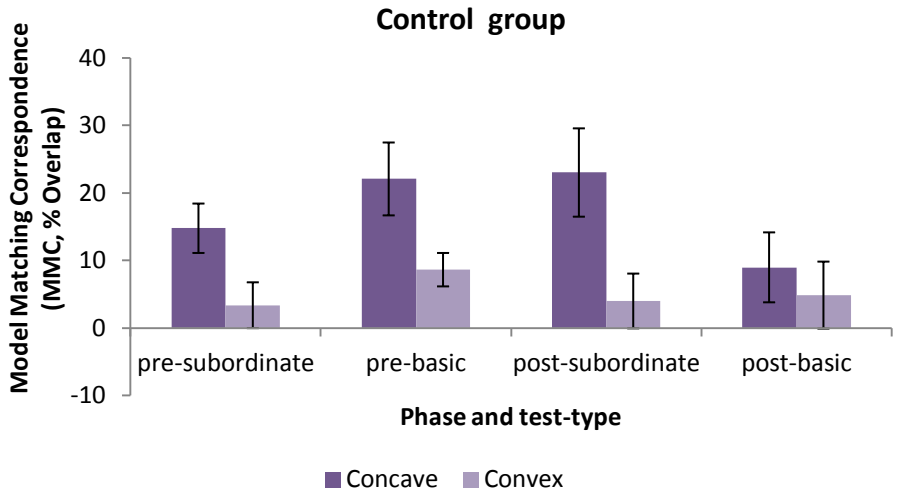


Figure 42 Mean MMC $_{Mx}$ – MMC $_{vs}$ measure of data-model correspondences between models (relative to visual saliency) for Control group for each phase (pre-test and post-test) and each Test-type (subordinate and basic). Bars show standard error of the mean (% overlap).

In order to test whether there is a difference in the data-model correspondence regardless of training, the data was collapsed across training group for each test (Subordinate, Basic) and phase (pre-test, post-test; see Figure 43). A 2 (Phase: pre, post) x 2 (Test-type: basic, subordinate) x 3 (Model: Concave, Convex, Visual saliency) ANOVA showed a significant main effect of Phase, F (1, 17) = 14.81, p = .001, η_p^2 = .466, and Model, F (2, 34) = 16.57, p < .0001, η_p^2 = .494, along with a three way interaction (Phase, Test-type, Model), F (2, 34) = 3.64, p = .037, η_p^2 = .176. Post-hoc paired comparisons between pre-tests and post-tests were significant for the basic (Concave), and basic (Visual saliency) tests, p = .003 and p = .033, respectively.

Post-hoc paired comparisons (Bonferroni) between each Test-type (Subordinate vs. Basic) were significant for pre-test (Convex), p = .024. Post-hoc pairwise comparisons (Bonferroni) between each model were significant for Concave vs. Convex, p =.015, Concave vs. Visual saliency, p <.0001, but not for Convex vs. Visual saliency, p =.063ns.

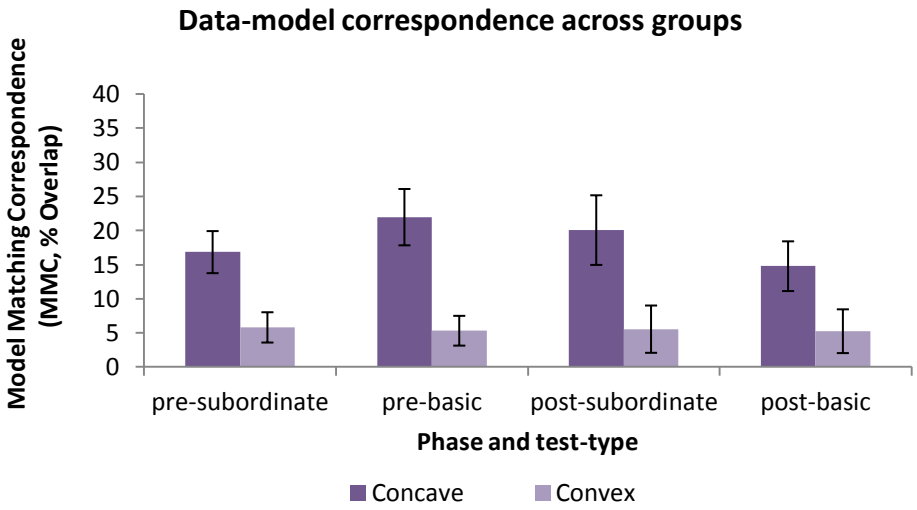


Figure 43 Mean MMC $_{Mx}$ – MMC $_{vs}$ measure of data-model correspondences between models (relative to visual saliency) collapsed across groups for each phase (pre-test and post-test) and for each Test-type (subordinate and basic). Bars show standard error of the mean (% overlap).

6.1.5. Saccade amplitude and dwell analyses

The data was further analysed in terms of saccade amplitude and dwell time for each group, test phase and Test-type. Saccade amplitude is the distance in visual angle/degrees between two fixations in a Euclidean plane, whereas the dwell time is the duration between saccades. Previous research (e.g. Tatler & Vincent, 2008) has shown that saccade amplitude and dwell time can be systematically related, thus analysing them here could provide additional information about perceptual processing during basic and subordinate categorisation tasks. More generally, it has been found that in scene viewing (e.g. Unema et al.,2005; Velichkovsky, Joos, Helmert, & Pannasch, 2005) large saccade amplitudes and short dwell times are associated with a global scanning approach, whereas small saccade amplitudes and longer dwell times are attributed to a local scanning approach. Two points of interest to examine here are the potential changes in viewing strategies between encoding (Image 1) and

recognition (Image 2) as well as saccade amplitude and dwell time changes before and after training.

Saccade amplitude analyses

The overall mean saccade amplitude for pre-mask (the first image presented in the sequential matching task) Image 1 for the Subordinate, Basic and Control groups can be seen in Table 10 below.

Table 10 The mean saccade amplitude (SA) for pre-mask Image 1 for the Subordinate, Basic, and Control group for each phase (pre-test and post-test), and Test-type (subordinate and basic). Standard error of the mean is shown in parentheses.

	Pre-test phase		**Post-test phase**	
	Subordinate	**Basic**	**Subordinate**	**Basic**
	Mean SA (SE)	Mean SA (SE)	Mean SA (SE)	Mean SA (SE)
Subordinate group	2.60 (0.23)	2.46 (0.21)	3.00 (0.28)	2.47 (0.27)
Basic group	2.93 (0.26)	2.85 (0.17)	3.11 (0.27)	3.04 (0.29)
Control group	2.96 (0.28)	2.96 (0.36)	3.52 (0.41)	3.16 (0.44)

The overall mean saccade amplitude for post-mask Image 2 (the image presented after the mask during the sequential matching task) for the Subordinate, Basic and Control groups can be seen in Table 11 below.

Table 11 The mean saccade amplitude (SA) for post-mask Image 2 for the Subordinate, Basic, and Control group for each phase (pre-test and post-test), and Test-type (subordinate and basic). Standard error of the mean is shown in parentheses.

	Pre-test phase		**Post-test phase**	
	Subordinate	**Basic**	**Subordinate**	**Basic**
	Mean SA (SE)	Mean SA (SE)	Mean SA (SE)	Mean SA (SE)
Subordinate group	2.18 (0.30)	1.74 (0.22)	1.99 (0.30)	1.96 (0.30)
Basic group	2.32 (0.38)	1.96 (0.23)	2.07 (0.31)	1.86 (0.24)
Control group	2.17 (0.25)	2.18 (0.50)	2.54 (0.50)	2.13 (0.52)

A mixed 3 x 2 x 2 x 2 analysis of variance (ANOVA) with a *Between Subject* factor of experimental Group and *Within Subject* of Phase (pre-test or post-test) x Test-type (Basic or Subordinate) x Image (pre-mask image 1 or post-mask image 2) showed a significant main effect of Phase, F (1, 29) = 39.43, p <.0001 η_p^2 = .576, and Image, F (1, 29) = 7.76, p =.009, η_p^2 = .211. The data also showed a two way interaction, between Phase and Test-type, F (1, 29) = 6.98, p =.013, η_p^2 = .194. (see Figure 43). To explain this further we have conducted within group analyses.

A within group 2 x 2 x 2 ANOVA for the Basic Group showed a significant main effect of Phase, F (1, 11) = 13.30, p =.004, η_p^2 = .547, but no other main effects or interactions. Paired t-tests between image 1 and image 2 showed a significant difference for the subordinate pre test, p =.018, basic pre test, p <.0001, subordinate post test, p =.026, and the basic post test, p =.007.

A within group 2 x 2 x 2 ANOVA for the Subordinate Group showed a significant main effects of Phase, F (1, 7) = 7.32, p =.030, η_p^2 = .511, and Image, F (1, 7) = 16.91, p =.005, η_p^2 = .707, but no other main effects or interactions. Paired t-tests

between image 1 and image 2 showed a significant difference for the basic pre test, p =.014, and subordinate post test, p =.029.

A within group 2 x 2 x 2 ANOVA for the Control Group showed a significant main effect of Phase, F (1, 11) = 27.25, p <.0001, η_p^2 = .712, but no other main effects or interactions. Paired t-tests between Image1 and Image 2 showed a significant difference for the basic pre test, p =.013, subordinate pre test, p =.001, basic post test, p <.0001, and subordinate post test, p <.0001.

All three groups showed significant differences between Image 1 and Image 2 in the sequential matching task, indicating that the participants changed their viewing strategies between the initial encoding of the image and subsequent recognition; however we cannot attribute these changes to the level of expertise, as there were no significant changes in fixation amplitude before and after training (see figure 44)

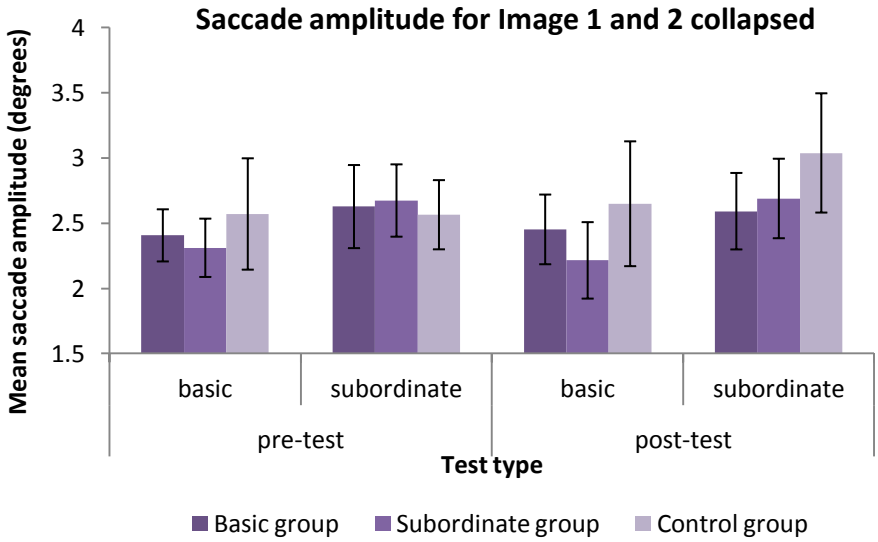


Figure 44 Mean fixation amplitude (degrees) for Basic, Subordinate, and Control group for each phase (pre-test and post-test) and Test-type (subordinate and basic) for Image 1 and Image 2 collapsed. Bars show standard error of the mean.

Given the lack of main effects or interactions between test-types, I have conducted additional analyses involving collapsing the data across test-type and phase for Image 1 and Image 2. The results showed a consistent pattern of longer saccade amplitudes for Image 1 and shorter saccade amplitudes in Image 2 for each group. The Basic group showed significantly longer saccade amplitude for Image 1 (M = 2.98; SE =.12) than for Image 2 (M = 2.05; SE =.14), p <.0001, R^2 =.067. Similarly, the Subordinate group saccade amplitude was significantly longer for Image 1 (M = 2.63; SE =.12) than Image 2 (M = 1.96; SE =.13), p <.0001, R^2 =.060. The Control group also showed significantly longer saccade amplitude for Image 1 (M = 3.15; SE =.18) than Image 2 (M = 2.25; SE =.22), p <.0001, R^2 =.079.

Dwell time analyses

The overall mean dwell time for pre-mask Image 1 for the Subordinate, Basic and Control groups can be seen in Table 12 below.

Table 12 The mean dwell times (DT) for pre-mask Image 1 for the Subordinate, Basic, and Control group for each phase (pre-test and post-test), and Test-type (subordinate and basic). Standard error of the mean is shown in parentheses.

	Pre-test phase		**Post-test phase**	
	Subordinate	**Basic**	**Subordinate**	**Basic**
	Mean DT (SE)	Mean DT (SE)	Mean DT (SE)	Mean DT (SE)
Subordinate group	187.86(6.45)	188.12 (7.10)	184.24 (6.96)	192.45 (8.24)
Basic group	164.87(8.38)	161.30 (8.05)	168.61 (9.42)	165.54 (9.25)
Control group	184.33 (8.71)	184.80 (8.06)	192.70 (9.15)	192.19 (9.49)

The overall mean saccade amplitude for post-mask Image 2 for the Subordinate, Basic and Control groups can be seen in Table 13 below.

Table 13 The mean dwell times (DT) for post-mask Image 2 for the Subordinate, Basic, and Control group for each phase (pre-test and post-test), and Test-type (subordinate and basic). Standard error of the mean is shown in parentheses.

	Pre-test phase		**Post-test phase**	
	Subordinate	**Basic**	**Subordinate**	**Basic**
	Mean DT (SE)	Mean DT (SE)	Mean DT (SE)	Mean DT (SE)
Subordinate group	224.30 (19.50)	231.62 (21.26)	200.62 (10.70)	213.22 (14.74)
Basic group	193.72 (14.98)	194.05 (15.78)	174.99 (9.88)	184.09 (10.46)
Control group	210.98 (19.16)	206.08 (15.95)	206.53 (13.04)	212.72 (21.28)

A mixed 3 x 2 x 2 x 2 analysis of variance (ANOVA) with a *Between Subject* factor of experimental Group and *Within Subject* of Phase (pre-test or post-test) x Test-type (Basic or Subordinate) x Image (pre-mask Image 1 or post-mask Image 2) showed a significant main effect of Phase, F (1, 29) = 23.58, p <.0001 η_p^2 = .448. The data also showed a two way interaction, between Test-type and Group, F (1, 29) = 3.59, p =.041, η_p^2 = .198, Phase and Test-type, F (1, 29) = 10.61, p =.003, η_p^2 = .268, and Phase and Image, F (1, 29) = 4.28, p =.048, η_p^2 = .129.

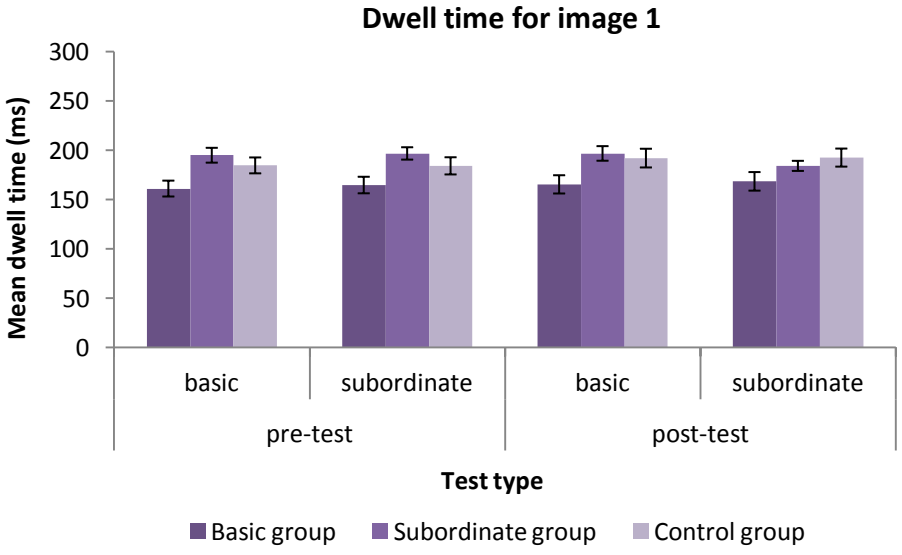
A within group 2 x 2 x 2 ANOVA for the Basic Group showed a significant main effect of Phase, F (1, 11) = 19.14, p =.001, η_p^2 = .635. The data also showed a two way interaction, between Phase and Image F (1, 11) = 5.46, p =.039, η_p^2 = .332, and Test-type and Image, F (1, 11) = 5.29, p =.042, η_p^2 = .325. Paired t- tests between Image 1 and Image 2 showed a significant difference for the subordinate pre-test, p =.007, basic pre-test, p =.003, and the basic post-test, p =.002

A within group 2 x 2 x 2 ANOVA for the Subordinate Group showed a significant main effect of Phase, F (1, 7) = 6.87, p =.034, η_p^2 = .495, but no other main

effects or interactions. Paired t-tests between Image 1 (see Figure 45 a) and Image 2 (see Figure 45b) showed a significant difference for the basic pre-test, p =.043, subordinate post-test, p =.032. There was also a significant difference between subordinate pre-and-post-tests for Image 1, p =.026. These results indicate that the participants changed their viewing strategies between the initial encoding of the image and subsequent recognition in the post subordinate sequential matching task as well as between pre and post subordinate test for Image 1. This pattern of results suggests that the subordinate group changed their viewing strategies as a function of training.

A within group 2 x 2 x 2 ANOVA for the Control Group showed no significant main effects or interactions. Paired t-tests between Image 1 and Image 2 showed a significant difference for the subordinate pre-test, p =.043, and subordinate post-test, p =.023. There were no other significant differences between pre-and-post-tests.

(a)



(b)

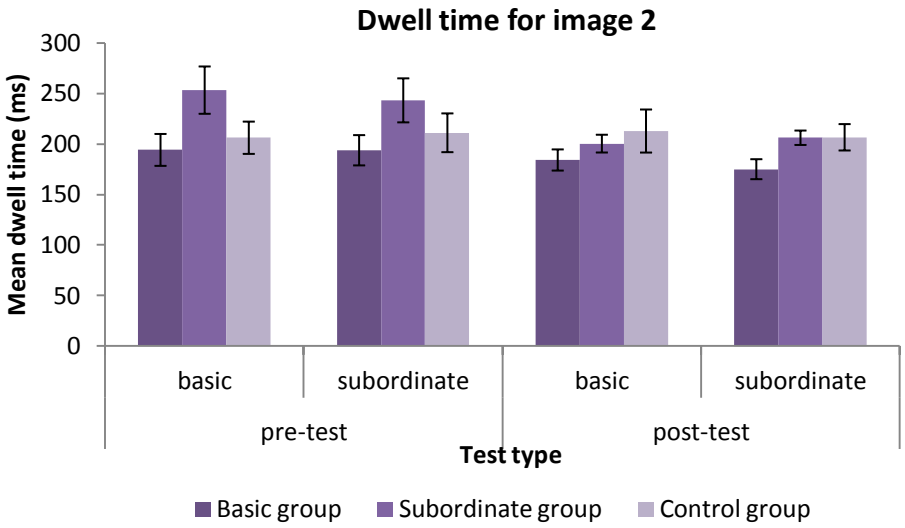


Figure 45(a) Mean dwell time (ms) for Basic, Subordinate, and Control group for each phase (pre-test and post-test) and Test-type (subordinate and basic) for Image 1. Bars show standard error of the mean. (b) Mean dwell time (ms) for Basic, Subordinate, and Control group for each phase (pre-test and post-test) and Test-type (subordinate and basic) for Image 2. Bars show standard error of the mean.

In further analyses I have collapsed the data across test-type and phase for Image 1 and Image 2. The results showed consistent pattern of shorter dwell time for Image 1 and longer dwell time for Image 2 for each group. The Basic group dwell time was significantly shorter for Image 1 (M =165.07; SE = 4.27) and Image 2 (M = 186.71; SE = 6.41), p <.0001, R^2 =.066. Similarly, the Subordinate group dwell time was significantly shorter for Image 1 (M = 188.17; SE = 3.52) and Image 2 (M = 217.43; SE = 8.43), p <.0001, R^2 =.056. The Control group also showed significantly shorter dwell time for Image 1 (M = 188.51; SE = 4.33) and Image 2 (M = 209.08; SE = 8.54), p <.0001, R^2 =.051.

Summary of results

Training data

- Both training groups significantly improved their RTs between Session 1 and Session 3, and demonstrated over 90% accuracy post training.

Behavioural data

- The RTs showed that each training group improved their post-test mean performance as a function of training, but only the Subordinate group showed a significant specific effect in the subordinate task.

Eye movement data

- The Basic training group showed a main effect of Model. There was also a two-way interaction between Model and Test-type. The data-model correspondence was significantly higher for the Concave model than the Convex model.
- The Subordinate training group showed a significant main effect of Model. The data-model correspondence was significantly higher for the Concave model than the Convex model.
- The Control group showed a significant main effect of Model. There was also a two-way interaction between Phase and Test-type and a three-way interaction between Model, Phase and Test-type. The data-model correspondence was significantly higher for the Concave model than the Convex model.
- Analysis of data-model correspondences across training groups showed a main effect of Phase and Model. There was also a three-way interaction between Phase, Test-type and Model. Post-hoc paired comparisons between models were higher for the Concave model than the Convex model. The Visual saliency model had the lowest data-model correspondence pattern.

Saccade amplitude analyses

- Within group analyses for the sequential matching tests for all three groups (Basic, Subordinate, and Control) showed that each group had a significantly longer saccade amplitude difference for Image 1 than for Image 2, suggesting that the participants changed their viewing strategies between initial encoding and subsequent recognition.

Dwell time analyses

- The Basic group showed a main effect of Phase, as well as a two-way interaction between Phase and Image as well as Test-type and Image. There was also significantly shorter dwell times for Image 1 than for Image 2, suggesting that the participants changed their viewing strategies between initial encoding and subsequent recognition.
- The Subordinate group analyses for the sequential matching task showed a main effect of Phase along with significantly shorter dwell times for Image 1 than for Image 2.
- The Control group analyses for the sequential matching task showed significantly shorter dwell times for Image 1 than for Image 2 for pre-and post subordinate tests.
- Overall, the saccade amplitude and dwell time analyses showed longer saccades and shorter dwell times during the encoding phase, (Image 1), which is typically characteristic of a more global scanning approach. During the recognition phase (Image 2) the results showed shorter saccades and longer dwell times, which is more characteristic of a local scanning approach.

6.1.6. Conclusions

In Experiment 3 we used fixational eye movement patterns to examine local shape analysis processes during object shape categorisation in experts and novices. In pre-test phases, observers sequentially matched visually similar novel objects to a basic and subordinate level of classification. Post-training, the basic and subordinate experts, performed the same tasks as in the pre-test. These results show that both

training groups (Subordinate and Basic) improved their post-test mean RT performance, as a function of training, but only the Subordinate group showed a specific training effect on the subordinate task. The Basic group showed a large but not significant improvement in the basic post-tests and a small and not significant improvement in the subordinate post-test. Two potential interpretations for the latter improvement could be, first this may be result of a greater basic level experience that proceeds and translates in to the next level of subordinate processing (e.g. Jolicoeur et al., 1984) and second, this could be interpreted as a priming effect in (pre-vs.-post test) subordinate level of categorisation.

The Control group showed a significant improvement in the basic post-tests and a small but not significant improvement in the subordinate post-test. Some possible explanations for this pattern are: First, the Control group showed a basic level priming effect which was not evident in the Subordinate group performance as was interrupted by training. Second, the improvement in the basic post-test may be due to the fact there were 6 stimulus-responses mapping in the basic sequential matching test, compared to 18 stimulus-response mappings in the subordinate sequential matching test. Third, this results appear to be consistent with previous findings (e.g., Rosch et al., 1976) showing that basic level of categorisation is indeed the preferred entry access of object classification in novices.

The analyses of the spatial distributions of fixations revealed a consistent pattern of data-model correspondences across tasks. During both basic and subordinate matching tasks in pre and post-test phases we found evidence that fixation patterns are predominantly driven by shape information defined by internal regions of concave surface discontinuity regardless of expertise or level of categorisation (basic or subordinate).

These findings are consistent with previous studies demonstrating the importance of curvature singularities in the visual perception of shape (e.g., Attneave, 1954; Barenholtz Cohen & Singh, 2003; De Winter & Wagemans, 2006; Feldman & Singh, 2005; Hoffman & Richards, 1984; Hoffman & Singh, 1997; Leek et al, 2012). Further, the finding of a preference for fixation at regions of concave surface discontinuity regardless of task (basic vs. subordinate) or phase (pre vs. post) provides information for a direct link between the encoding of information about surface concavity and object categorisation. Thus the central issues raised here concern our observation of similar fixation distributions, similar perceptual strategies for the acquisition of shape information, and later categorisation across pre and post tasks. Furthermore, our finding of a preference for fixation at regions of concavity is consistent with an influential hypothesis where concave regions are suggested to play a functional role in part segmentation, during the derivation of a structural description representation (e.g., Biederman, 1987; Hoffman & Richards, 1984; Marr & Nishihara, 1978).

A further aspect of the results that is of theoretical interest is the consistency of the patterns of data-model correspondences during pre and post-tests. This is perhaps surprising given that one might expect task requirements to affect the perceptual analysis of shape. Here, despite the fact that the participants were trained to classify objects into two different categories (basic vs. subordinate), along with a control group that received no training, the perceptual analysis strategies of the three groups, as evidenced by the patterns of data-model correspondences, were similar.

These results are consistent with models of object recognition which hypothesise a special functional status to concave regions in object shape representation (e.g. Biederman, 1987; Hummel & Stankiewitz, 1996; Leek, Reppa,

Rodriguez & Arguin, 2009; Leek, Reppa & Arguin, 2005; Marr & Nishihara, 1978), and they present a challenge to models which do not. Among the latter models are some recent Image-based models of recognition (e.g., Edelman & Weinshall, 1991; Riesenhuber & Poggio, 2006; Ullman & Basri, 1991), including HMAX (Serre, Oliva & Poggio, 2007; Serre, Wolf, Bileschi, Riesenhuber & Poggio, 2007). In order to account for these data, these models would require modification to include a level of feature representation which makes explicit the sign of curvature – for example, in the C2 layer of HMAX.

This raises the further issue of why regions of concave curvature should carry such functional significance. Perhaps the most influential hypothesis, following the seminal work of Hoffman and Richards (1984), is that concavities play a key role as local part boundaries supporting volumetric image segmentation. But this does not exclude the possibility that concave regions play other roles in shape recognition and image classification. One important implication of the current results (and those reported by Leek et al., 2012) is that object recognition makes use of local depth information at least to the level of the $2^1/_2$D sketch (Marr & Nishihara, 1978), and does not rely solely on 2D image features computed from the retinal input. In this respect also, these findings present a challenge to object recognition models that are based solely on 2D image-based representations.

One implication of this finding is that local shape analysis strategies during perception are 'hard-wired' in the sense of being invariant to task requirements - at least across the range of tasks tested here. Under the current context, the suggestion of 'hard-wired' mechanism is consistent with recent (Amir, Biederman & Hayworth, 2011) findings that adults and infants as young as 5 month-old, looked first and adults looked longer at simple volumetric shapes containing curved contours, as opposed to

straight contours, thus shapes containing high curve value produced larger blood oxygenated level dependent (BOLD) activity in the adult's shape selective cortex (Lateral Occipital Cortex). This finding implies that perceptual mechanisms directing attention to informative object segments exist from early infancy and are not affected by language, geometry training or cultural values.

This hypothesis is intuitively appealing in that during everyday recognition observers cannot entirely predict when unfamiliar objects might become relevant to their immediate or future goals and intentions.

The current pattern of results raises the question whether shape recognition is invariant to classification level, or whether the same kind of perceptual shape analyses processes/information underlies the classification of basic and subordinate levels. On the other hand, perceptual strategies could be different for basic and subordinate level of classification, but basic level of classification takes part before the subordinate level, which reflects in our findings.

However, it remains to be determined whether the observed patterns of shape-analyses found here will generalise across other tasks, including, for example, those related to the computation of shape representations for reaching and grasping (e.g., Land et al., 1999.

Chapter 7

7.1. Experiment 4

This experiment builds upon and extends the preceding study in order to determine whether extending the training time from three to four hours would result in a greater level of training effect, as measured by the Response time and Accuracy performance. In addition we modified the number of exposures for each object at pre- and post-test, thus each object was presented only once. This modification was made in order to allow us to measure the participant's performance by avoiding potential priming effects (increased sensitivity to previously presented stimuli). Also, we wanted to analyse the first exposure to each novel object in pre and post tests, which would give us a clearer picture of actual changes in performance (reaction time, accuracy and fixation locations) between pre and post tests.

Participants

Participants were 36 undergraduate and postgraduate students from Bangor University, participating in exchange for course and printer credits. Twelve of the participants were assigned to the Subordinate training group (11 females age M = 22.33, SD = 5.43), twelve to the Basic group (7 females age M = 20.92, SD = 3.91) and twelve to a No-training Control group (6 females, age M = 26.17, SD = 5.75). All the participants reported normal or corrected to normal vision and two were left handed.

The Stimuli, Apparatus, Design, and Procedure were similar to Experiment 3 with the difference being that the training phase was extended from three to four hours, and in the pre-test phase we presented each stimulus only once, equalling the total of 18 trials.

Training

The training procedure was extended by one hour and the total amont of individual training was four hours.

Pre- and post-tests

The pre-and post-tests were similar to the previous experiment (Classification 2), with the difference that each object was presented only once (18 trials).

7.1. Behavioural data analyses

Training data – The Basic and Subordinate group demonstrated learning and improved performance over the four training sessions. For the Subordinate group (see Figure 47), Session 2 produced a mean improvement of 19.96 ms (*SD* = 37.49), this performance gain increasing gradually to 135.01 ms (*SD* = 32.02) by Session 3; a total improvement of 221.45 ms (*SD* = 46.73) overall. Response accuracy also improved commensurately from 68%, to 91%, 96%, and 97% for Sessions 1, 2, 3 and 4 respectively. Wilcoxon tests showed this trend of improved response performance to be significant between Sessions 1 and 2, Z = -3.074, p = .002, Sessions 2 and 3, Z = -2.299, p = .021, and Sessions 1 and 4, Z = -3.063, p = .002. Paired sample t-tests showed a trend of faster RT performance to be significant between Sessions 1 and 3, t (11) = 3.24, p = .008; Session 3 and 4, t (11) = 2.47, p = .031; and overall between Session 1 and 4, t (11) = 3.50, p = .005.

A similar pattern of results was obtained for the Basic group (see Figure 46), with mean response time decreasing from 832.85 ms (*SD* = 170.70), to 700.93 (*SD* = 187.99), and over the four sessions showing an improvement of 131.91 ms between first and last sessions. Response accuracy also improved commensurately from 96%, in session 1 to 98% in sessions 2, 3, and 4. Wilcoxon tests showed this trend of

improved response performance to be significant between Sessions 1 and 2, Z = -2.986, p = .003. Paired sample t-tests of the mean reaction time showed a trend of improved response performance to be significant between Sessions 1 and 2, t (11) = 7.13, p < .0001, and overall between Sessions 1 and 4, t (11) = 4.92, p < .0001.

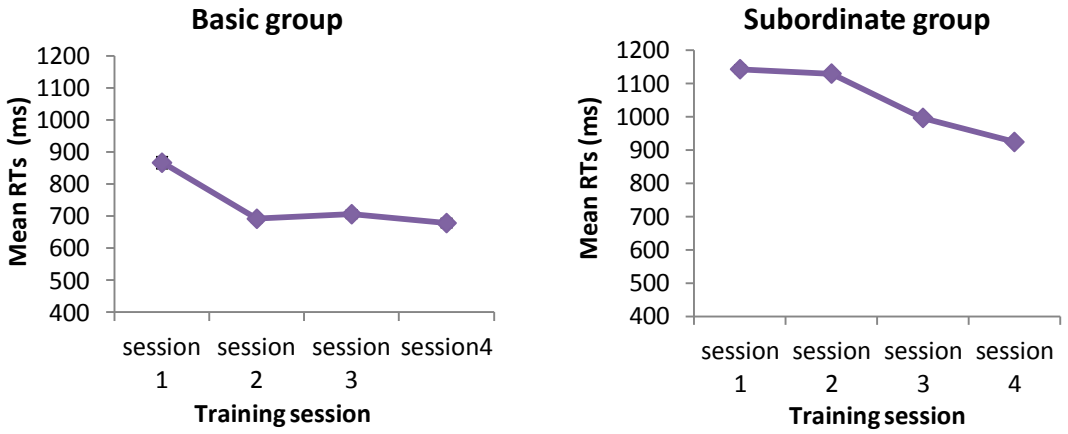


Figure 46 Mean RTs for Basic and Subordinate group across four training sessions. Bars show standard error of the mean.

A mixed 2 x 4 ANOVA with *Between Subject* factor of experimental Group and *Within Subjects* of Session showed a significant main effect of Group F (1, 11) = 21.91, p =.001, η_p^2 = .666, and Session, F (3, 33) = 14.88, p < .0001, η_p^2 = .575. There was also a significant interaction between Group and Session, F (3, 33) = 8.03, p <.0001, η_p^2 = .422. Post-hoc analyses of mean RTs between both groups were significant for session 1, 2, 3, and 4, p = .005, p < .0001, p = .001, p =. 003 respectively. These results show a significant improvement in RTs within and between the two groups. The group trained to categorise objects to a basic level was significantly faster than the group trained to subordinate categorisation in all the training sessions, in line with previous findings proposing that the basic level of categorisation is the entry level in object categorisation (e.g. Tanaka & Taylor, 1991).

Sequential matching task: pre and post tests

Accuracy

The data from four participants in the Subordinate group and one participant from Control group were excluded because of low accuracy rates (< 60%). The overall accuracy for Subordinate, Basic and Control group can be seen in Table 14 below. All three groups performed more accurately in the Basic test (82.8 %; *SE* =.036) than in the Subordinate test (89.7%; *SE* = .033) for both pre- and post-tests.

Table 14 The accuracy rates for Subordinate, Basic, and Control group for each phase (pre-test and post-test), and Test-type (subordinate and basic). Standard error of the mean is shown in parentheses.

	Pre-test phase		**Post-test phase**	
	Subordinate	**Basic**	**Subordiante**	**Basic**
	Mean accuracy (%)		Mean accuracy (%)	
Subordinate group	.76 (.043)	.83 (.030)	.88 (.035)	.92 (.061)
Basic group	.75 (.041)	.91 (.023)	.91 (.037)	.99 (.007)
Control group	.80 (.033)	.85 (.038)	.87 (.026)	.88 (.039)

Reaction times

The data from five participants from the Subordinate group was excluded from the analyses. The reason for this was low accuracy performance in four of the cases (< 60%), and lack of training effects in the remaining one case. Similarly, one participant with less than 60% correct performance from the Control group was also excluded from the analyses. For the remaining data only correct trials were analysed.

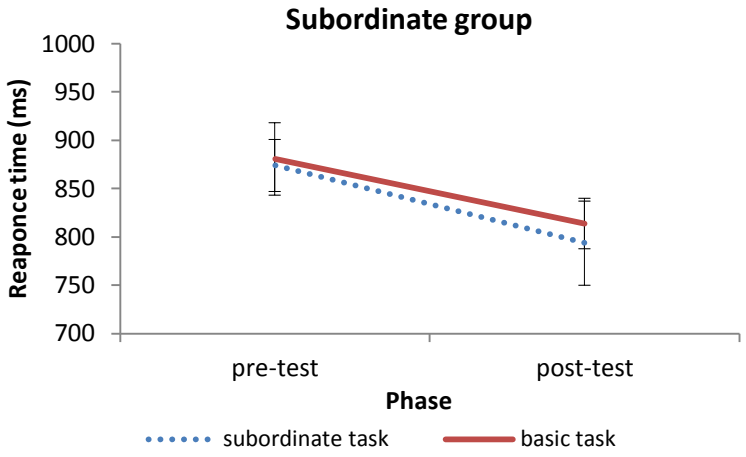
The response performance on pre-test and post test scores for Subordinate Basic, and Control groups can be seen in Table 15 below:

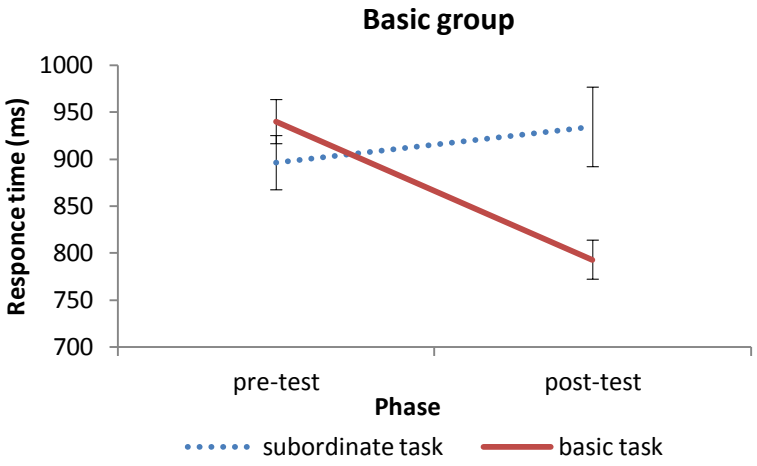
Table 15 The mean reaction time for Subordinate, Basic, and Control group for each phase (pre-test and post-test), and Test-type (subordinate and basic). Standard error of the mean is shown in parentheses.

	Pre-test phase		**Post-test phase**	
	Subordinate	**Basic**	**Subordinate**	**Basic**
	Mean RT (SE)	Mean RT (SE)	Mean RT (SE)	Mean RT (SE)
Subordinate group	873.93 (26.93)	880.68 (37.47)	793.50(43.58)	813.82 (26.15)
Basic group	896.10 (28.83)	939.76 (23.46)	934.19(42.34)	793.08 (20.77)
Control group	851.86 (61.59)	850.83 (31.49)	902.55(42.88)	869.83 (38.96)

The Subordinate group difference in performance for pre-post subordinate tests was 80.43 ms, and 66.86 ms difference for basic tests. The Basic group difference in performance for pre-post subordinate tests was -38.09 ms, and 146.68 ms for basic tests (see Figure 47).

This shows that both Basic and Subordinate groups had an effect of training as each group responded quicker and more accurately on the test-type relevant to their training. The Control group did not show improved performance in any of the post tests.


Subordinate group
Reaponce time (ms)
1000
950
900
850
800
750
700
pre-test
post-test
Phase
subordinate task
basic task


Basic group
Responce time (ms)
1000
950
900
850
800
750
700
pre-test
post-test
Phase
subordinate task
basic task

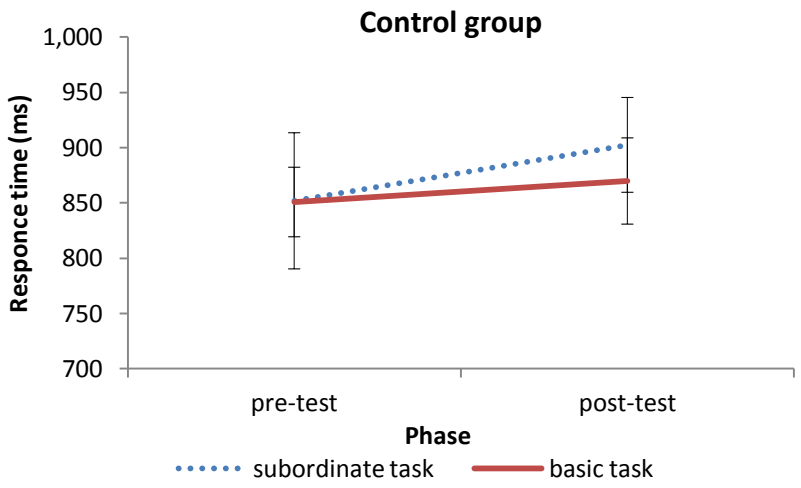


Figure 47 Mean RTs per group, for each phase (pre-test and post-test) for each Test-type (subordinate and basic). Bars show standard error of the mean (%).

A within group 2 x 2 ANOVA for the Subordinate Group showed a significant main effect of Phase, F (1, 6) = 11.31, p =.015, η_p^2 = .653. There were no other main effects or interactions. Paired t- tests showed no significant difference between the subordinate pre and post test, p = 0.190ns, and between the basic pre-and post-test, p =0.149ns.

A within group 2 x 2 ANOVA for the Basic Group showed a significant main effect of Phase, F (1, 11) = 5.02, p =.047, η_p^2 = .313, and an interaction between Phase and Test-type, F (1, 11) = 9.37, p =.011, η_p^2 = .460. There were no other main effects and interactions. Paired t-tests showed a significant difference between the basic pre and post test, p <.0001, which was not evident between the subordinate pre and post tests, providing evidence that the Basic group improved their performance selectively as a function of training.

A within group 2 x 2 ANOVA for the Control Group showed no significant main effect of Phase, F (1, 10) = .569, p =.469ns, η_p^2 = .054, Test-type, F (1, 10) = .161, p =.696ns, ηp^2 = .016, and no interactions.

A mixed 3 x 2 x 2 analysis of variance (ANOVA) with a *Between Subject* factor of experimental Group and a *Within Subject* factor of Phase (pre-test or post-test) x Test-type (Subordinate or Basic) showed no significant main effects or interactions.

These results provide evidence that both training groups changed their performance following training, whereas the Control group did not show any significant changes between the pre-test and post-test. The latter finding is different compared to the results in Experiment 3 where the Control group showed a significant change between pre and post tests despite the lack of training. One possible explanation is that the number of object presentations (218) for each pre and post test could have resulted in priming effects which are not evident in the current study due to reduction of the number of the objects presented (18) in both tests.

7.1.2. Analyses of eye movement data

Analyses between groups

A 3 (Group: Subordinate, Basic, Control) x 2 (Phase: pre, post) x 2 (Test-type: Subordinate, Basic) x 3 (Model: Concave, Convex, Visual saliency) ANOVA showed a significant main effect of Model, F (2, 102) = 68.41, $p < .0001$, $\eta_p^2 = .573$, and Test-type, F (1, 51) = 15.00, $p < .0001$, $\eta_p^2 = .227$, along with a two way interaction between Model and Phase, F (2, 102) = 10.94, $p < .0001$, $\eta_p^2 = .177$. There were no other main effects or interactions. Post-hoc pairwise comparisons (Bonferroni) between each model were all significant for Concave vs. Convex, $p < .0001$, Concave vs. Visual saliency, $p < .0001$, and Convex vs. Visual saliency, $p = .001$.

The above Model vs. Phase interaction implies that some of the three groups have an effect of training while the other did not. To investigate this further, we performed within group analyses.

Analyses within group

Basic group

A 2 (Model: Concave, Convex) x 2 (Phase: pre, post) x 2 (Test-type: Subordinate, Basic) within subjects ANOVA showed a significant main effect of Model, F (1, 17) = 19.81, $p < .0001$, $\eta_p^2 = .538$, along with a two-way interaction between Model and Phase, F (1, 17) = 7.84, $p = .012$, $\eta_p^2 = .316$. There were no other main effects or interactions (see Figure 48). Post-hoc pairwise comparison of pre-test phase (Bonferroni) between both models was significant for basic, $p < .0001$, and subordinate, $p < .0001$ tests respectively. This indicates that the concave preference was less at post-training.

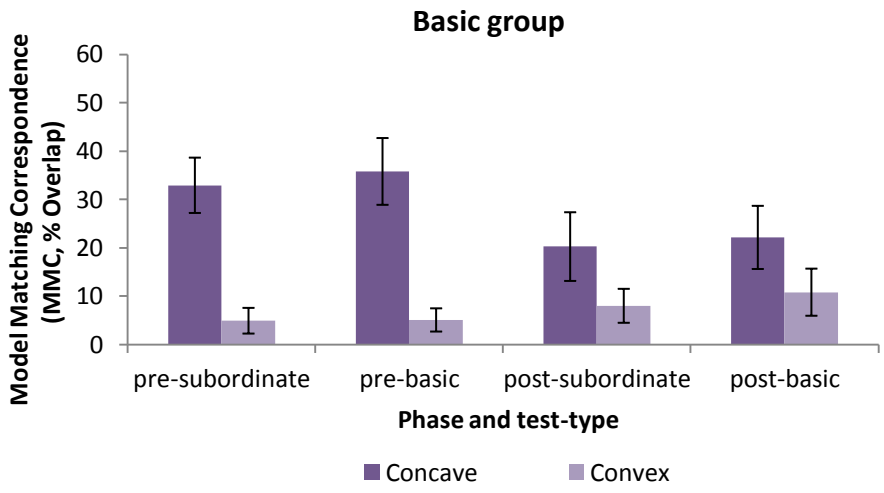


Figure 48 MMC for Basic group for each phase (pre-test and post- test) and each Test-type (subordinate and basic) for concave and convex models relative to visual saliency. Bars show standard error of the mean (%).

Subordinate Group

A 2 (Model: Concave, Convex) x 2 (Phase: pre, post) x 2 (Test-type: Subordinate, Basic) within subjects ANOVA showed a significant main effect of Model, F (1, 17) = 10.54, p = .005, η_p^2 = .383, along with two-way interaction between Model and Phase, F (1, 17) = 6.44, p = .021, η_p^2 = .275. There were no other main effects or interactions (see Figure 49). Post-hoc pairwise comparison for pre-test (Bonferroni) between both models was significant for the basic, p = .002, and subordinate, p = .006 tests respectively. This indicates that the concave preference was less at post-training.

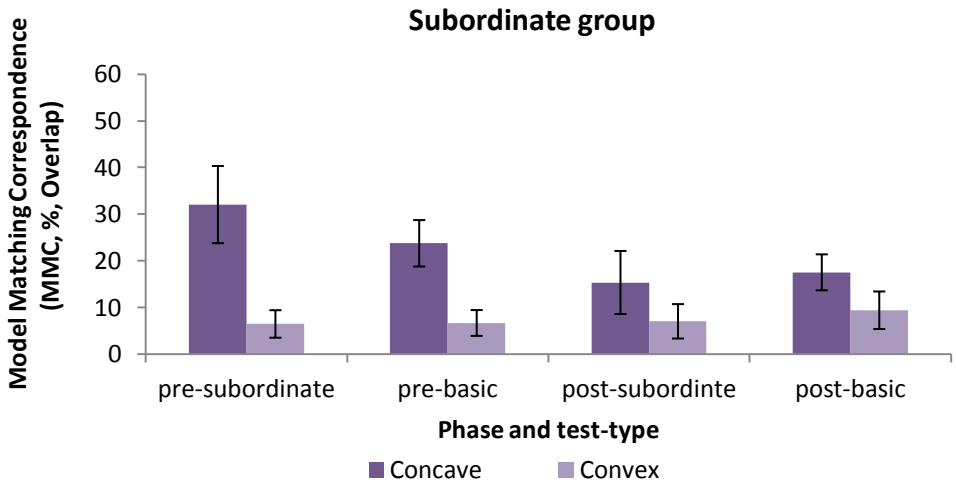


Figure 49 MMC for Subordinate group for each phase (pre-test and post-test) and each Test-type (subordinate and basic) for concave and convex models relative to visual saliency. Bars show standard error of the mean (%).

Control Group

A 2 (Model: Concave, Convex, Visual saliency) x 2 (Phase: pre, post) x 2 (Test-type: subordinate, basic) within subjects ANOVA showed a significant main effect of Model, F (1, 17) = 17.40, p < .0001, η_p^2 = .506. There were no other main effects or interactions (see Figure 50). Post-hoc pairwise comparison for pre-test (Bonferroni) between both models was significant for basic, p =.027, and subordinate,

p =.002 tests. There was also a significant pairwise comparison for the post-subordinate test, *p* =.002.

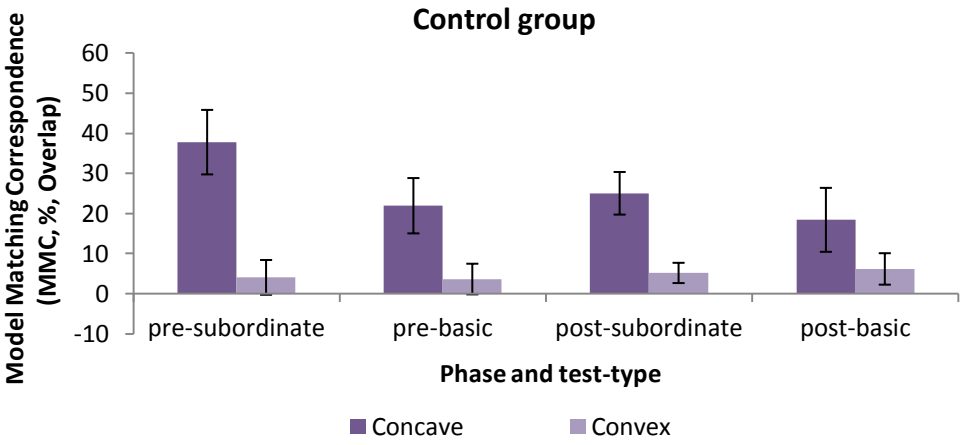


Figure 50 MMC for Control group for each phase (pre-test and post-test) and each Test-type (subordinate and basic) for concave and convex models relative to visual saliency. Bars show standard error of the mean (%).

In order to test whether there is a difference in the fixation data across models regardless of training, we collapsed the data across group for each test (Subordinate, Basic) and phase (pre-test, post-test), (see Figure 51). A 2 (Phase: pre, post) x 2 (Test-type: Subordinate, Basic) x 3 (Model: Concave, Convex, Visual saliency) ANOVA showed a significant main effect of Model, F (2, 34) = 37.07, p < .0001, η_p^2 = .686, and Test-type, F (1, 17) = 13.95, p = .002, η_p^2 = .451, along with a two way interaction (Model and Phase), F (2, 34) = 8.90, p = .001, η_p^2 = .344. Post-hoc pairwise comparisons (Bonferroni) between each model were significant for Concave vs. Convex, p <.0001, Concave vs. Visual saliency, p <.0001, and for Convex vs. Visual saliency, p =.009.

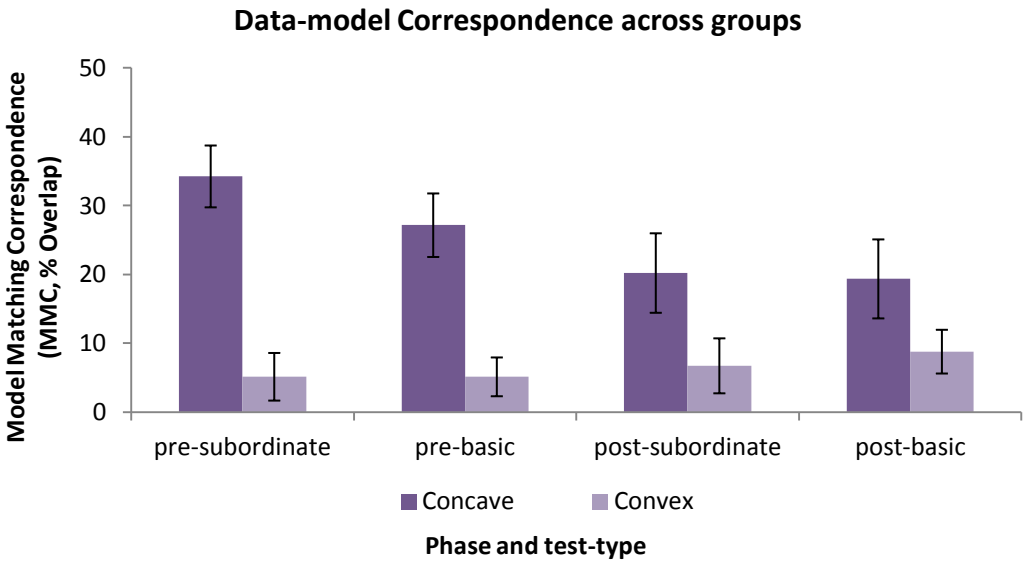


Figure 51 MMC collapsed across groups for each phase (pre-test and post-test) and for each Test-type (subordinate and basic) for concave and convex models relative to visual saliency. Bars show standard error of the mean (%).

7.1.3. Saccade amplitude and dwell time analyses

Saccade amplitude

The overall mean saccade amplitude for pre-mask Image 1 for the Subordinate and Basic groups can be seen in Table 16 below.

Table 16 The mean saccade amplitude (SA) for pre-mask Image 1 for the Subordinate, Basic, and Control group for each phase (pre-test and post-test), and Test-type (subordinate and basic). Standard error of the mean is shown in parentheses.

	Pre-test phase		**Post-test phase**	
	Subordinate	**Basic**	**Subordinate**	**Basic**
	Mean SA (SE)	Mean SA (SE)	Mean SA (SE)	Mean SA (SE)
Subordinate group	3.83 (0.47)	3.51 (0.45)	3.50 (0.41)	3.17 (0.50)
Basic group	4.11 (0.39)	3.40 (0.28)	3.70 (0.41)	2.99 (0.29)

The overall Mean saccade amplitude for post-mask Image 2 for the Subordinate and Basic groups can be seen in Table 17 below.

Table 17 The mean saccade amplitude (SA) for post-mask Image 2 for the Subordinate, Basic, and Control group for each phase (pre-test and post-test), and Test-type (subordinate and basic). Standard error of the mean is shown in parentheses.

	Pre-test phase		**Post-test phase**	
	Subordinate	**Basic**	**Subordinate**	**Basic**
	Mean SA (SE)	Mean SA (SE)	Mean SA (SE)	Mean SA (SE)
Subordinate group	3.94 (0.72)	1.97 (0.40)	3.40 (0.57)	1.78 (0.38)
Basic group	4.11 (0.56)	2.56 (0.53)	3.85 (0.53)	2.37 (0.38)

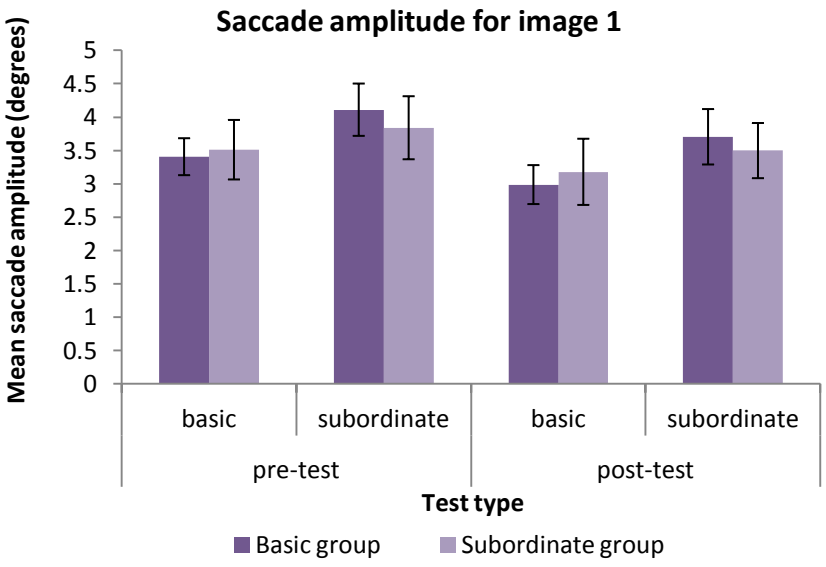
A mixed 2 x 2 x 2 x 2 analysis of variance (ANOVA) with a *Between Subject* factor of experimental Group and *Within Subject* of Phase (pre-test or post-test) x Test-type (Basic or Subordinate) x Image (pre-mask Image 1 or post-mask Image 2) showed a significant main effect of Phase, F (1, 17) = 6.66, p =.029 η_p^2 = .250, and Image, F (1, 17) = 17.63, p =.001, η_p^2 = .509.The data also showed a two way interaction, between Phase and Image, F (1, 17) = 16.85, p =.001, η_p^2 = .498.

A within group 2 x 2 x 2 ANOVA for the Basic Group showed a significant main effect of Image, F (1, 11) = 11.97, p =.005, η_p^2 = .521. The data also showed a two way interaction, between Phase and Image, F (1, 11) = 6.91, p =.023, η_p^2 = .386. Paired t-tests between Image 1 (see Figure 52 a) and Image 2 (see Figure 52 b) showed a significant difference for the basic post-test, (Image 1, M = 2.99, SD = 1.01, Image 2,

M = 2.37, *SD* = 1.29), *p* =.027. There were no other significant differences between pre-and post tests.

A within group 2 x 2 x 2 ANOVA for the Subordinate Group showed a significant main effect of Image, F (1, 7) = 7.31, p =.035, η_p^2 = .547. The data also showed a two way interaction, between Phase and Image F (1, 7) = 8.49, p =.027, η_p^2 = .586. Paired t- tests between Image1 and Image 2 showed a significant difference for the basic pre-test Image 1, (*M* = 3.51, *SD* = 1.55), and basic pre-test Image 2, (*M* = 1.78, *SD* = 1.37), *p* =.021. There were no other significant differences between Image 1 and Image 2 for pre-and post tests.

(a)



(b)

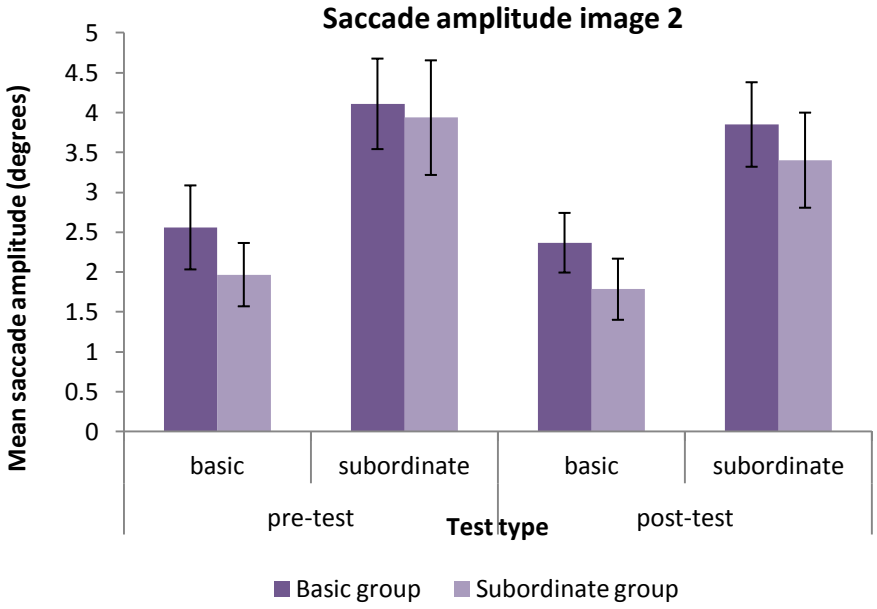


Figure 52 a) Mean saccade amplitude (degrees) for Basic and Subordinate group for each phase (pre-test and post-test) and Test-type (subordinate and basic) for Image 1. Bars show standard error of the mean (%). (b) Mean fixation amplitude (degrees) for Basic and Subordinate group for each phase (pre-test and post-test) and Test-type (subordinate and basic) for Image 2. Bars show standard error of the mean (%).

Dwell time analyses

The overall mean dwell times for pre-mask Image 1 for the Subordinate and Basic groups can be seen in Table 18 below.

Table 18 The mean dwell times (DT) for pre-mask Image 1 for the Subordinate and Basic for each phase (pre-test and post-test), and Test-type (subordinate and basic). Standard error of the mean is shown in parentheses.

	Pre-test phase		**Post-test phase**	
	Subordinate	**Basic**	**Subordinate**	**Basic**
	Mean DT (SE)	Mean DT (SE)	Mean DT (SE)	Mean DT (SE)
Subordinate group	179.25 (8.02)	182.63 (9.81)	188.95 (10.88)	209.63 (15.64)
Basic group	168.31 (5.71)	180.48 (8.67)	166.70 (6.90)	165.85 (7.61)

The overall Mean dwell times for pre-mask Image 2 for the Subordinate and Basic groups can be seen in Table 19 below.

Table 19 The mean dwell times (DT) for post-mask Image 2 for the Subordinate and Basic group for each phase (pre-test and post-test), and Test-type (subordinate and basic). Standard error of the mean is shown in parentheses.

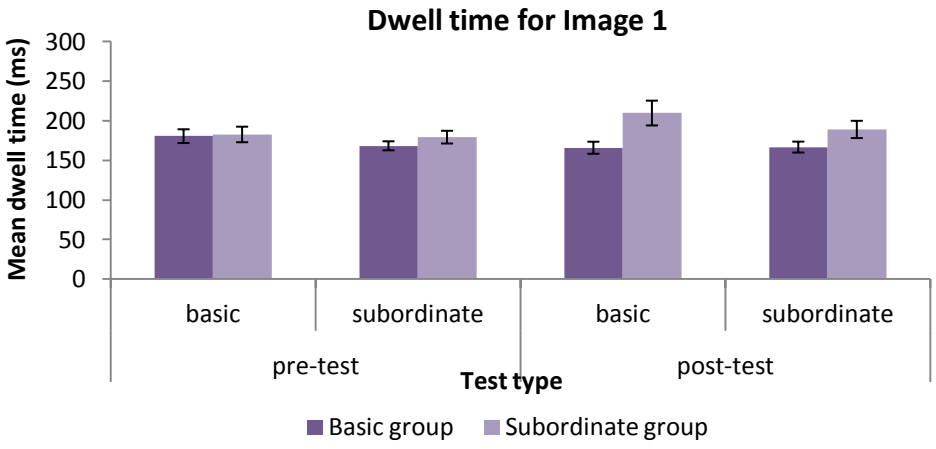
	Pre-test phase		**Post-test phase**	
	Subordinate	**Basic**	**Subordinate**	**Basic**
	Mean DT (SE)	Mean DT (SE)	Mean DT (SE)	Mean DT (SE)
Subordinate group	218.31 (8.90)	207.98 (12.25)	213.41 (14.92)	199.41 (16.72)
Basic group	195.69 (10.50)	237.64 (34.66)	193.46 (13.55)	207.02 (12.25)

A mixed 2 x 2 x 2 x 2 analysis of variance (ANOVA) with a *Between Subject* factor of experimental Group and *Within Subject* of Phase (pre-test or post-test) x Test-type (Basic or Subordinate) x Image (pre-mask Image 1 or post-mask Image 2) showed a significant main effect of Phase, F (1, 17) = 11.77, p =.003,η_p^2 = .409. The data also showed a two way interaction, between Phase and Test-type, F (1, 17) = 4.89, p =.041, η_p^2 = .223.

A within group 2 x 2 x 2 ANOVA for the Basic Group showed a significant main effect of Image, F (1, 11) = 9.34, p =.011, η_p^2 = .459. There were no other main effects or interactions. Paired t- tests between Image 1 and Image 2 showed a significant difference for the subordinate pre test (Image 1, M = 168.31, SD = 19.79, Image 2, M = 195.68, SD = 36.35), p =.004, and the basic post test Image 1, M = 165.84, SD = 26.36, Image 2, M = 207.02, SD = 42.42), p =.001. There were no other significant differences between pre-and post tests.

A within group 2 x 2 x 2 ANOVA for the Subordinate Group showed a significant main effect of Image, F (1, 7) = 20.50, p =.004, η_p^2 = .774. The data also showed a two way interaction, between Phase and Image, F (1, 7) = 6.82, p =.040, η_p^2 = .532. Paired t- tests between Image 1 (see Figure 53 a) and Image 2 (see Figure 53 b) showed a significant difference for the subordinate pre test, (Image 1, M = 179.25, SD = 27.76, Image 2, M = 218.31, SD = 30.84), p <.0001. There were no other significant differences between pre-and post tests.

(a)



(b)

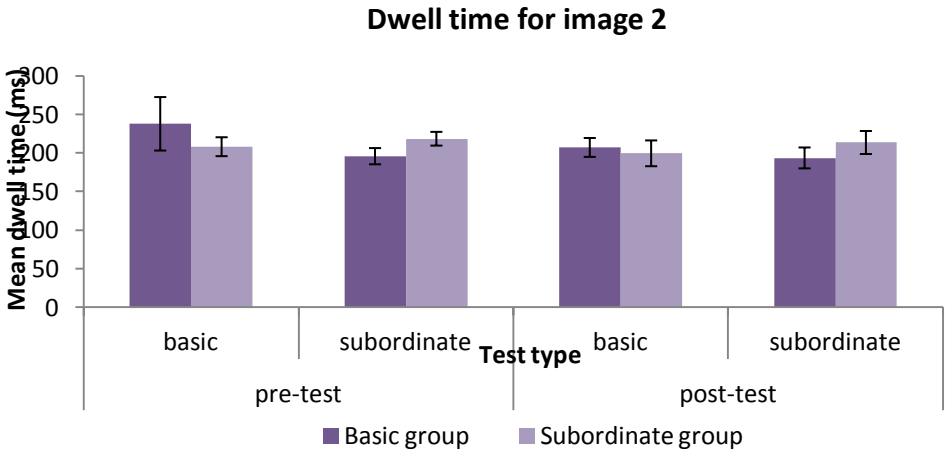


Figure 53 a) Mean dwell time (ms) for Basic and Subordinate group for each phase (pre-test and post-test) and Test-type (subordinate and basic) for Image 1. Bars show standard error of the mean (%). (b) Mean dwell time (ms)) for Basic and Subordinate group for each phase (pre-test and post-test) and Test-type (subordinate and basic) for Image 2. Bars show standard error of the mean (%).

Summary of the Results

Training data

- Both training groups significantly improved their RTs between Session 1 and Session 4, and demonstrated over 91% accuracy post training.

Behavioural data

- The RTs showed that each training group improved their post-test mean performance as a function of training, but only the Basic group showed a significant specific effect in the basic task. The Control group did not improve their performance in the post-basic test.

Eye movement data

- The Basic training group showed a main effect of Model. There was also a two-way interaction between Model and Phase. The data-model correspondence was higher for the Concave model than for the Convex model, but was only significant in pre-tests indicating less post-test concave preference.
- The Subordinate training group eye movement results showed a significant main effect of Model. There was also a two-way interaction between Model and Phase. The data-model correspondence was higher for the Concave model than for the Convex model, but was only significant in pre-tests indicating less post-test Concave preference.
- The Control group showed a significant main effect of Model. The data-model correspondence was significantly higher for the Concave model compared to the Convex model, in both pre-tests, as well as in the post-subordinate test.

- Across groups analyses for each test-type and phase showed main effects of Model and Test-type along with a two way interaction between Phase and Model.
- Data-model correspondences between each model were significantly higher for the Concave model, than the Convex and the Visual saliency models.
- **Saccade amplitude analyses**
- Within group analyses for the sequential matching tests for the two training groups (Basic and Subordinate) showed that each group had a significantly higher saccade amplitude for Image 1 than for Image 2, suggesting that the participants changed their viewing strategies between the initial encoding and the subsequent recognition.

Dwell time analyses

- Within group analyses for the sequential matching tests for the Basic group and Subordinate groups showed a main effect of Image as well as significantly shorter dwell times for Image 1 than for Image 2, showing different viewing strategies between the initial encoding and the subsequent recognition.

7.1.4. Conclusions

In the current study, similarly to Experiment 3, we used fixational eye movement patterns to examine local shape analysis processes during object shape categorisation in experts and novices. In pre-test phases, observers sequentially matched visually similar novel objects to a basic and subordinate level of classification. Post-training, the basic and subordinate experts performed the same tasks as in the pre-test. In this experiment the training was extended from three hours (in Experiment 3), to four hours. Observers at the basic-training group were faster and more accurate at making basic judgment following training. Observers at the subordinate-training group were faster and more accurate at making subordinate

judgment following training. The analyses of RT and accuracy data showed that both groups improved their performance as a function of training, although only the basic training group showed specific training effects. The Control group showed no improved performance in both post-tests.

The main finding revealed a consistent pattern of data-model correspondences across tasks. During both basic and subordinate matching tasks in pre-test we found evidence that fixation patterns are predominantly driven by shape information defined by internal regions of concave surface discontinuity regardless of expertise or level of categorisation (basic or subordinate). The data are consistent with models of object recognition which posit a special functional status to concave image regions in object representation (e.g., Biederman, 1987; Hummel & Stankiewitz, 1996; Leek, Reppa, Rodriguez & Arguin, 2009; Leek, Reppa & Arguin, 2005; Marr & Nishihara, 1978), and they present a challenge to models which do not. Among the latter models are some recent Image-based models of recognition (e.g., Edelman & Weinshall, 1991; Riesenhuber & Poggio, 2006; Ullman & Basri, 1991), including HMAX (Serre, Oliva & Poggio, 2007; Serre, Wolf, Bileschi, Riesenhuber & Poggio, 2007). In order to account for these data, these models would require modification to include a level of feature representation which makes explicit the sign of curvature – for example, in the C2 layer of HMAX.

Interestingly, a contradictory to Experiment 3 finding is that the concave preference was not evident at post-test where no significant difference was found between concave and convex models. One possible explanation of this finding is the suggested two-stage model of perceptual category learning (Ashby & Spiering, 2004; Nosofsky, 1986; Riesenhuber & Poggio, 2000; Sigala, 2004; Thomas et al., 2001). In this model, high level shape representation can be activated for different tasks using

the same stimuli in order to improve discrimination of these stimuli relative to untrained participants (Jiang et. al., 2007). In Experiment 4 it appears that both types of local image curvature information (concave and convex) have been employed during the post sequential matching task.

Although a large body of research supports the special status of curvature discontinuities in object representation, and a privileged role has been proposed for concave curvature discontinuities (e.g. Attneave, 1954; Barenholtz, Cohen, Feldman & Singh, 2003; Bertamini, 2008; Biederman, 1987; Cate & Behrmann, 2010; Cohen, Barenholtz, Singh & Feldman, 2005; Cohen & Singh, 2007; De Winter & Wagemans, 2006; Feldman & Singh, 2005; Hoffman & Richards, 1984; Hoffman & Singh, 1997; Lim & Leek, 2012) there is a number of studies proposing that convex curvature discontinues also have an important role. For example, resent psychophysical study demonstrated that high level visual cortex areas (LOC) are more sensitive to changes in convex than in concave shapes (Houshofer, Baker, Livingstone & Kanwisher, 2008). Moreover, previous research found that recognition performance (e.g. Driver & Baylis 1995), perceptual judgment of location (e.g. Bertamini, 2001), and shape similarity (Subirana-Vilanova & Richards 1996) are more accurate for convex, rather than concave image features.

However, possible factors for the apparent concave/convex post-test preference which was not evident in Experiment 3, could be attributed to the changes made in the current study in terms of extended training time and reduced number of image exposures in pre-and post-test. Nevertheless, investigating the exact magnitude and reasons for changes in preference between concave and convex curvature during different task requirements (e.g. number of trials, training trials) is out of the scope of this thesis, although it would be addressed in more detail in further studies.

These findings are to some extent consistent with the results from Experiment 3 and add to the importance of curvature singularities in the visual perception of shape (e.g., Attneave, 1954; Barenholtz Cohen & Singh, 2003; De Winter & Wagemans, 2006; Feldman & Singh, 2005; Hoffman & Richards, 1984; Hoffman & Singh, 1997). The previously found preference for fixation at regions of concave surface discontinuity regardless of task (basic vs. subordinate) or phase (pre vs. post) was not entirely replicated in this study, and this provides further information of the complexity during object shape representation.

The analyses of the saccade amplitude and dwell time showed no significant differences post-training. However, saccade amplitude was shorter and dwell time longer in Image 2, showing the expected pattern associated with recognition tasks. Moreover, the lack of significant change between pre and post tests was consistent in both types of analyses (e.g. saccade amplitude/dwell times and data-model correspondence), showing the systematic nature of the findings.

Chapter 8

8.1. Experiment 5

In the previous four experiments we have found how eye movement patterns can be used to elucidate shape analysis strategies during a variety of visual perception tasks (active learning, passive viewing, and basic or subordinate level of categorisation). However these tasks did not involve any form of motor interaction, such as imaginary reaching or grasping the objects. The goal of the current short study was to extend the analyses of the concavity effect across tasks and to examine how eye movements could be used to elucidate differences in shape analysis strategies in tasks of object recognition and the planning of imaginary grasping actions.

One dominant, but still controversial hypothesis attempting to explain how we transform the information from the visual input into motor acts is the two-stem hypothesis (Ungerleider & Mishkin, 1982; Goodale &Milner, 1992). This hypothesis was originally proposed by Ungerleider and Mishkin (1982) who suggested that the two visual pathways arising from V1 (ventral stream and dorsal stream) are involved in different aspects of the visual perception. The authors proposed that the ventral stream (also called the 'what' pathway) was involved in processing of visual features for object identification, whereas the dorsal stream (also called the 'where' pathway) was involved in processing the objects spatial location. Underleider and Mishkin (1982) proposed the theoretical possibility of objects identification and spatial location to be initially processed together followed by separate analyses in the two different pathways, and ultimately reintegrated. There is some neuropsychological evidence suggesting the possibility of two pathways for processing visual information. For

example, Levine, Warach and Farh (1985) reported dissociations between images for visual and spatial information in ventral and dorsal lesion patients.

Goodale and Milner (1992) proposed a somewhat different hypothesis, suggesting that the ventral stream is primarily involved in perceptual object representation, whereas the dorsal stream mediates the visual control for directing actions towards objects. The authors suggested that visual object information (i.e., object features) is processed by both streams, but used for different purposes; hence the input is the same but the output is different. Moreover, this suggests differential operations and transformations on the input. Hence, it can be assumed that in the ventral stream the object features are transformed to create a representation of an object, whereas in the dorsal stream the same object features are transformed in a different way in order to direct action to that object (James & Kim, 2010). A number of neuropsychological studies (Jennerod, 1988; Perenin & Vighetto, 1988) provided support for the dorsal stream function proposed by Goodale and Milner (1992). For example, patients with parietal lesions were reported to have difficulties in forming an appropriate grasp size, or orientation of objects, although they were able to correctly describe the object spatial location.

However, although both Ungerleider and Mishkin (1982) and Goodale and Milner (1992) disagree on the dorsal stream's function, they agree that the ventral stream is involved in object representation.

Relatively few studies have investigated gaze patterns of participants performing both perception and action tasks to the same objects. Brouwer, Franz, and Gegenfurtener (2009) investigated fixation locations during either grasping or viewing tasks to novel shapes. The authors reported that first saccades were directed towards the Centre of Gravity (COG) in both tasks whereas during the grasping task, they

focused on the upper part of the object associated with an index finger location (Desanghere & Marotta, 2008; de Grave, Hesse, Browler & Franz, 2008).

Similar findings were demonstrated in a recent study (Desanghere & Marotta, 2011) examining the fixation patterns for computer generated and real objects, in a reaching/grasping and object perception task. The authors' reported similar fixation patterns for both tasks. In the grasping task the participant's initial fixations were directed towards the top edge of the object, which is considered to be a part associated with index finger location. Subsequently the fixations were directed towards the COG. This fixation pattern was reversed during the perceptual task, thus the initial fixations were directed towards COG, followed by fixation to the top edge of the object. Although the sequence of the observed fixation patterns was different, the object areas covered occurred between COG and the top edge of the objects during all the tasks. The Desanghere and Marott (2011) study provides evidence for consistent fixation locations during object perception and action; however, it is not clear what local image properties were located in the areas fixated.

Gaze control is a considered to be an active process where relevant information is extracted from the environments in order to complete the task in hand (Henderson, 2003) and previous research has shown that eye movement in general precede hand movements in a variety of tasks such as pointing (van Donkelaar et al., 2004), or manipulation of objects (Hayhoe & Ballard, 2005).

Although it is to some extend an over-simplification of the true functional significance of the visual cortex, one interesting theoretical question is whether there is a single object representation and its information is used in a different way to fit with task demands (e.g. recognition or reaching), or whether there are separate object representations for different tasks.

Thus, the question we ask is whether fixation patterns will vary across the different tasks (e.g. motor-imagery, recognition) and what local image properties are predominantly observed in the motor-imagery task.

8.1.1. Method

Participants

30 students from Bangor University (21 female, mean age 22.00 years, *SD* = 7.07 (3 right handed) participated in the study for course credit. All participants had normal or corrected to normal visual acuity. Informed consent was obtained from each participant prior to testing in line with local ethics committee and BPS guidelines.

Stimuli

Each of the twenty four novel objects (for example see Figures 54 and 55) consisted of a unique spatial configuration of four volumetric parts. The parts were uniquely defined by variation among non-accidental properties (NAPs) comprising: Edges (Straight vs. Curved), symmetry of the cross section, tapering (collinearity) and aspect ratio (Biederman, 1987).

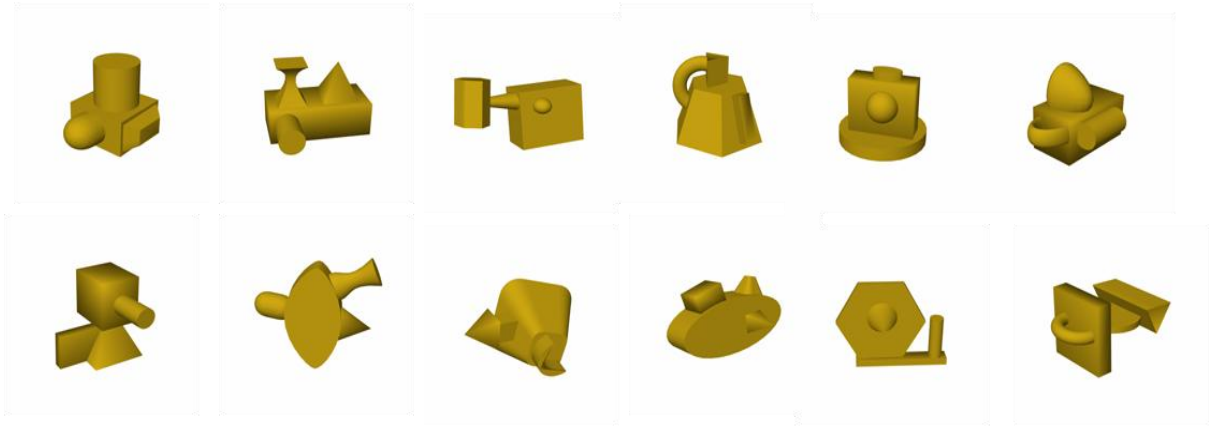


Figure 54 The 12 novel object stimuli used as targets in the current study.

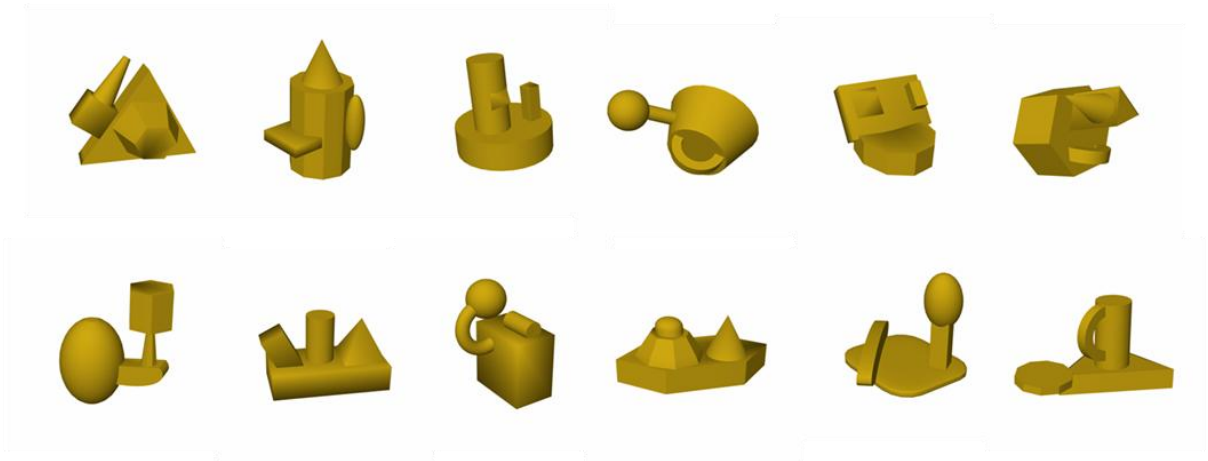


Figure 55 The 12 novel object stimuli used as distracters in the current study.

These object models were produced using Strata 3D CX software (Strata, USA) and rendered using a single light source (top left) and scaled to fit within an 800 x 800 pixel frame. All stimuli were uniformly coloured in mustard yellow: R = 227, G = 190, B = 43; Hue = 32, Saturation = 184, Luminance = 127. A ground shadow was removed from the images. Stimuli subtended 18 degrees of visual angle horizontally from a viewing distance of 60 cm. This scale was chosen to induce saccadic exploration over the stimuli. Versions of each model were created depicting the object from each of six different viewpoints at successive 60 degree rotations in depth around a vertical axis perpendicular to the line of sight. The zero degree viewpoint was a 'canonical' three-quarter view (see Figure 56). The 0, 120 and 240 degree versions served as training viewpoints, and the 60, 180 and 300 degree versions as novel test viewpoints.

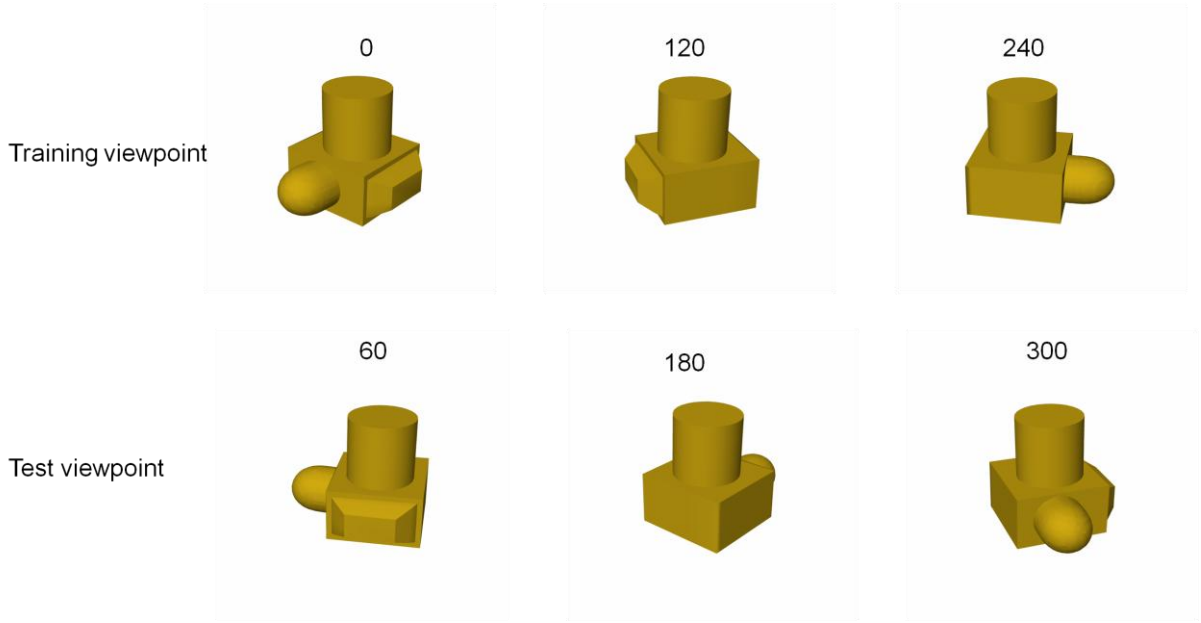


*Figure 56*An illustration of the three trained (0, 120, 240) and three novel viewpoints (60, 180, 300) used.

Apparatus

A Tobii 1750 eye tracking system was used to record eye-movement data. This apparatus allows for high precision binocular tracking with 0.5 degrees accuracy, 0.25 degrees spatial resolution, and drift < 1 degree. Stimuli were presented on a TFT monitor running at a resolution of 1280 x 1024 pixels and 60 Hz refresh rate. Mean surround luminance was 114.7cd/m2 (SD = 0.25cd/m2) – averaged over two independent runs using a Minolta CS-100 photometer. A chin rest was used to stabilize the participant's head at a 60cm viewing distance and a standard USB keyboard was used for response collection.

Design and Procedure

Initially each participant completed a calibration on the eye-tracker, where a static blue dot appeared randomly in each of 9 possible screen locations. Eye position and known screen position was recorded which allowed a transformation matrix to be constructed (via a linear interpolation method, which was used to determine gaze

position from eye position). Prior to continuing beyond the calibration stage, the calibration results were visually inspected to ensure that a sufficiently good calibration was performed.

The study comprised two phases: Planning and executing an imagined movement to novel objects (Task 1), and Memorisation and recognition of novel objects (Task 2). The Recognition Memory task was divided further into a Memorisation task and Test phase. All subjects completed both phases, and their eye movements were recorded during each task. For counterbalancing, targets were split into two groups. The participants in Group 1 (n = 8) used six of the stimuli (1 to 6) as targets in Task 1, and six stimuli (7 to12) as targets in Task 2. The remaining twelve stimuli were split between the two groups and served as distracters in the Test phase. For Group 2 (n = 8) this assignment was reversed. Participants were randomly assigned to each group.

In the Planning movement phase (36 trials each) participants viewed six objects each at six different viewpoints (0, 60, 120, 180, and 240 degrees). Following fixation at either left or right side of the screen, stimuli were presented centrally for 5 seconds and the participants were asked to plan and imagine picking up the object on the screen by 'using' their thumb and forefinger. Subsequently the participants were required to indicate (by two mouse clicks) where they have placed, first their thumb, and second forefinger (see Figure 57). The participants were told that accuracy is more important than speed.

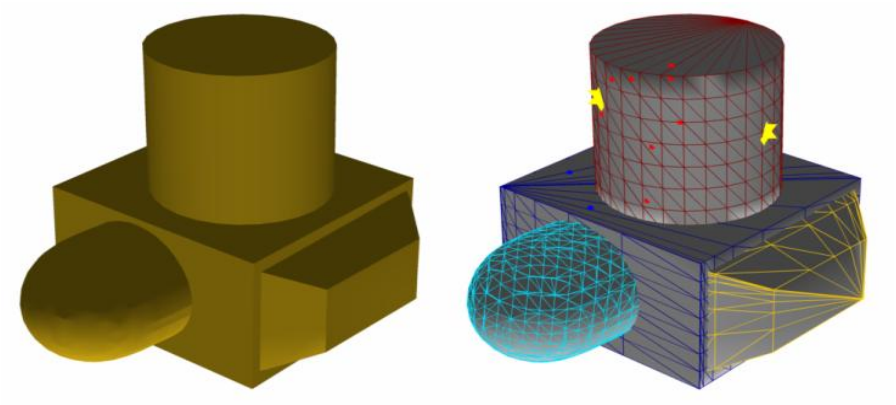


*Figure 57*a) Stimuli as seen by the participants (b) Eye movements (red dots) and mouse clicks (yellow stars) mapped onto the colour coded mesh (single colour per part).

The Recognition Memory phase consisted of two tasks, Active Learning phase and Test Phase. In the Active Learning phase (18 trials each) participants viewed six (target) objects each at three different viewpoints (0, 120 and 240 degrees). The stimuli were presented sequentially for 5 seconds each at the centre of the screen, following a two second fixation at a peripheral location randomly selected either on the left or the right side of the screen.

In the Test Phase (54 trials each) participants were presented with previously seen target objects, plus an additional set of six visually similar, but novel distracters were presented at the trained viewing angles of 0, 120 and 240 degrees and also untrained (60, 180 and 300 degree) viewpoints. Following fixation at either left or right side of the screen, stimuli were presented centrally and were displayed until a response was given. The participants were asked to determine whether the presented 3D object was one of the previously seen objects or not and respond using a key-press (k for 'yes'/ d for 'no'). The participants were told that accuracy is more important

than speed. Fixation, RT, and accuracy data were recorded as dependent measures. The experiment lasted about 30 minutes (including calibration).

8.1.2. Algorithmically generated model predictions

Generating Model Predictions

The predicted distributions for each model of image information content were algorithmically computed from the 3D object models using Matlab as in Experiment 1. An illustration of the predicted thresholded fixation region maps for the tested models can be seen in Figure 21.

Model 1: External (bounding) Contour

Model 1 examined the extent to which fixation patterns focus on external global shape features defined by bounding contour. This hypothesis derives from previous work showing that outline shape influences object recognition (e.g., Hayward, 1998; Hayward et al., 1999; Lloyd-Jones & Luckhurst, 2002). The bounding contour was computed using an edge detector on the image silhouette of the stimuli. It was then re-plotted using lines of 0.66 degrees width (see Figure 21). This value was used as it produced models of a similar size as the binarised eye movement data.

Model 2: Internal Convex Surface Discontinuity

Model 2 generated predicted fixation regions based on the locations of local features defined by *convex* surface curvature maxima. These were generated by applying a curvature estimation algorithm derived from Taubin (1995) to the object mesh models using the Peyre Matlab toolbox. From this we extracted edges along convex curvature maxima (see Figure 21). The convex features were re-plotted using lines of 0.66 degrees width. Edges on the exterior bounding contour were deleted. Due to the nature of our stimuli, convexities can be present both inside and on the bounding contour of an object but concavities are more likely to occur on the internal contour

(see Figure 21). By keeping internal features only, we are able to compute a bias free measure of the preference for convex or concave image features.

Model 3: Internal Concave Surface Discontinuity

Model 3 generated predicted fixation regions based on the locations of local features defined by *concave* surface curvature minima (see Figure 21). The same curvature estimation method was used as for Model 2, except that here we extracted edges along concave curvature minima. As with Model 2 edges falling on external bounding contour were removed.

Visual Saliency baseline

This model proposed that a separate map is computed for each visual feature (intensity, colour, and orientation) based on changes in these regions, thus increasing the saliency value for that particular region of the image. These separate maps are used to create a single saliency map where winner-take-all processes and inhibition of return are combined to produce a scan sequence of predicted fixations in order of decreasing saliency. However, since, the question of interest was whether specific models of shape analysis could account for fixation patterns beyond that explicable by visual saliency, I have used this model as a baseline contrast.

8.1.3. Analyses of eye movement data

In this study we extracted mean dwell time, fixation frequency (See Table 20) and saccade amplitude from the eye tracker and examined low level parameters of the eye movement data across the two tasks (memory vs. motor -imagery). Initial inspection of the data showed no significant differences between both participant groups and the data was collapsed across groups.

Table 20 The mean dwell time (DT) and fixation frequency (FF) for the three experimental phases (reaching task, memory task and test) collapsed across groups. Standard deviation (SD) of the mean is shown in parentheses.

	Dwell time	**Fixation frequency**
	Mean DT (SD)	Mean FF (SD)
Reaching task	366.00 (105.00)	6.66 (1.37)
Memory task	289.00 (57.00)	13.10 (3.25)
Recognition Test	336.00 (90.00)	5.59 (1.42)

A one way ANOVA on dwell times, F (2, 87) = 5.89, p = .004 revealed a significant main effect of task. Tukey Post-hoc tests showed a significant difference between the reaching task and the memorizing phase of the recognition task, $p < .001$ but not with the test phase, $p > 0.5$.

A one way ANOVA on fixation frequency revealed a significant main effect of task, F (2, 87) = 102.42, $p <$. 0001. Tukey Post-hoc analyses showed a significant difference in the number of fixations made during the reaching task and the memorizing phase of the recognition, $p < .001$ but no significant difference with the test phase, $p > 0.5$.

Further analyses of the saccade amplitude showed no significant main effect of task, but a Tukey Post-hoc test found a significant amplitude effect between the reaching task (M = 4.47, SD=1.08) and the memorizing phase of the recognition task (M = 4.98, SD = 0.78), p = .048.

In addition to the previous analysis methods, here I used a novel technique which enabled me to map fixations from a two dimensional rendered stimulus into the 3D mesh of the object itself (Figure 57). With this method, I computed the object parts on which fixations were made, and mapped the locations of the mouse clicks made by

the participants during the reaching task, which enabled us to relate the mouse clicks to a given object part. Interestingly, in 92% of the trials, participants clicked twice on a single part. From these 92% of trials, we identified an obvious preference for participants to fixate on the part of the object they planned to grasp, where 56% of the fixations were made within the part clicked (*M* = 56% *SD* = 17.8). During the memorising phase (*M* = 23.26, *SD* = 14.40) and test phase (*M* = 23.83%, *SD* = 11.33) of the recognition task, fixations to the clicked part decreased to a chance level (one in four – 25%).

I have also performed time bin analysis, where I re-computed the fixations mapped to object parts, but this time by using the first, second, third, fourth and so forth fixations (see Figure 58). The results showed that the first fixation was not predictive of the task. However, from the second fixation onwards, the strategies clearly changed and participants performing the motor imaginary task were much more likely to fixate on the single part with which they planned to interact. In contrast, during the recognition task the participants were equally likely to fixate any of the four parts.

In addition I have also examined the data-model correspondence for the Motor-imagery task (see Figure 58).

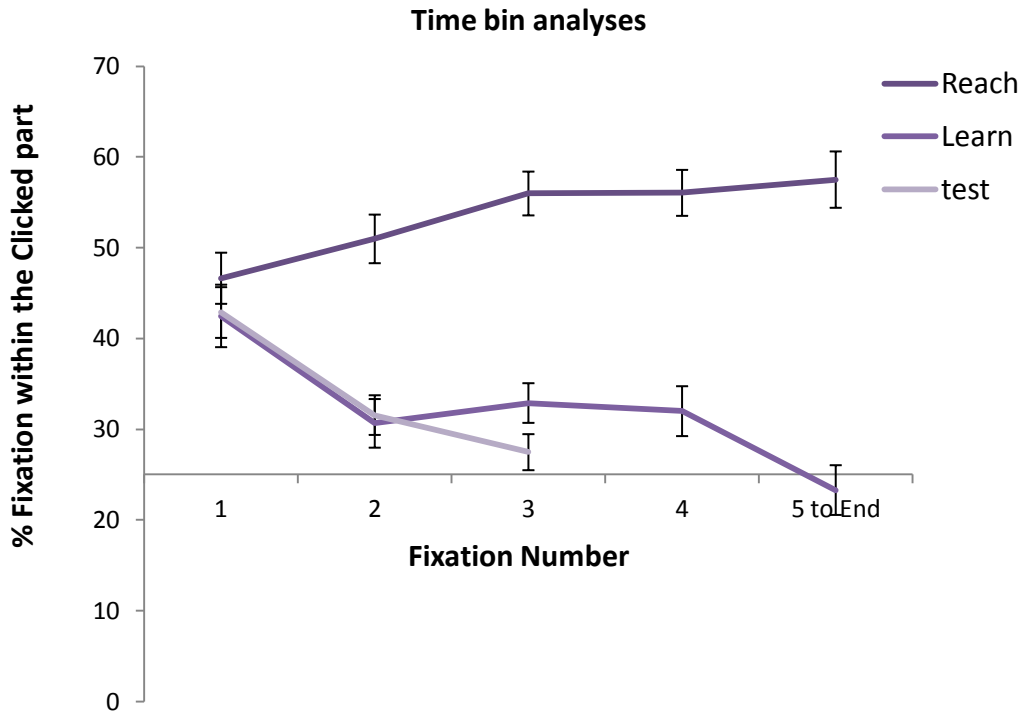


Figure 58 Time bin analysis of the % of fixations to clicked part. Bars show standard error of the mean (%).

Data-model correspondence for the motor-imagery task

A one-way ANOVA on mean MMC values across models (Convex, Concave, and External features) was significant, F (2, 215) = 21.10, p < .0001. Subsequent post-hoc analyses using the Bonferroni test showed that the pairwise contrasts between models were significantly different for External features vs. Concave, p < .0001, and External features vs. Convex, p < .0001 (see Figure 59). There were no other significant differences. These analyses show fixation data-model correspondence preference for both Concave and Convex models of shape analysis, but not for the External features model.

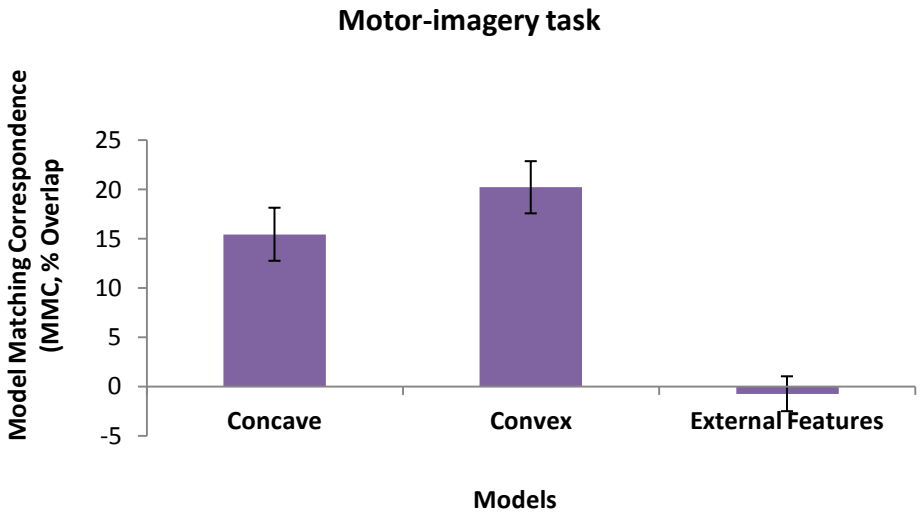


Figure 59 Mean MMC $_{Mx}$ – MMC $_{vs}$ measure of data-model correspondences between models (relative to visual saliency) for the reaching task phase. Bars show standard error of the mean (% overlap).

Summary of the Results

- Eye movement analyses revealed significantly more fixations, shorter dwell times and higher mean saccade amplitudes during the memory task than during the reaching task.
- The first fixation was not predictive of the task in hand. However, subsequent fixations showed that the participants in the motor imagery task preferentially fixated on a single part that they were planning to interact with, whereas during the recognition and memory tasks the participants were equally likely to fixate on any part of the object.
- Data-model correspondences for the motor imagery task were higher for Concave and Convex models, relative to the External features model.

8.1.4. Conclusions

In this study we recorded eye-movement patters during object recognition and a motor-imagery task. The results showed differences in eye movement patterns between the object recognition and motor-imagery task; for example, more saccades, with shorter dwell times and larger amplitudes, in the learning phase of the object

recognition task. During the object recognition task, the fixations were dispersed across all parts of the object, whereas during the motor-imagery task, the eye movements were predominantly localized to the object part where contact between the finger and the object was expected. This pattern of results was evident from the second fixation onwards, thus providing evidence that first fixation does not seem to be indicative of the task ahead, and can possibly only accounted for the object's centre of gravity (Vishwanath & Kowler, 2003). These findings show how eye movement patterns can be used to elucidate the perceptual analysis underlying our perception of shape appearance, and how this analysis differs between tasks. Additionally, the results show the importance of task differences in eye movement studies. Our current results suggest that even when planning an imagery movement to a 2D image observers tend to focus on specific local parts that contain potential grasp locations. In contrast, during the encoding of object shape for recognition, fixation patterns are more spread, indicative of a more global analysis of object configuration.

The analysis of data-model correspondence for the motor-imagery task showed that the eye movement data were best accounted for by models of shape analysis based on local regions of curvature extrema – there was no evidence that fixation distributions were determined by the external features model, despite the potential grip locations of the exterior. Interestingly, these results differ from the findings in Experiment 1 where we used the same algorithmically derived models data and the same stimuli. More specifically, the observers in the motor- imagery task showed no preference for fixating concave over convex regions, whereas the participants in the memory task (Experiment1) predominantly fixated concave regions. This dissimilar pattern of results provides information of how task demand influences local shape analyses during object representations.

Chapter 9

9.1. General discussion

The studies in this thesis examined eye fixation patterns during different visual perceptual tasks in order to investigate what kind of information is mediating the perception of 3D object shape. The main empirical finding revealed that the distribution of fixation patterns (1) varies between object perception and object recognition tasks (e.g. study, test), and (2) are generally consistent across object categorisation tasks. More specifically, in recognition tasks I found a reliable preference for fixation at internal regions of surface concavity. This finding is in line with previous studies demonstrating the importance of curvature singularities in the visual perception of shape (e.g., Attneave, 1954; Barenholtz et al., 2003; De Winter & Wagemans, 2006; Feldman & Singh, 2005; Hoffman & Richards, 1984; Hoffman & Singh, 1997), and also provides evidence for a direct link between the encoding of information about surface concavity and object recognition. One interpretation of this preference is consistent with the influential hypothesis suggesting that concave regions play an important role as segmentation points, allowing for the computation of parts-based structural descriptions (e.g., Hoffman & Richards, 1984; Marr & Nishihara, 1978). In this context, we should focus on an interesting aspect of the data which shows (1) fixation preference for concave surface minima along with (2) viewpoint-dependent performance in the recognition task. While the first finding is consistent with the claim that negative curvature minima play a functional role in part segmentation during the derivation of a structural description representation (e.g., Biederman, 1987; Hoffman & Richards, 1984; Marr & Nishihara, 1978), the second finding is in line with image-based view interpolation models (e.g., Bülthoff &

Edelman, 1992; Edelman & Weinshall, 1991; Reisenhuber & Poggio, 1999; Tarr & Bülthoff, 1998).

Fixation patterns and local object structure in shape representation

One particularly interesting finding here is the link shown between fixation patterns and curvature polarity in 3D objects. While the spatial distributions of fixation regions were constrained by curvature extrema, the data also showed a fixation preference for concave curvature minima – a finding that was robust across study phases, viewpoint change, targets and non-targets and tasks. It could be argued that observers specifically fixate those regions because they are the optimal locations for extracting global shape information, rather than because of their status as regions containing negative curvature minima. In fact, this possibility cannot be absolutely discounted. Nevertheless, this seems unlikely unless there happens to be a high correlation between those optimum locations and the presence of negative curvature minima. It is more likely, that the optimum location for extracting global shape attributes would be close to the centre of mass – which is not the case here (e.g., initial COG fixations were removed from the data prior to analyses). This is consistent with recent attempts to derive a formal derivation of information content along contour in which it has been demonstrated that segments of negative curvature minima are more informative about shape than regions of convex curvature maxima (Feldman & Singh, 2005). Although, it remains unclear whether this derivation can be equally applied to 3D curvature, our data provides support for the hypothesis that negative minima of curvature are more perceptually relevant for analyses of object shape information in the 3D domain.

An additional relevant finding came from the analysis of fixation duration. The mean durations of fixations falling within thresholded fixation regions were

significantly longer than those for non-thresholded regions within the bounding contour of the stimuli. This could be a sign of the differential processing of information content: where longer fixation times are taken as evidence for processing complexity. It is interesting to note that fixation durations were significantly longer in the learning and passive viewing phases of Experiment 1 than in the test phases. One interpretation of this is that fixation duration is at least partly determined by processing associated with the formulation of new long-term memory representations of object shapes, and the computation of local image features, during the initial viewing phase together with a reduction in the time taken to match local shape information between perceptual input and stored representations during the recognition phase. Shape perception during the test phase, and the speed of processing of local regions of curvature extrema, may be facilitated by 'top-down' activation of object-specific knowledge. This is consistent with results from other eye movement studies, for example, in the scene perception and visual search literatures, which show an influence of both stimulus-driven and top-down constraints on task performance (e.g., Foulsham & Underwood, 2007; Malcolm & Henderson, 2009).

Object recognition, view interpolation and concave surface minima

The findings are also consistent with another influential approach based on image based view interpolation studies. On some accounts shape classification is assumed to be based on class-specific appearance or image-based feature hierarchies computed across multiple spatial scales (e.g., Ullman, 2006; Ullman & Bart, 2004; Ullman, Vidal-Naquet & Sali, 2002). The current findings could be interpreted within the context of these models by assuming, for example, that curvature extrema, and in particular, negative curvature minima, form a class of local image features that might be used to constrain shape classification. Other image-based models have

hypothesized the use of 2D views, or aspects, that conjointly encode information about shape and the spatial locations of image features (e.g., Edelman & Weinshall, 1991; Riesenhuber & Poggio, 2006; Ullman & Basri, 1991). In contrast, structural description theories propose that shape perception depends on the decomposition of object shape into generic (rather than appearance-based) primitives (e.g., generalized cylinders, geons or surfaces) and that recognition is mediated by representations that independently encode information about these primitives and their spatial configuration (Biederman, 1987; Hummel & Stankiewitz, 1996; Leek et al., 2009; Leek et al., 2005; Marr & Nishihara, 1978). In these accounts, it has been suggested that negative curvature minima play a key role in the segmentation of complex objects into constituent generic parts (e.g., Cohen & Singh, 2007; De Winter & Wagemans, 2006; Hoffman & Richards, 1984; Hoffman & Singh, 1997). In this context, one interesting aspect of the data stems from the concurrent observations of a fixation preference for local shape regions containing concave surface minima, and viewpoint-dependent performance in the recognition task.

The present findings raise the possibility of an alternative hypothesis about the functional significance of negative curvature minima. In the first place, this fixation sensitivity to curvature polarity suggests that shape perception processes make use of local depth information at least to the level of the $2^1/_2$D sketch (Marr & Nishihara, 1978), and that they do not rely solely on 2D image features computed from the retinal input. Some supporting evidence comes from the recent demonstration by Wexler and Ouarti (2008) showing that saccadic eye movements during the spontaneous exploration of visual images follow surface depth gradients. These findings present a challenge to models of recognition that are based solely on 2D shape representations and view interpolation (e.g., Bulthoff & Edelman, 1992). At the same time, our

behavioural data support the use of viewpoint-dependent object representations. These two empirical observations raise the possibility that view interpolation is constrained by depth specified local features which include surface curvature extrema – and, in particular, regions of concave surface discontinuity. This shows two viewpoint-dependent stored representations of the same object. Recognition of an object from a novel viewpoint is accomplished by an interpolation process that indexes, or matches, local shape information – here, critically, at regions of concave surface discontinuity. On this hypothesis concavities do not derive their functional significance from being cues to volumetric part segmentation, but rather as salient local image features used to constrain interpolation across views.

This proposal is consistent, also, with our finding that the fixation preference for concave regions generalizes from trained to novel viewpoints – and would be expected if these features are used to interpolate stored viewpoint-dependent object representations. I do not claim that surface curvature extrema are the only local features or keypoints that are used in view interpolation. One might also speculate about the use of a large range of other potentially informative local image features including, for example, non-accidental properties (e.g., Biederman, 1987), fragments (Ullman, 2006) and other interest point operators (e.g., Mikolajczyk & Schmid, 2005). Finally, it should be noted that this hypothesis does not exclude the possibility that other forms of representation may also mediate recognition, as suggested by recent hybrid models (Foster & Gilson, 2002; Hummel & Stankiewitz, 1996), or that local regions of curvature minima may also be used in other ways by the visual system, notably, as image segmentation points (e.g., Cohen & Singh, 2007; De Winter & Wagemans, 2006; Hoffman & Richards, 1984; Hoffman & Singh, 1997).

A further aspect of the results that is theoretically interesting is the consistent pattern of data-model correspondences across the active learning and passive viewing tasks in Experiment 1. This is somewhat surprising given that one might expect task requirements to affect the perceptual analysis of shape. Here, despite the fact that only one group of observers were explicitly being told to memorise shape for a subsequent recognition task, the perceptual analysis strategies of the two groups, as evidenced by the patterns of data-model correspondences, were similar. One implication of this finding is that local shape analysis strategies during perception are perhaps 'hard-wired' in the sense of being invariant to task requirements –or at least across the range of tasks tested here. This hypothesis is intuitively appealing in that during everyday recognition observers cannot entirely predict when unfamiliar objects might become relevant to their immediate or future goals and intentions. However, it remains to be determined whether the observed patterns of shape-analyses found in Experiment 1 will generalise across other tasks and stimulus sets, including, for example, those related to the computation of shape representations for object categorisation.

The data in Experiment 2 revealed a statistically reliable fixation bias for concave surface discontinuities in the active learning phase and the test phase of the study. More specifically, the observers showed no preference for fixating concave over convex regions during passive viewing task, whereas in the active learning task the observers predominantly fixated concave regions. Although this pattern of results seems to be somehow intuitive in respect to the imposed task differences, it provides a dissimilar pattern of result to Experiment 1 where concave regions were preferentially fixated during both pre-tasks. There are a number of possible interpretations of these results. For example, local convex and concave curvature may both be processed to some extent during passive viewing (e.g. no task in hand), whereas concave curvature

regions are processed more during the active learning phase when an object shape representation is computed and stored in memory. Moreover, the models used in Experiment 2 were not algorithmically created but collected from a trained observer data incorporating the same amount of error measures as in the recognition task.

From more theoretical point of view, concave curvature discontinuities has been assigned a privileged role in object representation (e.g. Attneave, 1954; Barenholtz, Cohen, Feldman & Singh, 2003; Bertamini, 2008; Biederman, 1987; Cate & Behrmann, 2010; Cohen, Barenholtz, Singh & Feldman, 2005; Cohen & Singh, 2007; De Winter & Wagemans, 2006; Feldman & Singh, 2005; Hoffman & Richards, 1984; Hoffman & Singh, 1997; Lim & Leek, 2012) but there are number of studies proposing that convex curvature discontinues also play an important role in object recognition. For example, resent psychophysical study demonstrated that high level visual cortex areas (LOC) are more sensitive to changes in convex than in concave shapes (Houshofer, Baker, Livingstone & Kanwisher, 2008). Moreover, previous research found that recognition performance (e.g. Driver & Baylis 1995), perceptual judgment of location (e.g. Bertamini, 2001), and shape similarity (Subirana-Vilanova & Richards 1996) are more accurate for convex, rather than concave image features. Nevertheless, investigating the exact magnitude and reasons for changes in contributions between concave and convex curvature during changes in task requirements is out of the scope of this thesis, although it would be addressed in more detail in further studies.

The main empirical findings in Experiment 2 were: (1) Strong viewpoint-dependent pattern of identification latencies in the recognition memory task consistent with other previously reported studies supporting the use of viewpoint-dependent object representations (e.g., Bülthoff & Edelman, 1992; Edelman & Weinshall, 1991;

Tarr & Bülthoff, 1998). (2) Low level visual saliency as implemented by the Itti, Koch and Neibur (1998) model did not perform better than the random model of fixation region distribution in both pre-test and test phases of the study. (3) The fixation distributions were best modelled in terms of local shape analyses at regions of curvature extrema corresponding to concave or convex surface discontinuities. (4) While in the passive view phase there was no significant difference in the spatial distributions of data-model correspondences between the convex and concave surface discontinuity models, observers in the active learning phase and recognition memory test phase showed a fixation bias for regions of concave surface discontinuity. These findings support a large body of work in the psychophysics literature concerning the importance of surface curvature extrema in visual object recognition (e.g., Cohen & Singh, 2007; Cohen et al., 2005; De Winter & Wagemans, 2006; Feldman & Singh, 2005; Hoffman & Richards, 1984). Of particular interest here is that, unlike many previous studies that have reported perceptual biases for convex and/or concave contour curvature in 2D outline forms, here I report a processing bias revealed through fixation patterns determined by surface curvature extrema in 3D forms. The data did reveal a statistically reliable fixation bias for concave surface discontinuities in the active learning phase and the test phase of the study which supports the hypothesis that local regions of surface concavity play an important role in the indexing, encoding and /or matching of perceptual input to stored object shape representations. Furthermore, the analyses showed that similar local image regions were fixated during the active learning and test phases, and that observers tend to fixate regions of concave surface discontinuities across changes in stimulus viewpoint which both support the hypothesis that these local image regions are inherently linked to object recognition.

In Experiment 3 I have examined eye movement patterns during object shape categorisation tasks in experts and novices. The analyses of RT and accuracy data showed that both groups improved their performance as a function of training, although only the subordinate training group showed specific training effects.

The main finding revealed a consistent pattern of data-model correspondences across tasks. During both basic and subordinate matching tasks in pre and post-test phases we found evidence that fixation patterns are predominantly driven by shape information defined by internal regions of concave surface discontinuity regardless of expertise of level of categorisation (basic or subordinate). Thus, the central issues raised here concerns our observation of similar fixation distributions, and similar perceptual strategies for the acquisition of shape information, and later categorisation across pre and post tasks. This is perhaps surprising given that one might expect task requirements to affect the perceptual analysis of shape. Here, despite the fact that the participants were trained to classify objects into two different categories (Basic vs. Subordinate), along with a control group that received no training, the perceptual analysis strategies of the three groups, as evidenced by the patterns of data-model correspondences, were similar.

One implication of this finding is that local shape analysis strategies during perception are 'hard-wired' in the sense of being invariant to task requirements - at least across the range of tasks tested here. The suggestion of 'hard-wired' mechanism is consistent with recent (Amir, Biederman & Hayworth, 2011) findings that adults and infants as young as 5 month-old, looked first, and adults looked longer at simple volumetric shapes containing curved contours, as opposed to straight contours, thus shapes containg high curve value produced larger BOLD activity in the adult's shape selective cortex (Lateral Occipital Cortex). This finding implies that perceptual

mechanisms directing attention to informative object segments exist from early infancy and are not affected by language, geometry training or cultural values. This hypothesis is intuitively appealing in that during everyday recognition observers cannot entirely predict when unfamiliar objects might become relevant to their immediate or future goals and intentions. The current pattern of results raises the questions to whether shape recognition is invariant to classification levels, or whether the same kind of perceptual shape analyses processes underlie the classification of basic and subordinate levels.

As mentioned earlier, these results are consistent with models of object recognition which posit a special functional status to concave image regions in object representation (e.g., Biederman, 1987; Hummel & Stankiewitz, 1996; Leek, Reppa, Rodriguez & Arguin, 2009; Marr & Nishihara, 1978), and they present a challenge to models which do not. Among the latter models are some recent image-based models of recognition (e.g., Edelman & Weinshall, 1991; Riesenhuber & Poggio, 2006; Ullman & Basri, 1991), including HMAX (Serre, Oliva & Poggio, 2007; Serre, Wolf, Bileschi, Riesenhuber & Poggio, 2007). In order to account for these data, these models would require modification to include a level of feature representation which makes explicit the sign of curvature – for example, in the C2 layer of HMAX.

This raises the further issue of why regions of concave curvature should carry such functional significance. Perhaps the most influential hypothesis, following the seminal work of Hoffman and Richards (1984), is that concavities play a key role as local part boundaries supporting volumetric image segmentation. But this does not exclude the possibility that concave regions play other roles in shape recognition and image classification. One important implication of the current results (and those reported by Leek et al., 2012) is that object recognition makes use of local depth

information at least to the level of the $2^1/_2$D sketch (Marr & Nishihara, 1978), and does not rely solely on 2D image features computed from the retinal input. In this respect also, these findings present a challenge to object recognition models that are based solely on 2D image-based representations. However, this does not mean that evidence supporting the functional significance of concave regions during perception is theoretically incompatible with image-based models – such as HMAX. Within the context of viewpoint-dependent, image-based models, concavities may derive functional significance, not as cues to volumetric part segmentation, but rather as salient, local image features, keypoints or interest point operators that constrain generalization across views. In this sense, concave curvature singularities may support operations related to part segmentation on structural-description representations, as well as local keypoint matching, and view generalization, on image-based representations. This proposal is consistent with the use of both kinds of representational codes in human vision – as suggested by some recent hybrid approaches (e.g., (Foster & Gilson, 2002; Hummel & Stankiewitz, 1996).

More broadly, in the context of this hypothesis, it is also tempting to speculate about the interactions between these two representational codes and their roles at different levels of shape classification during object recognition. It has been assumed that entry-level object recognition occurs at an analytic, parts-based, basic-level of object classification potentially supported by a structural description representation in which parts, and their spatial configuration, are independently coded (e.g., Biederman, 1987). Elsewhere, recent work by Gauthier and colleagues (Gauthier & Tarr, 1997; Wong, Palmeri, Rogers, Gore, & Gauthier, 2009) has provided evidence that perceptual expertise in subordinate-level classification is supported by holistic representations which make explicit, amongst other properties, precise metric variation

in image features that permit the individuation of within category exemplars at this level of shape description (e.g., a Ford Kuga from a BMW X3). Thus, while the present results show a preference for fixation at concave regions in both subordinate and basic-level classification, it is possible that the same concave regions support different types of image analysis dependent on the classification level: that is, they serve as image segmentation points during basic-level classification, and as local keypoints to assist functions such as in view generalization during subordinate-level classification.

In Experiment 4 the extended training regime from three hours (in Experiment 3), to four hours revealed less consistent pattern of data-model correspondence across tasks. The RTs showed that each training group improved their post-test mean performance as a function of training, but only the basic group showed a significant specific effect in the basic task. Although the results showed that in pre-tests (basic and subordinate) fixation patterns are driven by shape information defined by internal regions of concave surface discontinuity, regardless of expertise or level of categorisation (basic or subordinate), this preference was not evident at post-tests. In both post sequential matching tasks, fixation patterns were driven by both types of local image curvature information (concave and convex). As mentioned previously, one likely explanation of this finding is the suggested two-stage model of perceptual category learning (Ashby and Spiering, 2004; Nosofsky, 1986; Riesenhuber and Poggio, 2000; Sigala, 2004; Thomas et al., 2001) where high level shape representation can be activated for different tasks using the same stimuli in order to improve discrimination of these stimuli relative to untrained participants (Jiang et. al., 2007). A possible reason for this different performance compared to the preceding Experiment 3 is that in the current Experiment 4 I have presented each stimulus image

only once in pre and post tests. This potentially allowed me to investigate the processes behind the initial perceptual stimulus encoding, and avoid potentially confounding variables such priming.

These findings are to some extent consistent with Experiment 3 results and add to the importance of curvature singularities in the visual perception of shape (e.g., Attneave, 1954; Barenholtz Cohen & Singh, 2003; De Winter & Wagemans, 2006; Feldman & Singh, 2005; Hoffman & Richards, 1984; Hoffman & Singh, 1997). The previously found preference for fixation at regions of concave surface discontinuity regardless of task (basic vs. subordinate) or phase (pre vs. post) was not entirely replicated in this study, and this provides further information of the complexity of object shape representation.

The analyses of the saccade amplitude and dwell time showed no significant differences post-training. However, saccade amplitude was shorter and dwell time longer in image 2, showing the expected pattern associated with recognition tasks. Moreover, the lack of significant change between pre and post tests was consistent in both types of analyses (e.g. saccade amplitude/dwell times and data-model correspondence), showing the systematic nature of the findings.

In my last study I recorded eye-movement patterns during object recognition and a motor-imagery task. The results showed differences in eye movement patterns between the object recognition and motor-imagery task; for example, more saccades, with shorter dwell times and larger amplitudes, for the learning phase of the object recognition task. One important finding is the fact that we found there characteristics despite not using real life objects but presenting 2D images. During object recognition task, the fixations were dispersed across all parts of the object, whereas during motor-imagery task, the eye movements were predominantly localized to the object part

where contact between the finger and the object was expected. This pattern of results was evident from the second fixation onwards, thus providing evidence that first fixation does not seem to be indicative of the task ahead, and can possibly only account for the object's center of gravity (Vishwanath & Kowler, 2003).These findings show how eye movement patterns can be used to elucidate the perceptual analysis underlying our perception of shape appearance, and how this analysis differs between tasks. Additionally, the results show the importance of task differences in eye movement studies.

The current results suggest that when planning prehensile grasping actions observers tend to focus on specific local parts that contain potential grasp locations. In contrast, during the encoding of object shape for recognition, fixation patterns are more spread, indicative of a more global analysis of object configuration.

There are a few additional aspects of the results that are worth discussing. One of these is the poor performance of the visual saliency algorithm of Itti, Koch and Niebur (1998) in predicting the spatial distributions of fixations. The visual saliency model performed no better than chance in accounting for fixation patterns in either of the tasks. These data contrast with results from some other studies that have provided support for the role of low-level image statistics in stimulus-driven eye movements and attentional capture in scene perception (Baddeley & Tatler, 2006; Foulsham & Underwood, 2007; Malcolm & Henderson, 2009; Rajashekar, van der Linde, Bovik & Cormack, 2007). For example, Baddeley and Tatler (2006) have shown that fixation patterns in natural scene viewing can be predicted from high spatial frequency edge information but not contrast. Although the analyses did not provide empirical support for visual saliency (as implemented by Walther and Koch, 2006) as a predictor of fixation patterns in object recognition, I cannot rule out that some other version of the

hypothesis, or some other combination of model parameters would fare better. However, in the context of shape perception, the conceptual notion of visual saliency may need to be broadened in light of the current findings, to include geometric features of objects in the visual scene, such as the magnitude and sign of surface curvature, in addition to purely low-level image properties.

The main empirical findings in this thesis provide evidence that: 1) eye movements can elucidate properties of internal mental representations of shape; 2) consistent fixation pattern to concave areas that generalises across tasks; and 3) different fixation patterns during recognition and motor imagery task.

9.2. Conclusions

The studies in this thesis provide some of the first evidence from measures of fixation patterns regarding the acquisition of higher level shape information during the perception and recognition of 3D objects. These studies demonstrate the considerable potential for quantitative analyses of fixation patterns to elucidate local shape feature processing during object perception and recognition. It is important to note that the results are preliminary in a number of ways as the experiments in this thesis were not designed to investigate the role of eye movements during object processing in daily environments (e.g., Land et al., 1999; Tatler, Gilchrist & Rusted, 2003), hence I do not claim that these fixation patterns necessarily characterize the role of eye movements in day to day object processing. Rather, my goal was to examine how fixation patterns may be used to inform hypotheses about shape analyses in vision under appropriately controlled experimental conditions. There also remains considerable on-going debate about the relative contributions of foveal and para-foveal processing in perception (e.g., Henderson et al., 1997), and about the extent, and conditions under which, fixation patterns are influenced by further variations in

task, display and stimulus properties. These issues, in relation to fixation patterns during object recognition, require systematic examination in future studies.

References

Amir, O., Biederman, I., & Hayworth, K. J. (2011). The neural basis for shape preferences. *Vision Research, 51*(20), 2198-2206.

Andrews, T. J., Schluppeck, D., Homfray, D., Matthews, P., & Blakemore, C. (2002). Activity in the fusiform gyrus predicts conscious perception of Rubin's Vase: Face illusion. *NeuroImage, 17*, 890-901.

Arguin, M., & Leek, E. C. (2003). Orientation invariance in visual object priming depends on prime-target asynchrony. *Perception and psychophysics, 65*(3), 469-77.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61,* 183–193.

Avidan, G., Harel, M., Hendler, T., Ben-Bashat, D., Zohary, E., & Malach, R. (2002). Contrast sensitivity in human visual areas and its relationship to object recognition. *Journal of Neurophysiology, 87,* 3102–3116.

Ashby, F. G., & Spiering, B. J. (2004). The neurobiology of category learning. *Behavioral Cognitive Neuroscience Review, 3,* 101–113.

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research, 46,* 2824–2833.

Barenholtz, E., Cohen, E. H., Feldman, J., & Singh, M. (2003). Detection of change in shape: An advantage for concavities. *Cognition, 89,* 1–9.

Bear, M. F., Connors, B. W., & Paradiso, M. (2007). *Neuroscience: Exploring the brain* (Third ed.). PA: Lippincott Williams & Wilkins.

Bertamini, M. (2001).The importance of being convex: an advantage for convexity when judging position. *Perception, 30,* 1295-1310.

Bertamini, M. (2008). Detection of convexity and concavity in context. Journal of Experimental Psychology: *Human Perception and Performance, 34,* 775–789.

Bertamini, M., & Farrant, T. (2005). Detection of change in shape and its relation to part structure. *Acta psychologica, 120*(1), 35-54.

Betz, T., Kietzmann T. C., Wilming N. & König, P. (2010). Investigating task-dependent top-down effects on overt visual attention. *Journal of Vision 10(15*), 1-14.

Bhatt, R. S., Hayden, A., Reed, A., Bertin, E., & Joseph, J. E. (2006). Infants' perception of information along object boundaries: Concavities versus convexities. *Journal of Experimental Child Psychology, 94*, 91–113.

Biederman, I. (1985). Human image understanding: Recent research and a theory. *Computer Vision, Graphics, and Image Processing, 32,* 29–73.

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

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

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

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19*(6), 1162-1182.

Biederman, I., Ju, G., & Clapper, J. (1985). The perception of partial objects. Unpublished manuscript, State University of New York at Buffalo.

Binford, T. O. (1971, December). Visual perception by computer. Paper presented at the IEEE Systems Science and Cybernetics Conference, Miami, FL.

Brouwer, A.-marie, Franz, V. H., & Gegenfurtner, K. R. (2009). Differences in fixations between grasping and viewing objects. *Journal of Vision, 9,* 1-24.

Brunelli, R. & Poggio, T. (1993). Face recognition: Features versus templates. *IEEE PAMI, 15,* 1042–1052.

Bülthoff, H. H. & Edelman, S. (1992). Psychophysical support for a 2-D view interpolation theory of object recognition. *Proceedings of the National Academy of Science, 89,* 60-64.

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

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *Trends in Cognitive Sciences, 10,* 159–166.

Cambell, F. W., & Wutrz, R. H. (1979). Saccadic omission: Why we do not see a gray-out during a saccadic movement. *Vision Research, 18,* 1297-1301.

Cate, D. & Behrmann, M. (2010). Perceiving parts and shapes from concave surfaces. *Attention, Perception and Psychophysics, 72,*153-167.

Cheung, O. S., Richler, J. J., Palmeri, T. J., & Gauthier, I. (2008). Re- visiting the role of spatial frequencies in the holistic processing of faces. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1327–1336.

Cohen, E. H., Barenholtz, E., Singh, M., & Feldman, J. (2005). What change detection tells us about the visual representation of shape. *Journal of Vision, 5(4):3,* 313–321, doi:10.1167/5.4.3.

Cohen, E. H. & Singh, M. (2007). Geometric determinants of shape segmentation:

Tests using segment identification. *Vision Research, 47*, 2825-2840.

Conlan, L. I., Phillips, J. C. & Leek, E. C. (2009): Negative priming of unattended part primes: Implications for models of holistic and analytic processing in object recognition, *The Quarterly Journal of Experimental Psychology, 62*:12, 2289-2297.

Cooper, L. A., Schacter, D. L., Ballesteros, S., &Moore, C. (1992). Priming and recognition of transformed three-dimensional objects: Effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 43–57.

Cristino, F. & Baddeley, R. (2009). The nature of the visual representations involved in eye movements when walking down the street. *Visual Cognition, 17(6),* 880-903.

de Grave, D. D. J., Hesse, C., Brouwer, A. M., & Franz, V. H. (2008). Fixation locations when grasping partly occluded objects. *Journal of Vision 8*(7):1–11.

Delabarre, E. B. (1898). A method of recording eye-movements. *American Journal of Psychology, 9(4),* 572-574.

Denisova, K., Singh, M., & Kowler, E. (2006). The role of part structure in the perceptual localization of shape, *Perception, 35*, 1073–1087.

Desanghere, L., & Marotta, J. J. (2011). "Graspability" of objects affects gaze pattern during perception and action tasks. *Experimental brain research, 212*(2), 177-87.

De Winter, J. & Wagemans, J. (2006). Segmentation of object outlines into parts: A large-scale integrative study. *Cognition, 99*, 275-325.

Driver, J. & Baylis, G. C. (1995). One-sided edge assignment in vision: 2. Part

decomposition, shape description, and attention to objects. *Current Directions in Psychological Science. 4:*201–206.

Dodge, R., & Cline, T. S. (1901). The angle velocity of eye movements. *Psychological Review, 8(2),* 145-157.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science (New York, N.Y.), 293*(5539), 2470-3.

Du Laurens, A. (1599). A Discourse of the Preservation of the Sight: of Melancholic Diseases, of Rheumes, and of Old Age, translated by R. Surphlet (London: The Sheaskspear Association).

Edelman, S. (1995). Representation of similarity in three-dimensional object discrimination. *Neural Computation, 7*, 408 – 423.

Edelman, S., & Weinshall, D. (1991). Biological Cybernetics A self-organizing multiple-view representation of 3D objects. *Structure, 219*, 209-219.

Epelboim, J., Booth, J. R., & Steinman, R. M. (1994). Reading unspaced text: Implications for theories of reading eye movements. *Vision Research, 34*, 1735–1766.

Epelboim, J., Booth, J. R., & Steinman, R. M. (1996). Much ado about nothing: The place of space in text. *Vision Research, 36*, 465–470.

Feldman, J., & Singh, M. (2005). Information along contours and object boundaries. *Psychological Review, 112*, 243-252.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and brain sciences, 22*(4), 661-721.

Foster, D. H., & Gilson, S. J. (2002). Recognizing novel three-dimensional objects

by summing signals from parts and views. *Proceedings in Biological Science, 269,* 1939–1947.

Fujita, T., Privitera, C. M., & Stark, L.W. (2007). Image-type dependent eigen-regions-of-interest define conspicuity operators for predicting human scan path fixation. *Computers in Biology and Medicine, 37(7):*965-74.

Gauthier, I., Anderson, A.W., Tarr, M. J., Skudlarski, P., & Gore, J.C. (1997). Levels of categorisation in visual object studied with functional MRI. *Current Biology, 7,* 645-651.

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience, 6,* 428–432.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision research, 37*(12), 1673-82.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behaviour. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 431–446.

Gilbert, C., Ito, M., Kapadia, M., & Westheimer, G. (2000). Interactions between attention,context and learning in primary visual cortex. *Vision Research, 40,* 1217-1226.

Gilford, J. P. & Hackman, R.B (1936). A study of the 'visual fixation' method of measuring attention value. *Journal of Applied Psychology, 20,* 44 -59.

Craddock, M., Martinovic, J., & Lawson, R., (2011). An advantage for active versus passive viewing aperture viewing in visual object recognition. *Perception. 40 (10),* 1154–1163.

Grill-Spector K (2003). The neural basis of object perception. *Current Opinion in Neurobiology,13(2),* 159-166.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: as soon as you know it is there, you know what it is. *Psychological science*, *16(2),* 152-60.

Goodale, M. A., & Milner, A.D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience, 15,* 20–25.

Haushofer, J., Livingstone, M. S., & Kanwisher, N. (2008). Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. *PLoS Biology, 6,* e187.

Harris, C. & Stephens, M. (1988). A combined corner and edge detector. *Fourth Alvey Vision Conference,* 147-151.

Haxby, J.V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and over- lapping representations of faces and objects in ventral temporal cortex. *Science 293,* 2425–30.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9(4*), 188–194.

Hayward, W. G. (2003). After the viewpoint debate: where next in object recognition? *Trends in cognitive sciences, 7*(10), 425-7.

Hayward, W. G. (1998). Effects of outline shape in object recognition. *Journal of Experimental Psychology: Human Perception & Performance, 24*, 427-440.

Hayworth, K. J. & Biederman, I. (2006). Neural evidence for intermediate representations in object recognition. *Vision Research, 46,* 4024-4031.

Hayward, W. G., Tarr, M. J., & Corderoy, A. K. (1999). Recognizing silhouettes and shaded images across depth rotation. *Perception, 28,* 1197-1215.

He, P. & Kowler, E. (1989). The role of location probability in the programming of

saccades: Implications for “center-of-gravity” tendencies. *Vision Research 29,* 1165–1181.

Henderson, J. M. (1993). Eye movement control during visual object processing: Effects of initial fixation position and semantic constraint. *Canadian Journal of Experimental Psychology, 47,* 79- 98.

Henderson, J. M. (2003). Human gaze control during real- world scene perception. *Trends in Cognitive Sciences, 7,* 498–504.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), Eye movements: A window on mind and brain (pp. 537–562). Oxford, UK: Elsevier.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50,* 243-271.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception and Psychophysics, 65,* 58-71.

Henderson, J. M., Williams, C. C., Castelhano, M. S., & Falk, R. J. (2003). Eye movements and picture processing during recognition. *Perception & psychophysics*, *65*(5), 725-34.

Hoffman, D. D. (1983). The interpretation of visual illusions. *Scientific American, 249,* 154-162.

Hoffman, J. (1998). Attention and eye movements. In H. Pashler (Ed.), *Attention* (pp. 257-295). East Sussex: Psychology Press Ltd.

Hoffman, D. D., & Richards, W. (1984). Parts of recognition. *Cognition, 18,* 65–96.

Hoffman, D. D. & Singh, M. (1997). Salience of visual parts. *Cognition, 63*, 29-78.

Hubel, D., &Wiesel, T. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology, 148*, 574–591.

Hubel, D. H. & Wiesel, T. N. (1968) Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology, 195,* 215-243.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 9,* 480 -517.

Hummel, J. E. & Stankiewicz, B. J. (1996). An architecture for rapid, hierarchical structural description. In T. Invi & J. McClelland (Eds.), Attention and Performance XVI: Information Integration in Perception and Communication (pp. 93-121). Cambridge, MA: MIT Press.

Hummel, J. E., & Stankiewicz, B. J. (1998). Two roles for attention in shape perception: A structural description model of visual scrutiny. *Visual Cognition, 5*, 49–79.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, *40*(10-12), 1489-506.

Itti, L., & Koch, C. (2001). Feature combination strategies for saliency-based visual attention systems. *Journal of Electronic Imaging, 10(1),* 161–169.

Itti, L., Koch, C. & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254-1259.

Ito, M., & Komatsu, H. (2004). Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *Journal of Neuroscience, 24*, 3313-3324.

James, K. H., Humphrey, G. K., Vilis, T, Baddour, R., Corrie, B. & Goodale, M. A.

(2002). Active and Passive learning of three-dimensional object structure within an immersive virtual reality study. *Behavioral Research Methods, Instruments and Computers, 34*(3), 383-390.

James, T.W., & Kim, S. (2010). Dorsal and ventral cortical pathways for visuo-haptic shape integration revealed using fMRI. In M.J. Naumer & J. Kaiser (Eds.), *Multisensory object perception in the primate brain*. New York: Springer

Javal, E. (1879). Essai sur la Physiologie de la Lecture. *Annales D'Oculistique, 81,* 61-73.

Jeannerod, M. (1988). The neural and behavioural organization of goal-directed movements. Oxford: Oxford University Press.

Johnston, S., & Leek, E. C. (2009). Fixation Region Overlap: A quantitative method for the analysis of fixational eye movement patterns. *Journal of Eye Movement Research, 1(3),* 1-12.

Jolicoeur, P., Gluck, M., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16*, 243–275.

Kanwisher, N. G., Chun, M. M., McDermott, J., & Ledden, P. J. (1996). Functional imagining of human visual recognition. *Brain Research Cognitive Brain Research, 5*, 55–67.

Kanwisher, N. G., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

Kayaert, G., Biederman, I., & Vogels, R. (2003).Shape tuning in macaque inferior temporal cortex. *Journal of Neuroscience 23*, 3016–3027.

Kayaert, G., Biederman, I., Op de Beeck, H. P., & Vogels, R. (2005). Tuning for

shape dimensions in macaque inferior temporal cortex. *European Journal of Neuroscience, 22,* 212-224.

Koch, C. & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry, *Human Neurobiology, 4,*219-227.

Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 20*(9), 3310-8.

Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception, 13*, 321–330.

Krueger, L. E. (1978). A theory of perceptual matching. *Psychological Review, 85,* 278-304.

Land, M.F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities. *Vision Research 41,* 3559-3565.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception,* 28(11), 1311-1328.

Lee, T. S. (2003). Computations in the early visual cortex. *Journal of physiology, Paris, 97*(2-3), 121-39.

Leek, E. C. (1998a). The analysis of orientation-dependent time costs in visual recognition. *Perception, 27,* 803–816.

Leek, E. C. (1998b). Effects of stimulus orientation on the identification of common polyoriented objects. *Psychonomic Bulletin & Review, 5,* 650– 658.

Leek, E. C., Reppa, I. & Arguin, M. (2005). The structure of 3D object shape representations: Evidence from part-whole matching. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 668-684.

Leek, E. C., Reppa, I., Rodriguez, E., & Arguin, M. (2009). Surface but not

volumetric part structure mediates three-directional shape representation: Evidence from part-whole priming, *Quarterly Journal of Experimental Psychology, 62***,** (4), 814-830.

Leek, E. C., Cristino, F., Conlan, L. I., Patterson, C., Rodriguez, E., & Johnston, S. J. (2012). Eye movement patterns during the recognition of three-dimensional objects: Preferential fixation of concave surface curvature minima. *Journal of Vision, 12*(1):7, *1–15*, doi:10.1167/12.1.7.

Levine, D. N., Warach, J., & Farah, M. (1985). Two visual systems in mental imagery: Dissociation of "what" and "where" in imagery disorders due to bilateral posterior cerebral lesions. *Neurology, 35*, 1010-1018.

Li, X. J., & Lisberger, S. G. (2011). Learned Timing of Motor Behavior in the Smooth Eye Movement Region of the Frontal Eye Fields. *Neuron, 69*, 159-169.

Lim, I. S., & Leek, E. C. (2011). Curvature and the Visual Perception of Shape: Theory on Information Along Object Boundaries and the Minima Rule Revisited. *Psychological Review.* Advance online publication. doi: 10.1037/a0025962

Liversedge, S. P. & Findlay, J. M. (2000). Saccadic eye movements and cognitive science. *Trends in Cogitive Science, 4*, 6-14.

Lloyd-Jones, T. J., Gehrke, J., & Lauder, J. (2010). Animal recognition and eye movements: The contribution of outline contour and local featural information. *Experimental Psychology, 57*, 117-125.

Lloyd-Jones, T. J., & Luckhurst, L. (2002). Outline shape is a mediator of object recognition that is particularly important for living things. *Memory and Cognition, 30*, 489-498.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Proceedings of the International Conference on Computer Vision, 2,* 1150–1157.

Lowe, D.G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision, 60*, 91–110.

Mack, M. L., & Palmeri, T. J. (2010). Decoupling object detection and categorization. *Journal of experimental psychology. Human perception and performance*, *36*(5), 1067-79.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision, 11(2),* 157–178.

Manor, B. R. & Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuo- cognitive tasks. *Journal of Neuroscience Methods, 128,*85-93.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Marr, D., & Hildreth, E. (1980).Theory of edge detection. *Proceedings of the Royal Society London B, 207,* 187–217.

Marr, D. & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional structure. *Proceedings of the Royal Society of London B, 200,* 269-294.

Matsukara, M., Brockmole, J. R., & Henderson, J. M. (2009). Overt attentional prioritization of new objects and feature changes during real-world scene viewing. *Visual Cognition, 17(6/7),*835-855.

Melcher, D. & Kowler, E. (1999). Shapes, surfaces and saccades. *Vision Research, 39,* 2929-2946.

Mikolajczyk, K., & Schmid, C. (2005).“A performance evaluation of local descriptors,” *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(10)*:1615–1630.

Mohan, A. (1999). Object detection in images by components. AI Memo 1664 CBCL and AI Lab, MIT, Cambridge, Massachusetts.

Monty, R. A. & Senders, J. W. (Eds.). (1976). Eye Movements and Psychological Processes, Hillsdale, N.J.: Lawrence Erlbaum.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading – evidence for parallel programming of saccades. *Journal of experimental psychology-human perception and performance, 10*(5), 667-682.

Murray, J. E. (1998). Is entry-level recognition viewpoint invariant or viewpoint dependent? *Psychonomic Bulletin & Review, 5*(2), 300-304. doi:10.3758/BF03212955.

Najemnik, J. & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434,* 387-391.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology General, 115,* 39–61.

O’Regan, J. K., & Levy–Schoen, A. (1987). Eye movement strategy and tactics in word recognition and reading. In M. Coldeart (Ed.), *Attention and Performance: Vol.12. The psychology of reading* (pp. 363-383), NJ: Erlbaum.

Osterberg, G. (1935). Topography of the layer of rods and cones in the human retina. Acta Ophthal.(suppl.) *6,* 1-103.

Palermo, R., & Rhodes, G. (2002). The influence of divided attention on holistic face perception. *Cognition, 82,* 225–257.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology, 9,* 441–474.

Parkhusrt, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision, 16,* 125–154.

Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42(1),* 107-123.

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of neurophysiology*, *82*(5), 2490-502.

Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *Journal of neurophysiology*, *86*(5), 2505-19.

Peissing, J. J., & Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology, 58*, 75-96.

Perenin, M.-T., & Vighetto, A. (1988). Optic ataxia: A specific disruption in visuomotor mechanisms. I: Different aspects of the deficit in reaching for objects. *Brain, 111*, 643-674.

Poggio, T. & Edelman, S. (1990). *Nurure (London) 343*, 263- 266.

Porterfield, W. (1737). An essay concerning the motions of our eyes. Part I. *Of their external motions. Edinburgh Medical Essays and Observations 3,* 160–263.

Privitera, C. M., & Stark, L.W.(2000). Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 9,* 970-982.

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review, 88*, 552-572.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372-422.

Rayner, K. (2009). The 35th Frederick Bartlett Lecture Eye movements and attention in reading, scene, perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62(8),* 1457-1506.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading:comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1188-1200.

Renninger, L.W., Coughlan, J., & Verghese, P. (2005). An information maximization model of eye movements. In L.K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems* (Vol 17). Cambridge, MA: MIT Press.

Renninger, L.W., Verghese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of Vision, 7(3*):6, 1-17.

Riesenhuber, M. & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019-1025.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature neuroscience, 3 Suppl*, 1199-204. doi:10.1038/81479

Riesenhuber, M. & Poggio, T. (2003). How Visual Cortex Recognizes Objects: The Tale of the Standard Model. *MIT Press.*

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes- Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

Rosin, P. L. (2000). Shape partitioning by convexity. *IEEE Transactions Systems, Man, and Cybernetics, .3 (2),* 202-210.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision, 7(14*), 1-20.

Schendan, H.E., & Kutas, M., 2003. Time Course of Processes and Representations Supporting Visual Object Identification and Memory. *Journal of Cognitive Neuroscience, 15(1):* 111-135.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America, 104*(15), 6424-9.

Senders, J. W., Fisher, D. F., & Monty, R. A. (Eds.). (1978). Eye Movements and the Higher Psychological Functions, Hillsdale, N.J.: Lawrence Erlbaum.

Sigala, N. (2004). Visual categorization and the inferior temporal cortex. *Behavioural Brain Research, 149,* 1-7.

Singh, M., & Hoffman, D. D. (1999). Completing visual contours: the relationship between reliability and minimizing inflections. *Perception and Psychophysics, 61,* 943–951.

Singh, M., Seyranian, G. D., & Hoffman, D. D. (1999). Parsing sil houettes: The short-cut rule. *Perception & Psychophysics, 61,* 636-660.

Schütz, A. C., Braun, D. I. & Gegenfurtner, K. R. (2011) Eye movements and perception: a selective review. *Journal of Vision, 11*(5), 9.

Stratton, G. M. (1906). Symmetry, linear illusion, and the movements of the eye. *Psychological Review, 13(2),* 82-96.

Subirana-Vilanova, J. B. & Richards, W. (1996). Attentional frames, frame curves and figural boundaries: the inside/outside dilemma. *Vision Research*, *36*: 1493–1501.

Tanaka, K., Saito, H., Fukada, Y. & Moriya, M. (1991). Coding visual images of

objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology, 66,* 170-189.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*(3), 457-482.

Tarr, M. J. & Bulthoff, H. H. (1995). Is human object recognition better described by geon-structural-descriptions or by multiple views? *Journal of Experimental Psychology: Human Perception and Performance, 21*(6), 1494–1505.

Tarr, M. J. & Bülthoff, H. H. (1998) (Eds). Object recognition in man, monkey and machine. *MIT Elsevier.*

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection:Effects of scale and time. *Vision Research, 45(5),* 643-659.

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: spatial statistics at fixation vary with saccade amplitude and task. *Vision Research, 46,* 1857–1862.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research, 2(5),* 1-18.

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition, 17(6/7),* 1029-1054.

Tatler, B. W., Gilchrist, I. D. & Rusted, J. (2003). The time course of abstract visual representation. *Perception, 32,* 579–592.

Taubin, G. (1995). Estimating the tensor of curvature of a surface from a polyhedral approximation. *In Proceedings of IEEE International Conference on Computer Vision. IEEE, Los Alamitos*, CA, 902–907.

Thoma, V., Davidoff, J., & Hummel, J. E. (2007). Priming of plane-rotated objects depends on attention and view familiarity. *Visual Cognition, 15,* 179–210.

Thoma, V., Hummel, J. E., & Davidoff, J. (2004). Evidence for holistic representations of ignored images and analytic representation of attended images. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 257–267.

Thomas, E., Van Hulle, M. M, & Vogels, R. (2001). Encoding of categories by noncategory-specific neurons in the inferior temporal cortex. *Journal of Cognitive Neuroscience,13*, 190–200.

Tversky, B. (1989). Parts, partonomies, and taxonomies. *Developmental Psychology, 25*, 983-995.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113*, 169-193.

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in cognitive sciences*, *11*(2), 58-64.

Ullman, S., & Bart, E. (2004). Recognition invariance obtained by extended and invariant features. *Neural networks : the official journal of the International Neural Network Society*, *17*(5-6), 833-48.

Ullman, S., & Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelli- gence, 13*, 992–1005.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002).Visual features of intermediate complex- ity and their use in classification. *Nature Neuroscience 5*, 682–687.

Underwood, G., Foulsham, T., & Humphrey, K. (2009). Saliency and scan patterns in the inspection of real-world scenes: Eye movements during encoding and recognition. *VisualCognition, 17(6/7)*, 812-834.

Underwood, G., Foulsham, T., vanLoon, E., Humphreys, L., & Bloyce, J. (2006).

Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology, 18,* 321–343.

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition,12*(3), 473-494.

Ungerleider, L.G., & Mishkin, M. (1982).Two cortical visual systems. In: Ingle DJ, Goodale MA, Mansfield R. J. (eds) The analysis of visual behaviour. MIT Press, Cambridge, MA, pp 549–586.

van Donkelaar, P., Siu, K. C., & Walterschied, J. (2004). Saccadic output is influenced by limb kinetics during eye-hand coordination. *Journal of Motor Behaviour, 36(3*):245–252.

van Oeffelen, M. P., & Vos, P. G. (1982). Configurational effects on the enumeration of dots: counting by groups. *Memory & Cognition, 10(4),* 396–404.

Velichkovsky, B. M., Joos, M., Helmert, J. R., & Pannasch, S. (2005). Two visual systems and their eye movements: Evidence from static and dynamic sceneperception. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the XXVII Conference ofthe Cognitive Science Society* (pp. 2283-2288).Mahwah, NJ: Lawrence Erlbaum.

Vergilino-Perez, D., Collins, T., & Dore-Mazars, K. (2004). Decision and metrics of refixations in reading isolated words. *Vision Research, 44*, 2009-2017.

Vishwanath, D.,& Kowler, E. (2003). Localization of shapes: eye movements and perception compared. *Vision Research, 43*, (15), 637-1653.

Vishwanath, D. & Kowler, E. (2004). Saccadic localization in the presence of cues to three-dimensional shape. *Journal of Vision, 4*, 445-458.

Wade, N. J. (2010). Pioneers of eye movement research. *i-perception, 1,* 33 - 68.

Walther, D. & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19,* 1395-1407.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron, 56,* 366–383.

Wexler, M., & Ouarti, N. (2008). Depth affects where we look. *Current Biology, 18,* 1872–1876.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shape less bundles of basic Features. *Vision Research, 37,* 25-43.

Wong, A. C. N., Palmeri, T. J., Rogers, B. P., Gore, J. C., & Gauthier, I. (2009). Beyond shape: How you learn about objects affects how they are represented in visual cortex. *PLoS One, 4,* e8405.

Wong, A.C.-N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for face-like expertise with objects: Becoming a Ziggerin expert -- but which type? *Psychological Science, 20,* 1108-1117.

Xu, Y., & Singh, M. (2002). Early computation of part structure: Evidence from visual search. *Perception & Psychophysics, 64,* 1039-1054.

Yarbus, A. L. (1967). Eye Movements and Vision, New York: Plenum Press.