UNIVERSITY OF ADELAIDE LIBRARY
16-12-02

# THE RECOGNITION OF PLANAR ROTATED AND SCALED FORMS: NORMALIZATION VERSUS INVARIANT FEATURES

**Marcus A. Butavicius, B.A. Hons (Adelaide)**

*Thesis submitted to the University of Adelaide in fulfillment of the conditions for the degree of Doctor of Philosophy.*

**Department of Psychology**
**University of Adelaide**

**July, 2002**

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ABSTRACT

There are two major views on how humans recognize objects despite changes in viewpoint. On the one hand, an internal model of an object may be normalized (e.g., rotated) until there is a template match between views of the stimulus and internal model. On the other hand, matching may be achieved by comparing geometrical features of the stimulus and internal model that are invariant under perspective changes. Although generally seen as competing paradigms, we argue that both processes are employed by the visual system. On the basis of previous literature and two experiments on the recognition of planar rotated and scaled images, we propose that there are several variables that influence which of the two mechanisms is employed. Several existing algorithms for normalization and invariant feature recognition are discussed and a novel model of invariant recognition is also proposed that is inspired by neural mapping in the human visual system. This project takes a multi-disciplinary approach to the problem by looking at not only traditional psychological paradigms but also explicit computational models and neurophysiological data on visual processing.[1]

---

[1] Selected findings of this thesis have been presented at the 36th Annual Australian Psychological Society Conference (Butavicius, 2001) and the 29th Annual Australasian Experimental Psychology Conference (Butavicius, 2002).

# STATEMENT

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

12/11/02

# ACKNOWLEDGMENTS

I would like to thank my supervisor Douglas Vickers for his friendly advice and guidance throughout my candidature, Bob Wilson and Mark Brown for their technical and statistical advice, Nick Burns, Chris Cooper, Michael Lee, Gerard O'Brien, George Galanis and Armando Vozzo for their support, Carmen and Josh for administrative help, all my friends (especially Dames, Tom, Meg and all the Sheridans, Sunshine, Cathy, Theo, Cam, Josh, Kylie, Poppy, Dom, Maria, Lucy, Michael, Helen, Katerina, Dan and Shona) for never losing the faith and last, but certainly not least, all my family especially Mum and Dad for their unlimited support and patience. This thesis is dedicated to you, Mum.

# CHAPTER 1: Introduction

*The visible world only becomes the real world by the operation of thought. (Jean Metzinger)*

## 1.1 The problem of object recognition

In a hostile and competitive world, identifying the objects in our environment is a fundamentally important task. Being able to recognize quickly that a nearby object is a tiger has an obvious survival advantage. Given the evolutionary importance of this task, our visual systems have evolved such that object recognition is usually achieved quickly, accurately and effortlessly. Yet understanding exactly how this is achieved is no mean feat. Our visual systems have to make sense, and make sense quickly, of the mass of information such as shading, colour and contrast that greets our eyes.

For one thing, the image of a shape that projects onto the retina is constantly changing as we, or the object we are viewing, move within the environment. This means that we often see objects from a variety of perspectives and from a range of distances. This project investigates how a subset of these changes to the images in our visual field – i.e., alterations in size and some changes in orientation – affect our ability to recognize the objects that created them. How do we decide that two images depict the same object despite such changes in retinal stimulation? Although the problem is deceptively simple to state, it is difficult to solve. Shedding light on possible solutions, or at least illuminating paths along which such solutions may be found, will have a significant influence on our understanding of our internal representation of the visual world.

## 1.2 Two competing theories

Our ability to determine object identity despite such image transformations can be considered as a solution to a perceptual problem. There are at least two distinct ways in which a solution may be provided. On the one hand, we may use what is known as an *invariant* strategy. Even though the size and orientation of an object's image on our retina can change, the shape of the object in the real world remains constant. Therefore, one way to recognize an object is to pick out from the retinal image the properties that are unchanged under transformations such as rotation and scaling that identify the shape of the object. For example, a parallelogram is a four-sided polygon with opposite sides parallel. When you look at a depiction of a parallelogram on a page, none of these properties change when you bring the page closer to your eyes or if you turn the page upside down. An invariant strategy relies on properties of this nature. The representation of these objects in long-term memory is therefore relatively abstract – they are not pictorial images but representations that are free of scale and orientation information (e.g., Biederman, 1987; Cavanagh, 1989; Cooper, Biederman, & Hummel, 1992; Marr, 1982; Marr & Nishihara, 1978; Schwartz, 1980a). This type of solution is an economical one – orientation and size are irrelevant to the problem so they are eliminated from the representational framework.

On the other hand, we may use what is known as a *normalization* strategy. Under this alternative approach, the representations of objects in our long-term memory possess a specific orientation and size (e.g., Bundesen, Larsen & Farrell, 1981; Huttenlocher & Ullman, 1987; Rock, 1973; Ullman, 1989). When we are faced with a different perspective on one of these objects in the visual world, we must alter the orientation and size of our representation of this stimulus until it matches the one we have in memory. In other words, we transform the stimulus representation by rotating and / or scaling it. For example, if the object in the environment is upside down, we need to rotate the representation of this stimulus until it matches the version we have in memory of this object in its upright position. In contrast to the invariant strategy, the long-term memory store of objects under the normalization hypothesis is less abstract and more pictorial.

These two strategies are traditionally seen as competing hypotheses in the psychological literature. However, we will examine the idea that the visual system may have both strategies at its disposal in order to recognize objects and that certain experimental variables influence the choice of strategy. More specifically, we will look at the evidence for the invariant and normalization strategies in recognizing planar rotated and scaled objects. Planar rotations involve rotations that occur in the plane perpendicular to the line of sight of the viewer. They do not include rotations in other planes. However, where appropriate we will discuss the relationship between different types of rotations in the context of object recognition.

This project will approach this problem from at least four different perspectives. Firstly, we will examine not only previous psychological theory and

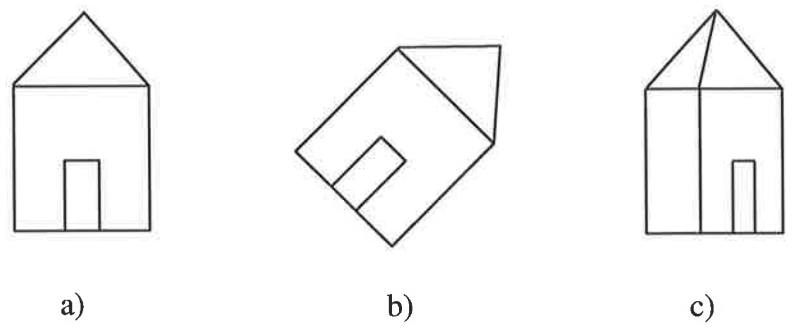


*Figure 1.1. Planar versus in-depth rotations. Image a) represents a frontal view of a simple, house-like object. Image b) represents the view after a planar rotation from the view in a). Image c) depicts the view after an in-depth rotation from the view in a).*

empirical data but conduct experiments into visual processing. This will involve alternative accounts and theories for a variety of empirical phenomena. Secondly, where appropriate, we will examine data from neurophysiological and neuropsychological studies in considering possible explanations for psychological effects and possible neural mechanisms for the visual processes observed. Thirdly, we will examine the question from an implementational perspective – can the commonly proposed psychological theories be realized computationally and, if so, in what manner? Does examining a more complete, working model of a particular theory actually support or condemn it? Fourthly, we will examine the philosophical implications of such theories and concepts in the broader context of understanding visual processes.

## 1.3 Definition of terms

The word *rotation*, unless otherwise specified, will refer to rotations within the picture plane. Rotations that occur in other planes relative to the viewer are referred to as *rotations in depth* (see figure 1.1).

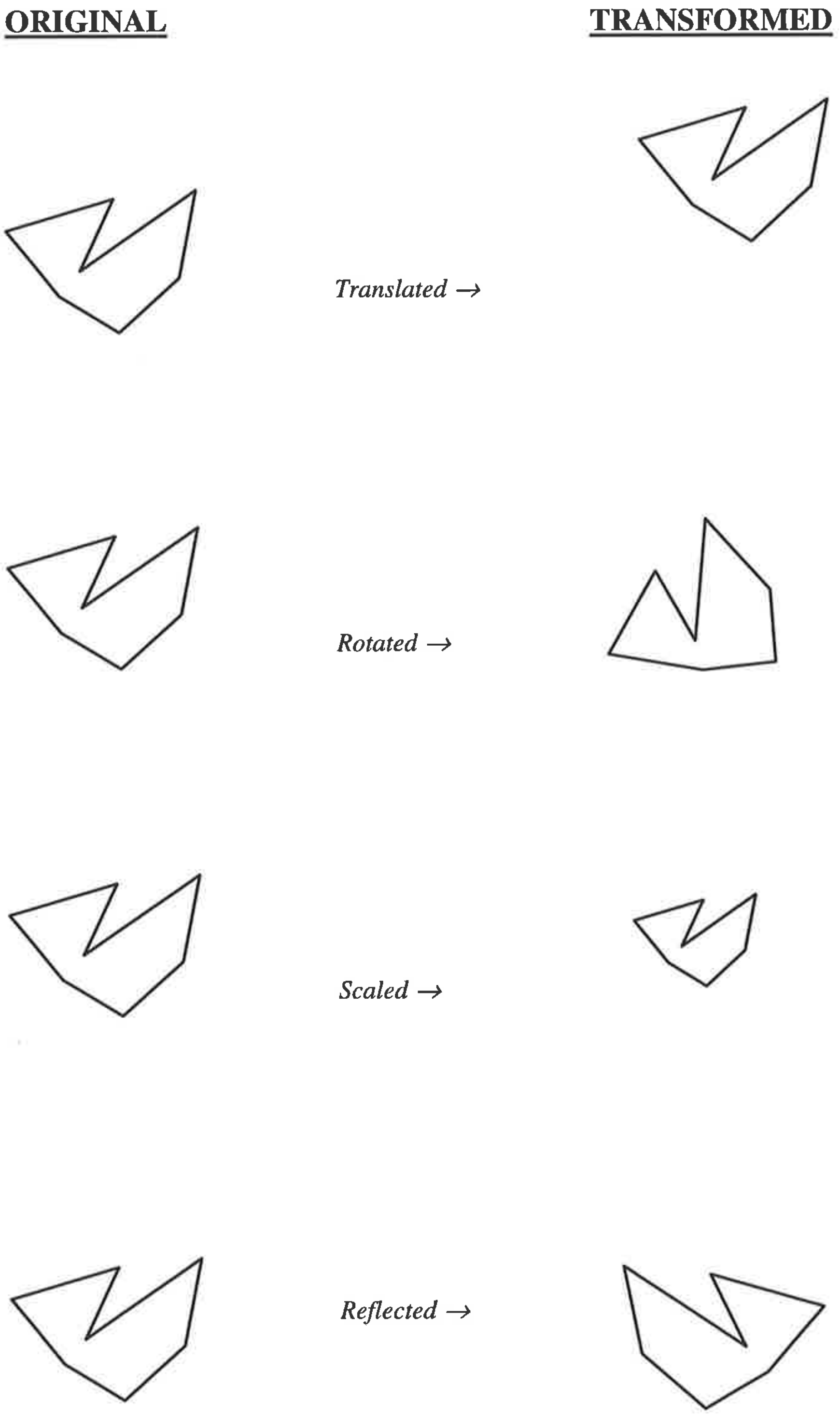


Figure 1.2. Examples of individual transformations from the similarity group.

The definition of *shape* is a contentious issue in the field of experimental psychology. Can the notion of perceived shape be defined purely in terms of the geometrical properties of the object being viewed (e.g., Shepard, 1994) or does it require considerations of subjective properties as well (e.g., Palmer, 1999)? If shape is a psychological property then it must incorporate a subjective element. However, it is difficult to define it completely in this manner. Therefore, for this project we will use a working definition of shape that refers only to the objective properties of the object being viewed. We will employ the definition of *objective shape* which is defined by Palmer (1989) (see also Shepard, 1994, p.3) as the spatial structure of an object that remains invariant or unaltered when any of the following transformations are applied to it:

1) ***translations*** – changes in an object's position
2) ***rotations*** – changes in an object's orientation
3) ***dilations*** – changes in an object's size or scale
4) ***reflections*** – changes in the "handedness" or left-right directionality of an object
5) any combinations of the above transformations

Examples of shape constancy under the transformations 1 through 4, and combinations thereof, are depicted in figures 1.2 and 1.3 respectively. This particular set of transformations form what is known in mathematics as the *similarity group* and defines the notion of shape used in this project.

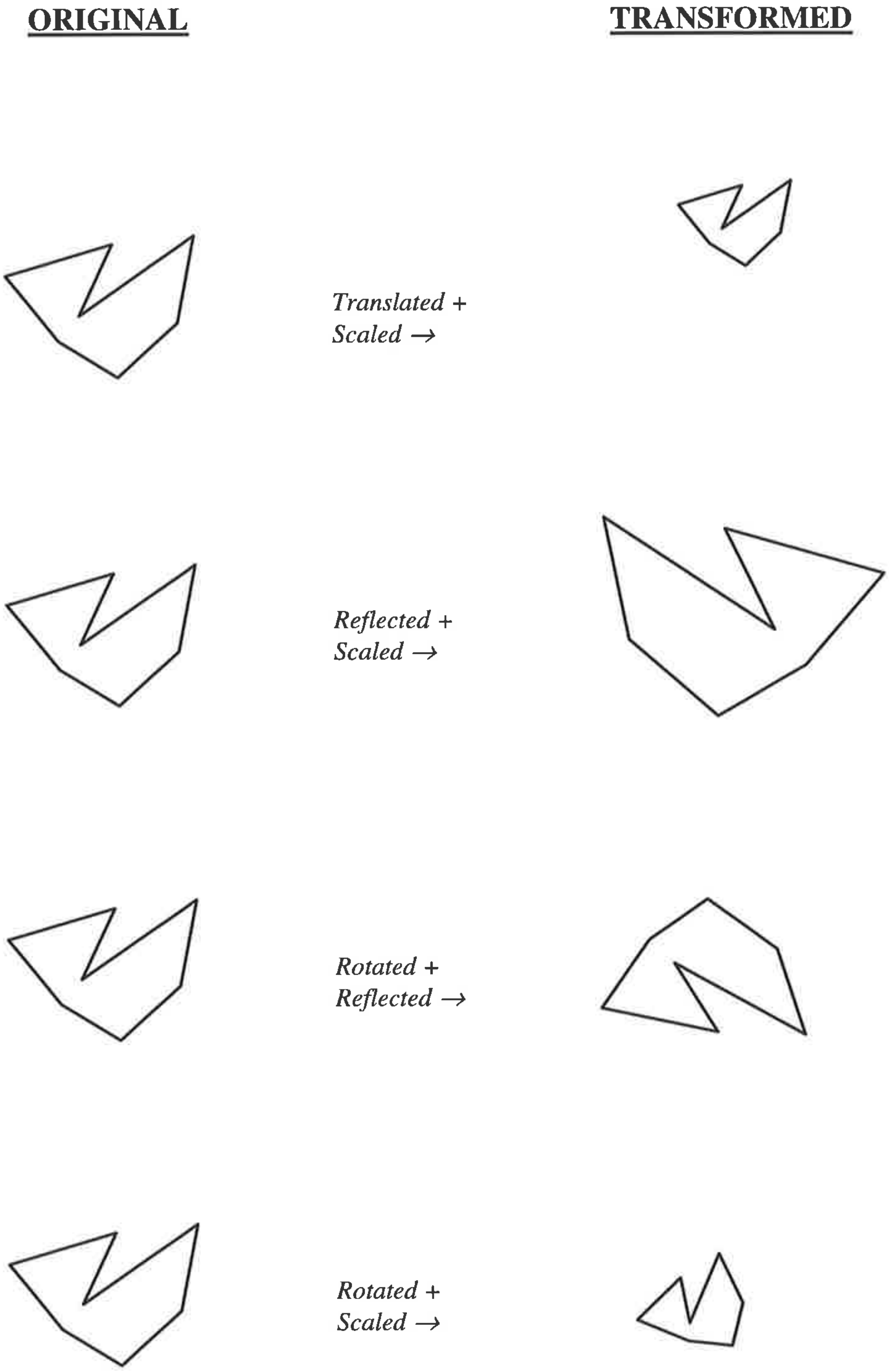


*Figure 1.3. Examples of combinations of transformations from the similarity group.*

## 1.4 Overview

This project is divided into a number of chapters. The following chapter examines a variety of possibly related perceptual phenomena that involve, not only the recognition of shape, but also some understanding of how our relationship with respect to the object in the environment can change. The third chapter discusses models of normalization, i.e., recognition mechanisms that involve compensatory orientation and size changes. The fourth chapter looks at the alternative approach to object recognition, namely models relying on invariant strategies. This also includes a model we have developed based on ideas from both neurophysiology and psychology. The fifth chapter investigates the effects of rotation on recognition in previous literature while the sixth chapter tests some of these ideas empirically. The seventh and eighth chapters look at the effects of size and the combined influence of rotation and size respectively in the literature on object recognition. The ninth chapter tests the combined influence of size and rotation on shape recognition using the same paradigm as the first experiment. The final chapter discusses the limitations in conventional psychological investigations on identifying objects and suggests ways of interpreting our findings in the broader context of visual processing. We will also propose a variety of further experiments that reflect the multi-disciplinary perspective of this thesis.

# CHAPTER 2: Motion-recognition continuum

*This chapter details a number of closely related visual phenomena that involve a) transformations of scale and / or rotation of an image in one or more static displays and b) knowledge of the object depicted by the images despite changes in these properties. We demonstrate how these phenomena can be meaningfully organized along a continuum. Traditional object recognition tasks are situated at one extreme of this continuum where processing is less automatic and emphasizes knowledge of the object over understanding the transformations applied to the image. All of these phenomena occur in remarkably similar experimental paradigms and may reflect similar mechanisms or principles in visual processing.*

## 2.1 A continuum of motion-recognition mechanisms

> *We visually perceive both a persisting object and its relation to us. We also recognize both the face of a friend and its momentary expression, both what has been written and the format in which it has been written, both what has been said and the emotional state of the speaker, and both a particular melody and the pitch and timbre at which it has been played. (Shepard, 1984, p. 421)*

This chapter explores some of the family resemblance between the perceptual mechanisms of *Apparent Motion, Representational Momentum, Backward Alignment, "Mental Transformation"* and *Transformation-to-match in other recognition tasks.*

We will describe a manner in which they can be meaningfully arranged along a continuum. All of these different perceptual phenomena are triggered by static displays that involve knowledge both of an object's shape and the way in which the retinal image has been transformed.

There are two main dimensions that can help us in categorizing these phenomena. Firstly, we can describe a mechanism as being influenced by *bottom-up* and / or *top-down* processing. Bottom-up processing refers to the direct influence of sensory information in triggering a response or mechanism. Such processing is highly automated. In contrast, top-down processing reflects the influence of higher cognitive involvement and is more controlled. Secondly, we can look at the relative emphasis of the mechanism on a) the recognition of the object itself and b) the motion or transformation that is applied to it. For example, a mechanism may place greater emphasis on analyzing the implied movement of an object than recognizing the object itself. This dimension echoes a common theme in the neurophysiological literature on visual processing - namely, that one system of the brain is specialized to recognize objects while the other analyses positional and movement information of objects. We will discuss this distinction, known as the 'what' versus 'where' division, in greater detail later in this chapter where there is evidence relating to the individual mechanisms.

These two descriptive dimensions may or may not be independent. However, for the closely related phenomenon that we are investigating, the two dimensions appear to be aligned. For this reason, it is convenient to order the phenomena along a single continuum that combines both dimensions. This *motion-recognition continuum* is represented in figure 2.1. This orders mechanisms with respect to a) the emphasis on either identifying an object versus interpreting the respective transformation applied to it and b) the extent to which a mechanism is automatic and stimulus instigated as opposed to more conscious and less data driven. As one proceeds from left to right along this dimension, each mechanism becomes gradually less associated with motion (and consequently more concerned with recognition of shape) and less automatic. Wherever there is comparable data, all the mechanisms display strong similarities in the way they deal with the symmetries of size and planar rotation.

All the transformational processes contained in this continuum reflect two of Shepard's (1981, 1984, 1994) principles of perception. Through the process of evolution, Shepard claims that our perceptual and cognitive apparatus has developed in a manner that reflects important characteristics of our physical environment. Such characteristics include the notions that material objects are generally conserved when they move (i.e., they do not change shape) and that the movements between any two positions tends to be the shortest (the *principle of least action*). The mechanisms of the *motion – recognition continuum* comply with these two principles because (a) the transformations inferred from the static display(s) conserve shape (i.e., the transformations are all shape preserving symmetries from the symmetry group) and (b) the transformations all involve the simplest, shortest transformational path to match. Although Shepard analyses only the mechanisms of apparent motion and "mental transformation" from our *motion - recognition continuum*, the other members of this group are largely bound by the same principles. These universals will be indicated in the description of each mechanism.

In all of these tasks, what the subjects are asked to do or report may be different. For example, in an *apparent motion* task, subjects are asked whether they have perceived motion whereas in an *object recognition* task subjects are asked whether the object is the same in both displays. However, what the subjects actually see in these experiments is often remarkably similar and there are broad similarities in human performance in these tasks. We suggest that there may be similar principles or possibly even common strategies in visual processing at work in all these phenomena. Therefore, a greater understanding of processing in pure object recognition tasks may be gained by looking at processing that occurs in related tasks.

## 2.2 Apparent motion

Apparent motion refers to the perceptual experience of movement from the successive presentation of at least two static displays. For this phenomenon to occur, the images are presented without any interstimulus interval (ISI), i.e., they are presented in direct succession without any pause during which the screen or display is blank. In addition, the rate at which the display switches between the two images (called the stimulus onset asynchrony or SOA) is relatively fast, ranging from 25 to 550ms. For example, Farrell and Shepard (1981) used this paradigm to present images of random polygons separated by a rotation in the plane. The subjects in this experiment reported the perceptual experience of a single, rotating object.

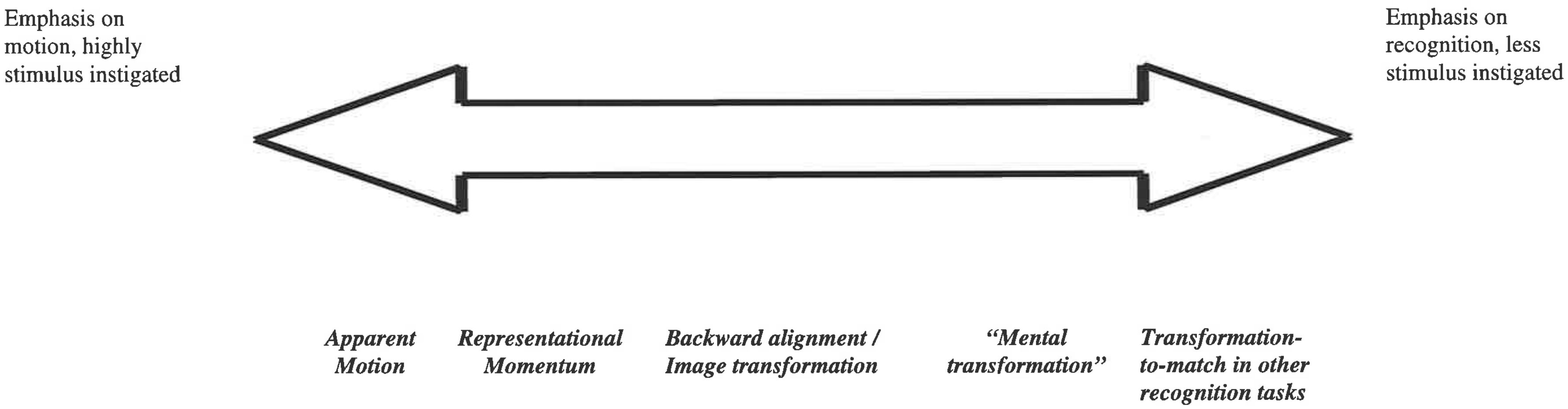


*Figure 2.1. Motion-recognition processes in perception as conceived along a continuum. This continuum represents the degree of automation and relative emphasis on motion versus recognition.*

Certain conditions must be met with regards to the timing of stimulus presentation and the size of the transformation applied to the object in order for the apparent motion effect to occur. More precisely, there is a linear increase in the time of presentation of alternate views with the angular difference in orientations of the successive displays. This relationship has been illustrated both for depictions of two-dimensional objects (Farrell & Shepard, 1981) and three-dimensional objects (Shepard & Judd, 1976) rotated in the picture plane. This reflects a common feature of apparent motions known as *Korte's Third Law* which states that, within upper and lower bounds, the alternation rate and the degree of transformation required for the illusion of motion are inversely related (Korte, 1915). This suggests that there is either a preferred rate or a limited range of rates implied by the static displays necessary for the phenomenon. For example, this rate is 1000 deg/s (*degrees per second*) in Shepard and Judd's (1976) task. The direction of rotary motion experienced is always along the shortest rotational path between the two images. For example, if the angular deviation is shorter in the clockwise direction then the apparent motion occurs along this path and not anticlockwise. In this way, apparent motion is consistent with Shepard's principle of least action.

These results for rotated images strongly parallel those for apparent motion induced by size differences. Using the same experimental paradigm, Bundesen, Farrell and Larsen (1983) induced an illusion of translation in depth of a constant sized object using differently scaled images of the same shape. The minimum SOA necessary for apparent depth motion was found to increase linearly with implied distance in the third dimension (as depicted by the different sizes of the images). This is again concordant with Korte's Third Law and suggests that there is a preferred range of rates for perceived depth movement. Just like the apparent rotational effects, those for size comply with the two principles of Shepard mentioned previously. Firstly, the variation in image size is resolved as a change in the distance from the observer of an object of a constant distal size. In other words, the object is conserved, i.e., its real world size is unchanged, as opposed to the perception of an object expanding or contracting like a balloon. Secondly, the apparent motion always occurs along the shortest transformational path to match between the two images.

These principles were upheld when Bundesen et al. (1983) examined the apparent motion effects associated with images that were both rotated *and* scaled. Shepard (1981) demonstrated that the shortest transformation path to match between two images distinguished by changes along both of these dimensions was a “helical or screwlike rigid motion in depth” (p.314). This involves transformation along both dimensions simultaneously much like screwing the top on or off a bottle. Correspondingly, subjects in Bundesen et al.’s (1983) experiment experienced the illusion of a single object concurrently rotating in the plane and translating in depth such that when it “appeared at intermediate distances, its apparent angular orientation was intermediate between those of the two stimuli” (Bundesen et al., 1983, p.551). Interestingly, the transformation times for the two symmetries were additive rather than interactive in the combined condition. The authors argue that the path to match is in fact composed of alternating, small steps of either simple rotations or simple depth translations. However, when these smaller steps are integrated the overall path approximates a helical motion.

One issue faced by theories of apparent motion is how the visual system knows which object in one display is a transformation of an object in another. This is known as the *correspondence problem of apparent motion*. In fact, this type of problem is common to all of the perceptual mechanisms in this *motion - recognition continuum*. Palmer (1999, p.474-477) claims that the shortest transformational path to match can sometimes influence the determination of object constancy between presentations, as demonstrated by the *wagon wheel illusion*. This involves the unusual apparent motion effects associated with the wheels of a stage coach often seen in Western movies. The film itself is made up of static images presented at a rate fast enough to provide, for the most part, good apparent motion (known as the *stroboscopic effect*). However, as the wagon begins moving and accelerating in a uniform direction, the wheels appear to (a) go forward, then (b) stop moving and then (c) travel in a direction opposite to that of real life motion.

Palmer (1999) argues that the shortest rotational path to match between any of the identically shaped spokes of the wheel determines these apparent motion effects. In (a), this path is consistent with the real world direction of travel and accurately captures the movement of the spokes. However, as the speed of rotation increases, the limited sampling rate of the film begins to induce inaccurate perceptions of the real-life movement. In the snapshots presented in (b), the path between any of the identical spokes is negligible, resulting in no motion detection. And in (c), as the rate of rotation increases further still, the shortest path to match between the identical spokes is opposite in direction to the real world motion resulting in the appearance of the wheels "going backwards". Palmer's (1999) explanation suggests that, when faced with a group of moving objects that are all physically similar, perceptual identity can be governed by the shortest transformational path to match. However, in other cases it is not clear how the *correspondence problem of apparent motion* is solved.

This phenomenon also demonstrates the automaticity of the apparent motion experience – despite a high level understanding of the real world direction of the wheel's travel, the illusion which contradicts this is unavoidable. Another way in which the bottom-up nature of apparent motion processing can be established is by drawing similarities between apparent motion and real motion which is generally regarded as highly automatic. For example, Braddick (1974, 1980) claims that the illusion of motion is caused by the triggering of motion detectors early in the visual pathway. In other words, the phenomena of apparent and real motion may share the same perceptual apparatus. This notion has received support from Anstis, Giaschi and Cogan (1985). In their task, adaptation to real motion suppressed the experience of apparent motion which suggests a common underlying neural mechanism for the two effects. Similarly, Green (1983) has found interaction effects between the two phenomena. For example, the apparent motion effect is enhanced when real motion occurs in the same direction as apparent motion. Similarly, when the two phenomena imply opposite directions of movement the apparent motion effect is effectively cancelled. Therefore, there is some compelling evidence that apparent and real motion are closely related in psychological and neurophysiological terms and that both are strongly influenced by bottom-up processes. However, it should be noted that recent research has suggested that motion analysis in general can be the product of both top-down and bottom-up processing. Lu and Sperling (1995) have demonstrated that there are three different types of motion analysis – two of which are purely bottom-up processes and one, relatively slower mechanism, that is influenced both by bottom-up and attentional influences.

## 2.3 Representational Momentum

The next effect in the continuum is Freyd's (1987) *Representational Momentum*. This involves a distortion in the memory of static images that is consistent with an analysis of the dynamic information in these displays. It is apparent when subjects are presented with two or more successive displays separated by ISIs of up to 2s and where these displays contain images differentiated only by certain, shape preserving transformations (Finke & Freyd, 1985). Subsequent testing reveals a distortion in memory of the last image that is consistent with the implied movement between the images in the two displays.

For example, Freyd and Finke (1984) [Experiment I] presented four depictions of the same rectangle presented at different planar rotations around the center of the display. Each image was presented for 250ms and the ISI was also 250ms. While the first three images depicted 'snapshots' of the rectangle at a constant rate of rotation (e.g., 0.3 rad, 0.6 rad, 0.9 rad), the final image could either be a) in the same orientation as the third (e.g., 0.9 rad), b) further along the implied trajectory (e.g., 0.10 rad) or c) oriented in such a way that it depicted a direction of rotary movement relative to the third image that was opposite to the direction implied in the first three displays (e.g., 0.8 rad). Subjects were asked to make a same / different discrimination between the third and fourth images and the final display remained visible until such a decision was made. The results from Freyd and Finke's (1984) [Experiment I] experiment demonstrated that subjects were more far more likely to incorrectly choose the image in b) whose orientation was consistent with the implied trajectory of the first three displays (43.9% of trials) than that in c) where the implied trajectory is in the opposite direction (6.4%). In other words, the internal representation of the object appeared to possess its own momentum associated with the implied trajectory in the static displays. This momentum is analogous to the physical inertia of external physical objects. Freyd (1987) claims that this memory distortion is evidence that dynamic information has a “special status in mental representation as well as in perception” (p.427).

The effect of representational momentum has also been successfully demonstrated for changes in size using a similar experimental paradigm. Kelly and Freyd (1987) [Experiment 5] used four successive presentations of identical squares, the first three of which depicted samples from a constant rate of translation in depth – i.e., either increasing or decreasing in size at a constant rate along the line of sight of the observer. The representational momentum phenomenon was again evident with the expected memory distortion along the path of implied motion, i.e., the participants were more likely to incorrectly match an image depicting a size greater than that evident in the third display if the implied motion was that of translation in depth towards the viewer (i.e., increasing in size) and vice versa. In addition, all 16 subjects reported that the sequence of images in a trial invoked a subjective experience of the object retreating or looming in depth.

The representational momentum effect for size has also been demonstrated using computer-generated stereoscopic images, i.e., with the influence of perspective and binocular disparity information (Hayes, Sacher, Thornton, Sereno & Freyd, 1996). One study by Hubbard (1995) demonstrated that Freyd's (1987) phenomenon can interact with other perceptual phenomena. The author argued that the reduction in the representational momentum effect associated with approaching targets in their study was the result of a perceptual effect known as *boundary extension*. Under this effect, subjects remember the scene as if it had been presented in a wider-angle view than the actual image. In other words, the boundaries in a scene appear to be expanded such that stationary objects are remembered as being further away than in the pictorial representation.

The phenomenon of representational momentum is reliable only for images with the same objective shape – if the objects that are depicted are entirely different (e.g., rectangles, triangles and rhombi) the effect does not occur (Kelly & Freyd, 1987 [Experiment 2]). In fact, even for different rectangles (i.e., distinguished by varying height to width ratios), the representational momentum effect is reduced (Kelly & Freyd, 1987 [Experiment 3]). This suggests that only certain transformations can reliably invoke this phenomenon (including the shape-preserving symmetries of planar rotation and scale) and that the determination of identity appears to be a requirement for the dynamic element of the mental representation. Because the phenomenon is available only when the transformations are shape-preserving, and not for other transformations that would deform the object in ways that would change the objective shape, the effect is consistent with Shepard's notion of object conservation.

Freyd's (1987) effect possesses some important similarities to apparent motion. Representational momentum is highly data driven as demonstrated by its immunity to error feedback and practice (Finke & Freyd, 1985). In addition, the phenomenon occurs quickly and is evident after extremely short retention intervals (10 to 100ms) between the test and preceding stimulus (Freyd & Johnson, 1987). As with all the phenomena in this continuum, the perceptual illusion is always associated with the shortest transformational path to match implied by the procession of images and is therefore consistent with Shepard's principle of least action. However, Freyd (1987) has argued that representational momentum occurs at a higher level of processing than apparent motion. Whereas apparent motion is largely the product of bottom-up processes (Lu & Sperling, 1995), representational momentum involves a higher level of control as evidenced by the long retention intervals (Finke & Freyd, 1985) and the use of ISIs, both of which are absent from conditions necessary for apparent motion. In apparent motion, the dynamic effect is far more transitory and the images are presented in direct succession without any blank gap in-between.

## 2.4 Backward Alignment

Apparent motion and representational momentum involve either a direct perception, or memory, of movement from consecutive, static images. As we move on to the next phenomenon in the continuum, *backward alignment*, the emphasis shifts more clearly onto behavioural responses involving object knowledge that imply, in a more indirect manner, an underlying transformational process or motion. Backward alignment is a phenomenon first demonstrated by Koriat and Norman (1988). Like representational momentum, this effect requires more than one successive presentation of images that are differentiated only by a shape preserving transformation. The results also suggest that the perceptual system utilizes a similar internalized transformational process between the images. However, the evidence is demonstrated indirectly through response latencies because the task itself is one of form recognition or verification.

In the backwards alignment task, subjects are required to choose one of two possible responses. If the two images in the sequence depict an object that has been only planar rotated then subjects are required to provide a 'match' response. However, if the two images depict an object that has been both planar rotated and mirror-reflected, subjects are required to provide a 'foil' response. Evidence for a transformational process is derived from response latencies in providing these responses. The results from this type of experiment have revealed that the time to respond to the second pair of the sequence often reflects the angular difference between the two images (Koriat & Norman, 1984, 1988; Koriat, Norman, & Kimchi, 1991). In other words, response latencies are proportional to the *angular deviation from the preceding orientation (ADP)*. This suggests that subjects may have verified the objects identity via the shortest transformational path to match between the successive pair of images. Note that the rotational effect is evident for both standard versions of the images and their mirror-reflected counterparts suggesting that, for the 'foil' images as well, subjects rotate the image before testing for a match. The backward alignment phenomenon has been investigated in only a small number of studies (Koriat & Norman, 1984, 1988; Koriat, Norman & Kimchi, 1991).

## 2.5 "Mental Transformation"

### 2.5.1 Mental rotation

The terms *mental transformation, mental size scaling* and *mental rotation* have been used in a variety of contexts in the literature, often in a very general sense. In this project, we will define the term mental rotation with respect to the classical 'mental rotation' experiments first conducted by Shepard & Metzler (1971). Mental size scaling will be defined as the phenomenon that is ostensibly similar to mental rotation with the difference that it involves size changes and not rotations. Given the similarities between these mechanisms, we will subsume both under the more general term "mental transformation". This term is presented in inverted commas because it is used widely in the literature to describe what we believe may be different, albeit closely related, phenomena.

Koriat and Norman (1988) argue that mental rotation is closely related to backward alignment. In fact, they found evidence that both mechanisms can act simultaneously within the same experiment. According to their research, the two procedures utilize similar rates of rotation and occur in experimental paradigms characterized by verification ('match' and 'foil') responses. In these tasks, the foil images (also called distractors) are mirror-images of the standard versions and both types of images are shown at different orientations in the picture plane. In the mental rotation and backward alignment effects, both the standard and mirror-reflected versions are mentally rotated. The main difference between the two phenomena is the image with respect to which the current stimulus image is aligned. In backward alignment, the current stimulus image is aligned with the short-term representation of the previous stimulus image. However, in mental rotation the current stimulus is brought into alignment with the long-term representation of the same object in its upright position - i.e., the way in which it was presented and memorized during the memory stage of the experiment. Koriat and Norman (1988) were able to discriminate between the effects of the two mechanisms by determining whether the time to respond was consistent with rotation to the upright (i.e., mental rotation) or rotation to the former stimulus orientation (i.e., backward alignment). In other words, both mechanisms appeared to act in parallel and the mechanism which finished first was that which prescribed the shorter angle to match. Given the differences in memory storage between the two mechanisms, Koriat and Norman (1988) propose that mental rotation is a higher level procedure than backward alignment and that it is both more data driven and more closely associated with apparent motion (p.95). This statement is consistent with our ordering of these mechanisms along the *motion – recognition continuum*, i.e., that apparent motion is more automated than backward alignment which, in turn, is more automated than mental rotation.

### 2.5.2 The basic mental rotation effect

The "mental transformation" effect can occur with testing of a single presentation in what is commonly known as the *handedness task*. An early study by Cooper (1975) demonstrates the classic mental rotation effect within a handedness task that utilizes the types of stimuli and rotations that will feature in the experiments of this project - random polygons under rotations in the picture plane. At the beginning of the experiment, the subjects were asked to study a diagram, featuring the eight forms of the experiment, for approximately 10 minutes. These shapes were presented in both standard and reflected (through the vertical axis) or ‘mirror image’ versions at their standard, upright orientations. Then, before each test block, subjects practised the speeded task. This involved discriminating between the two types of forms (“standard” or “reflected”) by pressing a right-hand or left-hand button respectively. In the practice trials, all forms were shown in their standard orientation only, whereas in the test trials the form could occur at six equally spaced orientations – the standard orientation plus $60^0$, $120^0$, $180^0$, $240^0$ and $300^0$ angular departures of clockwise rotation from the trained orientation.

The resulting data for correct responses demonstrated a striking linearity in the increase of reaction time with angular departure from the test stimulus. In addition, “standard” responses were consistently shorter than “reflected” responses although both response profiles were of a similar slope. This pattern of results was replicated for two- and three- dimensional objects by Metzler and Shepard (1982). Cooper (1975) claimed that these results support the idea that (a) an identical visual image or “copy” of the currently presented test form is mentally rotated from its current orientation into the trained orientation (a notion supported by the subjects’ introspective reports) and (b) that this image is compared to the memorial representation of the standard image – a match is initially tested for but if this fails additional time is required to switch to the “nonpreferred ‘reflected’ response” (Cooper, 1975, p.28). In this experiment, the rate of rotation was estimated to be a constant rate of $460^0$ per second.

### 2.5.3 The analogue nature of mental rotation

The same stimuli used in Cooper's (1975) experiment were also utilized in a study that demonstrates the analogue nature of mental rotation (Cooper, 1976). In this context, the word analogue refers to a similarity between the mechanism of mental rotation and the actual perception of a rotating object. This means that, during the mental rotation procedure, the stimulus is represented at intermediate angles between the original stimulus orientation and the final image orientation. The same subjects from the earlier experiment were asked to imagine a stimulus rotated clockwise in the picture plane, following an initial presentation of the stimulus for 2 seconds. Subjects were then probed at certain intervals of time with the same shape or its mirror-reflection at various orientations and asked to provide a verification response (i.e., "standard" or "reflected").

There were two types of trials in the experiment. *Probe-expected* trials consisted of stimuli oriented in accordance with the predicted orientation of the subject's mental image at that particular point in time. This orientation was estimated from the previous experiment's (Cooper, 1975) data on rotation rates and was calculated for each individual subject. *Probe-unexpected* trials consisted of stimuli at some orientation that deviated from this prediction. For probe-expected trials, there was no significant difference between response times across the different orientations. This suggests that handedness verification was quick and independent of angular departure from the trained orientations when there was a correspondence between the orientation of the imagined and probed stimuli. In contrast, the times to respond to probes occurring at 'unexpected' orientations were always longer and displayed a linearly proportional relationship with angular deviation from the expected orientation. This linear trend was in agreement with the rotation rates found in the original experiment on mental rotation (i.e., Cooper, 1975).

These finding are consistent with the notion that in mental rotation, the process passes through a series of intermediate states that are comparable to the intermediate stages in the perception of a rotating object. In addition, at each of these intermediate states there is a preparedness to respond to an image presented at the particular orientation. Cooper (1976) points out that although the process appears to be analogue, this does not necessarily entail a structural isomorphism between a mental image and physical object. What the results do demonstrate is that there is a controlled preparedness across orientation changes that is analogue in nature. As Shepard and Judd (1976) have pointed out:

> *"to say that the internal process is a mental analogue of an external process is, in part, to say that the internal process is similar in important respects to the perceptual process that would take place if a subject were actually to watch the corresponding physical rotation" (p.952)*

### 2.5.4 Mental rotation as a higher level procedure

In comparison to the mechanisms of apparent motion and representational momentum that appear earlier in this continuum, mental rotation is relatively less automatic and stimulus instigated. For example, Just and Carpenter (1985) have proposed that it is a high-level procedure that is at least partially under the subject's control. They have demonstrated both individual and strategic differences between subjects on mental rotation tasks. Similarly, the fact that mental rotation is unaffected by luminance, colour, texture, relative orientation and binocular disparity suggests that it occurs at a level of processing higher than low-level motion analysis which is highly stimulus instigated and automatic (Jolicoeur & Cavanagh, 1992). A more extreme view has been taken by Pylyshyn (1981) who has argued that, not only is mental rotation a deliberate strategy, but it is one based on the conscious manipulation of abstract symbols which is guided by an understanding of naturally occurring transformations.

However, mental rotation and lower level motion illusions such as apparent motion still appear to share the same internal framework for representing internalized motion or transformation (Corballis & McLaren, 1982; Jolicoeur & Cavanagh, 1992; Jolicoeur, Corballis & Lawson, 1998). For example, Jolicoeur, Corballis and Lawson's (1998) study investigated the influence of two motion effects on a handedness task. In the first experiment, subjects adapted to rotary motion using a wheel stimulus presented before each trial. Adapting to such a rotating image causes a motion illusion on a static stimulus presented shortly afterwards and the direction of this implied motion is opposite in direction to the wheel's rotation. In the second experiment, the shape was presented in a manner to induce an apparent rotation effect, i.e., 8 views of the same shape presented for 14ms apiece without any ISI and each separated by $2^0$ along a rotational path of uniform direction (clockwise or anticlockwise). The results demonstrated that the perceived low-level rotary effects influenced the response times to the handedness task. Specifically, the motion illusions appeared to alter the rate of mental rotation. This finding suggests that the mechanisms of mental rotation, apparent motion and the rotary after-effect may share some common structures or resources within the brain (see also Corballis & Blackman, 1990).

There is also evidence that mental rotation and representational momentum may be linked in the visual processing system. Munger, Solberg and Horrocks (1999) found that when the rotational axes were aligned with the viewers' coordinate system, mental rotation rates and the size of memory distortions in a representational momentum task were both increased. Similarly, there was evidence of a correlation between rates in the two tasks with participants with slower rates of mental rotation also using smaller memory distortions in the representational momentum task. Such evidence may indicate important similarities in the way the visual system performs mental rotation and representational momentum.

Studies into the neurophysiology of mental rotation have also provided evidence that it is related to lower-level motion phenomena. For example, two studies have implicated the role of the premotor and supplementary motor areas in the mental rotation task (Cohen, Kosslyn, Breiter, Di Girolamo, Thompson, Anderson, Bookheimer, Rosen & Belliveau, 1996; Richter, Somarjai, Summer, Jamarsz, Menon, Gati, Georgopoulos, Tegeler, Vgurbill & Kim, 2000). Cohen et al. (1996) have noted that the areas recruited for the procedure are similar to those involved in direct perception and in the detection of motion (p.92). In other words, some common neuronal structures are used in these different tasks. Interestingly, this finding echoes Shepard's sentiments that mental rotation reflects some kind of internalization of the principles of physical motion - it appears to involve similar perceptual apparatus. In summary, while there is evidence that mental rotation is a higher level procedure than apparent motion, there is psychological and neurophysiological evidence that it is nevertheless closely related to lower-level motion phenomena.

### 2.5.5 Mental size scaling

There has been comparatively little study into the effects of scale in similar experimental paradigms to the classic mental rotation task. However, while some of the phenomena associated with mental rotation effects have yet to be investigated for scale changes, there is nevertheless some evidence of a comparable "mental transformation" effect for size differences. For example, Howard and Kerst (1978) examined performance on a successive presentation task where the pair of stimulus figures varied systematically in overall size. As with Cooper's (1975) task, the foil images were simply mirror-reflected versions of the first form of the pair in each trial. Response latencies were proportional to the linear increase in size ratio for both response types. Thus, like the mental rotation experiments, there is evidence for a transformational procedure that operates for both match and foil images when the foil images in the task are mirror-reflected versions of the target stimuli. Such similarities in findings for rotation and scale changes have prompted Bregman (1977) to propose a common transformational process operating on the stimulus image to test for a match with the expected image in memory. Presumably, this expected image is based on the image presented in either the first pair of the sequence (e.g., Howard & Kerst, 1978) or at the start of the experiment (e.g., Cooper, 1975).

As with mental rotation, the analogue nature of mental size scaling has also been established (Rösler, Heil, Bajric, Pauls & Hennighausen, 1995). Rösler et al.'s (1995) study is important for two reasons. Firstly, it found fundamental similarities in psychological processing between the simple effects of rotation and scale within the same experiment. Secondly, it not only examined their effects on behavioural responses, but it also measured corresponding patterns of cerebral activation. This was achieved via event-related brain potentials (ERPs) which measure electrical activity in the brain through electrodes placed on the scalp. The results from this study demonstrate a degree of correlation between "mental transformation" for planar rotation and size, both in psychological *and* neurophysiological measures.

### 2.5.6 "Mental transformation" and neurophysiological correlates

The experimental paradigm used by Rösler et al. (1995) was similar to that used in Cooper's (1976) study described previously. All subjects took part in both size and rotation conditions which were separated into different blocks. Standard stimuli (smallest size and upright orientation) were first presented for 3 seconds followed by a tone of one of three different frequencies (1046 Hz, 1397 Hz or 1976 Hz). Depending on which condition subjects were participating in, they were required to either rotate or scale an image of the stimulus in one of three ways indicated by the pitch of the sound – $0^0$, $60^0$, or $120^0$ for the rotation condition and 1:1, 1:3 or 1:5 for the scale condition. In the pilot study, subjects responded when this imagery process was complete. The mean latencies for these responses were linearly proportional to the degree of transformation required for *both* rotation and scale conditions. In addition, matching times to a probe stimulus that was identical to the image subjects were asked to prepare revealed no differences in latencies associated with the various transformations. In other words, it appears that for size as well as rotation, images can be prepared and subsequently used in a matching task and that the time to prepare the image is proportional to the degree of transformation required.

In the main experiment, the ERP data also showed an interesting similarity in brain activity during the rotating and scaling imagery tasks. In both conditions, there was a negative potential over the occipital cortex that remained constant over the task duration. Meanwhile, negative activity over the parietal cortex appeared related to the process of image transformation. In general, the amplitude of the negative potential was related to the amount of transformation required although the amplitude associated with the largest size scaling condition of 1:5 was relatively less than expected. Overall, Rösler et al.'s (1995) study demonstrates both a psychological and neurophysiological similarity in the way orientation and scale transformations are accounted for in such "mental transformation" experiments.

Rösler et al.'s (1995) study is part of a growing body of research that implicates the role of the parietal cortex in the "mental transformation" process. Several studies have also found a relationship between the magnitude of the negative potential associated with the parietal cortex and the degree of transformation required in a mental rotation task (Farah & Peronnet, 1989; Peronnet & Farah, 1989; Rösler, Schumacher, & Sojka, 1990; Yoshino, Inoue & Suzuki, 2000). Like Rösler et al.'s (1995) study, these experiments have also measured brain activity using ERPs. However, the experimental paradigm used in these studies does not involve explicit instructions to transform an image but rather the discrimination of standard and reflected shapes shown at different orientations as in Cooper's (1975) experiment. Another measure of cortical activity known as functional magnetic resonance imaging (fMRI) has also demonstrated the role of the parietal cortex in mental rotation as well as the involvement of motor areas of the brain (Cohen, Kosslyn, Breiter, Di Girolamo, Thompson, Anderson, Bookheimer, Rosen & Belliveau, 1996; Richter, Somarjai, Summer, Jamarsz, Menon, Gati, Georgopoulos, Tegeler, Vgurbill & Kim, 2000). This technique records neuronal activity, not by monitoring electrical activity, but by measuring changes in elements such as iron and oxygen.

Interestingly, the parietal cortex which is implicated in the "mental transformation" process and, in particular, the transformational aspects of this task, is not generally associated with object recognition. Rather, it forms part of what is known as the dorsal stream of visual processing in the brain (Goodale & Milner, 1992; Goodale, Milner, Jacobsen & Carey, 1991; Mishkin, Ungerleider & Macko, 1983; Ungerleider & Mishkin, 1982). This pathway connects the striate, prestriate and inferior temporal areas. The other stream of visual processing is known as the ventral system. This pathway connects the striate, prestriate and inferior parietal areas. These two streams contain qualitatively different representations of visual images. On the one hand, the dorsal stream appears to be associated with visuo-motor processes and our ability to manipulate objects. On the other hand, the ventral stream of visual processing involves pure object recognition and the long-term store of object representations (Ashbridge & Perrett, 1998; Goodale & Milner, 1992; Goodale et al., 1991). The association of mental transformation with the dorsal system suggests that the representation used for these tasks may be more similar to that used in visuo-motor processing rather than pure object recognition.

Accordingly, there is evidence that the dorsal system and the ability to perform mental rotation with which it is associated, are not necessary for the recognition of planar rotated objects. Support for this notion comes from neuropsychological studies that have investigated the abilities and disabilities associated with damage to one or other of the dorsal and ventral pathways. Farah and Hammond (1988) reported a subject (known as R.T.) with damage primarily to the dorsal system (specifically a fronto-parietal lesion) who performed normally on the recognition of disoriented alphanumeric characters, words and line drawings but performed very poorly on three different mental rotation tasks. This dissociation between the ability to recognize rotated objects and the ability to mentally rotate such objects has also been supported by Farah, Hammond, Levine and Calvanio's (1988) study. They reported a patient (known as L.H.) who showed no deficiency in mental rotation after ventral damage (suffering a bilateral occipitotemporal lesion) despite severe deficiencies in object recognition. Taken together, the findings of Farah and Hammond (1988) and Farah et al. (1988) present a complementary pattern of abilities and disabilities associated with a different lesion site. Technically, this is known as a *double dissociation* and suggests that the two functions are functionally distinct and anatomically separate. These studies not only lend further support to the notion that the dorsal system is fundamental to the mental rotation process but that both the dorsal system and mental rotation are not necessary for the recognition of planar disoriented objects. Presumably, the system most likely to be primarily involved is the ventral one - a system that many believe contains object representations that are invariant to properties such as orientation and size (Mishkin, Ungerleider & Macko, 1983; Ungerleider & Mishkin, 1982).

## 2.6 Transformation-to-match in other recognition tasks

We will use the term *normalization* to describe the alignment phenomenon for "mental transformation" (including its close relative backward alignment) and transformation-to-match in other recognition tasks. Normalization simply refers to the process whereby an incoming stimulus image is transformed to the same orientation and / or size as a previously stored representation of that object. Normalization can also occur in the opposite direction, for example in backward alignment, where the image in memory is aligned with the present stimulus image.

At first glance, the "mental transformation" and *transformation-to-match in other recognition tasks* (which we will refer to as TMORT) phenomena appear remarkably similar. Both appear to involve some kind of alignment procedure whereby the current stimulus image is altered in size or orientation in order to bring it into alignment with a stimulus image in memory of the very same object. This is mostly apparent in the response time data where the time to respond is proportional to the angular or size difference between the two images. Occasionally, it is evident in error rates where the response accuracy is inversely proportional to the amount of transformation required.

However, there is a difference in the experimental paradigms used to demonstrate the "mental transformation" and TMORT phenomena. In the "mental transformation" experiments, subjects are required to discriminate between standard and reflected versions of rotated images. In addition, some experiments of this type also involve requesting the subject to “mentally rotate” or “mentally scale” an image. Neither of these properties is present in experiments where the TMORT phenomenon is observed. Such experiments may instead require subjects to produce a semantic label (e.g., “dog” or “fish”) or to discriminate between pairs of images depicting the same objective shape from those depicting entirely different objective shapes.

In addition to such differences in the experimental paradigms (and probably because of them), there is also some evidence of a difference in the way the two perceptual phenomena are produced. Although both “mental transformation” and TMORT can be described as normalization mechanisms, there is evidence of a difference in the implementation of this principle. As mentioned above, Jolicoeur et al. (1998) have demonstrated that perceived rotary motion, whether in the form of a rotation after-effect or apparent motion, influences performance on the handedness task. For both rotary effects, the direction of implied motion influences the direction subjects choose to “mentally rotate” the image. However, when the task was one of naming using the very same stimuli, neither effect influenced the transformation-to-match process involved in naming. While the response latencies still supported the notion that subjects used some kind of normalization strategy for identification, the exact mechanism used in TMORT may be different from the one used in mental rotation. In other words, while mental rotation is closely related to lower level rotary effects in perception, the normalization procedure used in TMORT may be distinct from these phenomena and may occur at a higher processing level. As such, the TMORT phenomenon is placed at the latter end of the continuum where processing is, relatively speaking, least influenced by bottom-up processing. Jolicoeur et al.’s (1998) experiment demonstrates that, while the two phenomena appear outwardly similar, there is a difference in the underlying processing.

Another fundamental difference between the two phenomena is that while "mental transformation" is, for the most part, reliably and consistently evident in the "mental transformation" paradigm, the same cannot be said for the normalization effects in other recognition or identification tasks. There is convincing evidence of processing that does not require size and / or orientation normalization in these recognition tasks. In other words, object recognition is possible that is invariant to the stimulus properties of orientation and / or size (i.e., an *invariant* recognition strategy). In chapters 5,7 and 8 we will examine evidence for such mechanisms for recognizing objects under planar rotations, scalings and combinations of the two transformations and discuss the conditions necessary for the use of such procedures. We will also discuss exactly *how* the differences in paradigms and other experimental factors might influence the processing strategy used. For example, there are several factors that appear to determine whether or not normalization takes place in recognition tasks outside of the mental transformation paradigm. For example, there is evidence for individual differences in the use of either strategy in these experiments (McKone & Grenfell, 1999). In addition, repeated practice on these tasks can bring about a qualitative change in visual processing that is much harder to achieve, and perhaps impossible in mental rotation tasks. Both of these factors support the idea that processing in these paradigms is not purely data-driven and occurs at a higher processing level than that in the "mental transformation" experiments.

The presence of a mechanism that is orientation and size-invariant in these tasks is consistent with another trend in the continuum. We mentioned at the start of this chapter that, as we moved along the continuum, the emphasis shifted from the influence of possible motion from the static images to recognition of the shapes themselves. A recognition mechanism that is size and orientation invariant does not involve any normalization or analysis of possible motion or the internalization of any such motion. What it extracts is information that relates only to the property of shape which, according to our working definition, is unaffected by changes in orientation and size. In contrast, any normalization mechanism involves an analysis of the transformation that has been performed on the image or at least some kind of internal transformation of the stimulus image. Therefore, as we find ourselves at the end of the continuum, there is evidence of perceptual processing which reflects, not the transformations of rotation and size that have been applied to an image, but only the unchanging property of shape contained in it. In other words, recognition of shapes shown at different sizes and orientations that does not require any normalization.

In summary, using stimuli that are rotated and / or scaled, performance in a traditional object recognition task may be related to performance in other visual tasks. Not only is what the subjects sees in the different tasks often remarkably similar but there are gross similarities in how subjects respond. The analysis above raises the possibility that the transformational procedure sometimes evident in human performance when subjects are identifying objects may be related to the analysis of implied motion. It also suggests that such a procedure is not always necessary, i.e., processing is less automatic in object recognition tasks and there is a greater emphasis on object knowledge rather than image transformation. In other words, in addition to a normalization strategy, there is also evidence of invariant object recognition that is independent of orientation and scale.

# CHAPTER 3: Normalization theories of object recognition

*This chapter investigates how the idea of normalization for object recognition can be realized computationally (i.e., as a complete, working model) and details one popular approach to this problem. Unlike underdetermined notions of recognition such as "mental transformation" (e.g., mental rotation and mental size scaling) which also propose a transformation to match procedure, this approach has the ability to account for an initially perplexing feature in the data – namely, that we seem to know what the stimulus image represents before we have even transformed it. In addition, it solves many of the problems that vex other theories of normalization for object recognition. However, the theory is not without its limitations and we will discuss these in turn.*

## 3.1 Some problems in defining a working model of normalization

As mentioned previously, there are a large number of experiments investigating object recognition that are traditionally regarded as being supportive of the normalization hypothesis (for examples see chapters 5,7 and 8). In these experiments, the objects are presented either at varying angles from the upright or at various sizes and subjects are asked to provide a recognition response. For example, subjects might be asked to name the item (e.g., "dog"). Alternatively, they may be asked to respond whether an image that is currently visible represents the same object as the preceding image, whether it represents the same object as the image placed next to it or whether it represents the same object as an image shown earlier in the experiment (e.g., whether it belongs to a set of images which the participants were asked to remember). In these experiments, the time to respond (and sometimes the accuracy) with which an object is recognized is often a linear function of the transformation applied to the image. In other words, the time to respond and the number of errors increases the further the image is rotated from the upright or the extent to which it is enlarged or reduced in size. Faced with this pattern of results, most researchers hypothesize that subjects are recognizing the objects via a normalization procedure, i.e., that a representation of the stimulus image is gradually transformed into alignment with an internal model to determine identity[2].

However, as originally noted by Corballis (Corballis, 1988; Corballis, Zbrodoff, Shetzer, & Butler, 1978), such a description of the underlying processing is a ‘Catch 22’. In order to test for a match between a stimulus image and an internal model, this account proposes that we must first adjust the properties of size and orientation of the stimulus image (or the internal model) until the two representations are equivalent with respect to these properties. But how can we know the manner in which we need to transform the stimulus representation in order to bring it into alignment with the internal model without already knowing what the internal model is? On the one hand, if we do not know which internal model possesses the same shape as the stimulus image how can we possibly know which transformation will align the two? On the other hand, if we *do* know what the appropriate internal model

---

[2] Note that this normalization process can also occur in the opposite direction, i.e., such that the internal model is transformed in a way that brings it into alignment with the stimulus image. For the sake of simplicity, we will discuss only the case where the stimulus image is transformed because the arguments we will present in this section are equally relevant to both versions of normalization.

is before normalization, and can therefore calculate how to align the two representations, why do we need to transform the representation at all given that its identity has already been determined?

It is possible that the visual system does rely entirely on normalization to determine identity, i.e., that we do not 'know' beforehand which model can be brought into alignment with the stimulus shape. However, the computational costs of this reliance would be extreme. To identify an image, the visual system would be required to proceed in an arbitrary manner through some or all of the internal models in an attempt to find a match for the stimulus image. In some paradigms, the number of internal models that need to be considered can be limited by the experimental design. For example, in a successive presentation paradigm where a trial consists of a pair of images presented one after the other, subjects might be asked to respond whether or not the second image of the sequence is a rotated version of the first (e.g., Pierret & Peronnet, 1994). In such cases, subjects need only consider the internal model of the first image of the pair. However, in other paradigms, the number of internal models that would need to be considered, i.e., where a normalization procedure would need to be attempted, is not limited in this way. For example, if the task involves naming the stimulus shape (e.g., "dog") then there is no such limit to the number of internal models in memory that could provide a match. Therefore, with no way of limiting the search, we would need to proceed in an arbitrary fashion through the entire store of models attempting normalization on each model we consider until a match is found. In the worst-case scenario, the system would attempt normalization on all of the models in memory. Such a modus operandi would be computationally expensive and time-consuming. In fact, researchers such as Corballis (Corballis, 1988; Corballis, Zbrodoff, Shetzer, & Butler, 1978) have not even discussed this hypothesis.

Another thorny issue that must be addressed by any account of normalization is how the time to respond so closely reflects the shortest transformational path to match between stimulus image and model. One possibility is that this search uses a ‘brute force’ method whereby we perform simultaneous, unguided transformations on multiple copies of the representation. If the stimulus image and the internal model differ in terms of planar orientation and size, then four copies of the stimulus image must be created and concurrently manipulated in each of the following ways – (1) increasing its size and rotating clockwise, (2) increasing its size and rotating anti-clockwise, (3) decreasing its size and rotating clockwise and (4) decreasing its size and rotating anti-clockwise. These transformations occur at a uniform rate and the manipulations cease when one of them provides a match with an internal model. This method guarantees that the response times will always reflect the smallest necessary transformation to match. However, this solution is an inelegant one and is very computationally expensive. In fact, the description given above assumes unrealistically that we already know that the image is only transformed in planar orientation and size. In real life, this assumption is implausible. Therefore, one should also include other transformations from the similarity group, e.g., translations, which would result in a combinatorial explosion of possible transformations that need to be performed.

However, there are ways in which the transformations applied to the stimulus image can be limited or guided in some manner. For example, Ullman (1989, p. 211-

212) has suggested that the calculation of simple image attributes can sometimes be used to guide the transformations required. Calculating the center of mass of both the stimulus image and the internal model can guide any required translation transformation – the difference between the two values indicates the translation required to bring the two into alignment. If the stimulus image and internal model differ in terms of size then calculating the property known as the *convex hull* (defined by Preparata and Shamos (1985) as the smallest convex envelope that will enclose an object) of the two representations can similarly be used to calculate the scale operation required. However, compensating for orientation differences is far more complicated. Ullman (1989) argues that orientation can sometimes be determined by calculating bilateral symmetry. However the axis of bilateral symmetry is not necessarily aligned with the upright direction. And, even if it is aligned in this manner, deriving this axis will not determine where the top of the object is. In fact, none of these techniques is reliable when rotations occur outside of the picture plane because the global properties they rely on (e.g., center of mass) cannot be reliably calculated from a single perspective.

Research by Vickers (2002), Vickers and Preiss (personal communication) and Huttenlocher and Rucklidge (1992) has investigated the use of a measure known as the *Hausdorff distance* to guide the search for a matching transformation between identically shaped images. This measure is a generalization of the normal Euclidean distance between any two points to the distance between any two sets of points where each set defines a separate form (Peitgen, Jürgen, & Saupe, 1992, p. 167). Consider two forms described by the coordinates set A and set B respectively. For every point in A, there is a distance to each point in B and there is one least distance. Therefore, for all points in A there is a set of least distances to points in B. Likewise, for all the points in B there is a set of least distances to points in A. In these two sets of least distances (i.e., A to B and B to A), there is one greatest least distance. The Hausdorff distance is the maximum of these two greatest least distances. Identical images will of course have a Hausdorff distance of zero. The more any two similar shaped images are separated by transformations such as rotation and translation, the greater the Hausdorff distance will be. Given this characteristic, the measure can be used to guide a transformation to match procedure.

For example, Vickers and Preiss (personal communication) have used the Hausdorff distance to guide the transformations performed in representational momentum. To bring two consecutive images in a sequence into alignment, the first image is transformed in a myriad of different ways (e.g., simple transformations from the similarity group as well as combinations of these transformations) by only a small amount. After each different transformation, the new Hausdorff distance is calculated between this new image and the following image in the stimulus sequence. For the most part, the transformation that is of the same type and is in the same direction as the one which will bring the two images into complete alignment will produce the lowest Hausdorff distance. For example, if the two images are separated by a $90^0$ clockwise rotation, then a rotation in this direction, even if by only a few degrees, should produce the shortest Hausdorff distance out of all the other types of transformations considered (e.g., translations, scale changes and even anticlockwise rotations).

This procedure is repeated iteratively until the Hausdorff distance is zero (or a small arbitrary value) at which point there is a match between original and transformed image. The magnitude of the Hausdorff distance is a good predictor of the size of the transformation needed to match such that it can be used to gradually reduce the size of the transformations applied the closer the image is to a match. Although not all test transformations applied to the original image will occur along the shortest transformational path to match, the overall transformational path that is traversed should be close to the shortest between the two images.

## 3.2 Huttenlocher and Ullman's (1989) model

One way to circumvent the 'Catch 22' is to propose a preliminary analysis of the stimulus image that derives a limited amount of shape information prior to any normalization. This partial description or 'soft-match' can be likened to a perceptual hypothesis about the shape's identity. In order to test this hypothesis and to derive a 'hard-match', normalization may be required. This account eliminates the need for a potentially exhaustive search through all the internal models.

Corballis (Corballis, 1988; Corballis, Zbrodoff, Shetzer, & Butler, 1978) has proposed just such an account for the recognition of planar rotated images although it is easily extended to accommodate scale differences as well. He suggests that there are, at most, three stages in the recognition process of rotated images. Firstly, an orientation-free description is invoked that may or may not be 'strong' enough to recognize the object. If this description is 'strong' enough, the object is recognized and no normalization is required. However, if it is not 'strong' enough, the second stage proceeds whereby orientation information is extracted from the 'weak' representation. In the final stage, a process of analogue transformation of the image occurs whereby the representation is brought into its upright position in order to match an internally generated template of the internal model. This theory, also known as the *double-checking* account, avoids the logical paradox behind conventional theories of normalization - by making a first guess based on orientation-free information, it explains how we appear to "know" what the identity of an image might be before any normalization is performed.

Corballis' (Corballis, 1988; Corballis, Zbrodoff, Shetzer, & Butler, 1978) theory is presented in relatively broad terms with no mention of how these ideas might be implemented in a working model. For example, there is no mention of what features might be initially extracted from the stimulus image or how the normalization procedure should be carried out. However, there is another model which is very much in the spirit of Corballis' (Corballis, 1988; Corballis, Zbrodoff, Shetzer, & Butler, 1978) ideas and that *is* spelt out at an implementational level. Jolicoeur (1992) has pointed out important similarities between the alignment system developed by Huttenlocher and Ullman (Huttenlocher & Ullman, 1987; Ullman, 1989) and Corballis' (Corballis, 1988; Corballis, Zbrodoff, Shetzer, & Butler, 1978) theory. Huttenlocher and Ullman's (Huttenlocher & Ullman, 1987; Ullman, 1989) model provides a means of recognizing objects using an initial analysis of shape information and a subsequent image transformation process.

The first stage of their model involves the derivation of a 'key' from the stimulus image that is composed of three simple geometric features and the spatial relationships between them. These features can be arbitrary (but non-collinear) points, known as *anchor points*, on the object's surface. Points of maximum convexity or concavity along the contour and the center of the figure are ideal candidates for such anchor points. An example of anchor points in an image is given in figure 3.1. Alternatively, one can also use combinations of three co-planar points and lines as the 'key' (Shoham & Ullman, 1988).

The representation of objects in long-term memory also consists of a particular 'key' as well as a link between this key and a more complete object description. Just like the weak orientation-free interpretation proposed by Corballis (Corballis, 1988; Corballis et al., 1978), the key derived from the stimulus image is used to pick out a key, or a number of keys, stored in long-term memory. Labels assigned to the points in a key are used to determine possible matches between stimulus and model keys. The labels include information on how they were derived – e.g., deep concavities, cusps, tips, sharp curvature maxima, center of a closed or almost closed blob and inflection points – and may also include a rough description of location. The link between the key of a stimulus and a matching key of a model stored in long-term memory is comparable to the shape hypothesis formed by the weak orientation free description described by Corballis' (Corballis, 1988; Corballis et al., 1978) theory. Importantly, the geometrical features that define the points in a key are independent of the orientation and scale of the image. In other words, the preliminary analysis of the image is based only on properties that are invariant to the orientation and size of the stimulus image.

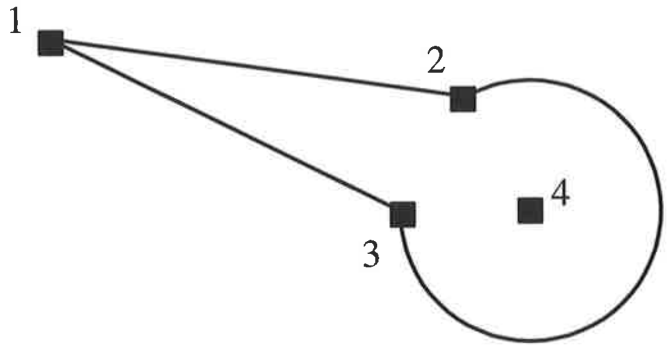


*Figure 3.1. An example of an image with anchor points determined by simple geometric properties. In this case labels 1, 2 and 3 are inflection points while label 4 is the center of an almost closed blob.*

Having isolated one or more potential matches in the store of models in long-term memory, we then proceed to the second stage of Huttenlocher and Ullman's (Huttenlocher & Ullman, 1987; Ullman, 1989) algorithm that involves image transformation. The first step involves calculating what transformation, if any, will bring the key of the stimulus and the key of the model into alignment. Mathematically, Huttenlocher and Ullman (1987) and Ullman (1989) have demonstrated that a unique transformation can be calculated using just the anchor points on the image and those on the internal model and that this transformation is the shortest possible. If no transformation can be calculated, then the model cannot be a match to the stimulus image and is therefore rejected. However, if the anchor points can be aligned, then the transformation that will align the two sets of keys is performed on a more complete description of the stimulus image to test for correspondence with the more complete object representation in long-term memory.

To test for a match between transformed stimulus image and model is then a relatively simple affair. For example, Ullman (1989) has successfully used a simple distance metric that is similar to a measure known as the Hamming distance. This measure requires that the images are coded in a binary fashion such that each point in the image is a point in a grid that is "on" and all other points in the grid are "off". The Hamming distance is the number of points in the grid that differ between the model and the image, i.e., the number of points that need to be switched "on" or "off" in order for the two to be identical. Therefore, if the images are identical the Hamming distance is zero.

Huttenlocher and Ullman's (Huttenlocher & Ullman, 1987; Ullman, 1989) model bears some similarity to Lowe's SCERPO (*Spatial Correspondence, Evidential Reasoning, and Perceptual Organization*) alignment system for object recognition (Lowe, 1985, 1986). Both theories calculate the transformation required to bring a stimulus into alignment with an internal model by using the method of three non-collinear points. Both theories also propose an initial analysis of the stimulus image that uses only properties that are invariant to the perspective from which the object is viewed. However, this initial analysis of the image in the SCERPO model is more complex than that used in Huttenlocher and Ullman's (Huttenlocher & Ullman, 1987; Ullman, 1989) model. This involves perceptual grouping variables such as collinearity, parallelism, and endpoint proximity or connectivity. These variables are then used to form higher level groupings which are ranked in order of significance for the purposes of finding a model in memory that matches the stimulus image.

Huttenlocher and Ullman's (Huttenlocher & Ullman, 1987; Ullman, 1989) model solves all the problems facing computational theories of normalization mentioned previously. It gives normalization a proper computational role in the recognition process. It also explains how we appear to know what the stimulus represents before normalization by postulating an initial analysis based on invariant features of the stimulus image. Cleverly, the key derived from this analysis is also used to calculate the shortest transformational path to match between stimulus and internal model. Huttenlocher and Ullman's (Huttenlocher & Ullman, 1987; Ullman, 1989) model not only solves the 'Catch 22' but it also provides a way of restricting the search for a matching model and calculating the required transformation for normalization. In other words, the system does not need to resort to arbitrary or random methods in either looking for the right model or seeking the right transformation needed for alignment.

## 3.3 Possible refinements

However, to more accurately model the performance of humans, Huttenlocher and Ullman's (Huttenlocher & Ullman, 1987; Ullman, 1989) account of normalization needs further refinement. This algorithm can explain why the perceptual system normalizes the image. This account also explains how the system knows which transformation is appropriate. But what is also needed is an explanation for why the degree of transformation is reflected in error rates and response latencies in the first place. As Edelman and Weinshall (1998) point out, there is no reason, in computational terms, why the transformation could not be achieved in a single chronometric step. If this were the case, there would be no latency variability or error relating to the degree of transformation required.

One way of dealing with this problem is to posit a limited capacity normalization device that can only carry out transformations by a fixed, discrete amount (Edelman & Weinshall, 1998). For transformations larger than this amount, an iterative process is required which chains together a succession of these steps by feedback. Each of these individual steps takes a fixed amount of time and the number of steps (and therefore the total time) required to perform a transformation will be linearly related to the size of the transformation.

Similarly, we can introduce a small amount of noise in each step in order to emulate psychological data with respect to accuracy. Although the overwhelming majority of studies taken as supportive of normalization cite evidence in response times only, several researchers have noted that the number of errors is sometimes linearly related to the size of the transformation required. Rock (1973) was the first to propose that the normalization process for planar rotation requires a taxing element proportional to the angle of rotation. Kosslyn (1980) has also proposed that, in mental imagery, error is relative to the degree of rotation away from the target orientation. Therefore, an additional refinement of Huttenlocher and Ullman's (Huttenlocher & Ullman, 1987; Ullman, 1989) mechanism might be the addition of a small amount of noise at each step. Hypothetically speaking, the amount of noise that is added at each step could also be proportional to the complexity of the stimulus being transformed. This would explain the relatively severe effects of planar disorientation on error rates in the recognition of complex shapes like faces. Empirical research has shown that even very familiar faces can become unrecognizable when inverted (Rock, 1974; Thompson, 1980).

## 3.4 Limitations

However, such an alignment procedure is not without its difficulties. For one thing, the number of possible alignment keys available in real world objects would be extremely large. As Palmer (1999) has pointed out, natural objects "do not come conveniently marked with black and white dots" representing the alignment key (p. 368). Elements such as blob centres, points of maximum convexity or concavity may provide clues to alignment features. However, there may be more than three of these elements in one figure such that analyzing these simple geometric features may not be adequate to pick out a unique key or even a small number of keys. The economy and the psychological plausibility of Huttenlocher and Ullman's (Huttenlocher & Ullman, 1987; Ullman, 1989) procedure comes from the fact that the number of potentially matching models that need to be transformed is drastically constrained by the search for matching keys. In addition, when image normalization does occur the keys are used to calculate the transformation required. However, this account has shifted much of the computational burden onto the preliminary stage of key extraction and search. In a limited, constrained environment of objects this may be workable. However, for real-world perception this may not adequately constrain the search for potential models.

# CHAPTER 4: Invariant theories of object recognition

*This chapter discusses different models for recognizing objects that do not involve any normalization - i.e., the object representations are size and orientation free. It is divided into two main sections. The first focuses on two popular theories of object recognition - David Marr's theory (Marr, 1982; Marr & Nishihara, 1978) and the Recognition-by-components or geon theory (Biederman, 1985, 1987; Cooper, Biederman & Hummel, 1992). The second section presents a Connectionist model for size and planar orientation invariant visual processing that incorporates ideas from Schwartz (1980a, 1980b) and Uttal (1975) and is inspired by both neurophysiological and psychological research. Although this model is far less powerful and sophisticated than the conventional theories outlined in the first part of the chapter, it offers a radically different perspective on visual processing and may complement more conventional mechanisms. In addition, we argue that a more complete model of human object recognition may contain components from a number of different theories.*

## 4.1 Introduction

In the previous chapter, we examined recognition mechanisms that rely on an adjustment of the size and rotation of the stimulus image. In this chapter, we look at the mechanisms that do not require normalization. According to these accounts, the object representation remains unchanged despite alterations in the orientation and size of the stimulus image. We will refer to such theories as *invariant* theories of object recognition. In this chapter, we will describe and compare two influential theories of this type in the literature. In addition, we will present a simple Connectionist model for size and rotation invariant object recognition that provides a novel solution to the problem at hand. At present, there is only a small number of studies that have sought to answer the difficult question as to what invariant mechanism may be used as opposed to the more common question of whether any invariant mechanism is used at all. Therefore, there is little empirical evidence that allows us to discriminate between the competing theories. Our approach will not be to make any firm conclusions regarding the validity of these models but to discuss the advantages and disadvantages associated with each and to suggest ways in which aspects of the various models can be integrated into a more complete model of invariant object recognition.

## 4.2 Marr's theory of object recognition

One of the most influential and comprehensive theories of visual perception is that of David Marr (Marr, 1982; Marr & Nishihara, 1978). This account outlines a cohesive flow of processing from lower level perceptual processing right up to higher level object recognition which ultimately involves object-centred shape representations (i.e., free from global orientation and size information). An early representational form is known as the 2 ½ - D sketch. This consists of a representation of the visible object surfaces via a large number of locally flat pieces. Each of these fragments outlines the local surface orientation, distance from the viewer and colour at a particular point in the visual scene and this representation is viewer-centred. However, the next stage consists of a three-dimensional representation that is object-centred and contains information about unseen surfaces. This consists of one or more primitives known as generalized cylinders or cones - an idea that was first introduced by Binford (1971) in the context of computer vision. These primitives consist of the surface created by moving a cross-section along a particular axis and an example of a generalized cone is given in figure 4.1.

The next stage consists of an axis-based description of the object. These axes are defined by properties such as elongation, symmetry and motion (i.e., direction of movement) and include the axes of the generalized cones which make up the object. One of these axes is the *model axis*. In most cases, this model axis also serves as the *principal axis* which defines the local coordinate system. This means that the arrangements of the other axes, called the *component axes*, are defined relative to the principal axis (by a mathematical description Marr calls *adjunct relations*). This representational system is therefore object-centred and not viewer-centred. To simplify this description, we will consider the most common case where the model axis *is* the principal axis.

The process of recognition is an interactive process between a description of the visual image and stored 3-D models. A more general analysis of the visual image guides the choice of possible models. In turn, these models guide further analysis of the visual image. Therefore, while the shape representation can often be derived in a purely bottom-up fashion, there is also scope for top-down involvement. Because the object representations are size and orientation free, there is no normalization. However, later in the recognition process when a homology between a 3-D model and the image has been established, Marr suggests that there may be some kind of procedure that is described as a "relaxation process that adjusts the orientation of the [principal] axis incrementally" (Marr & Nishihara, 1978, p.290). This relaxation

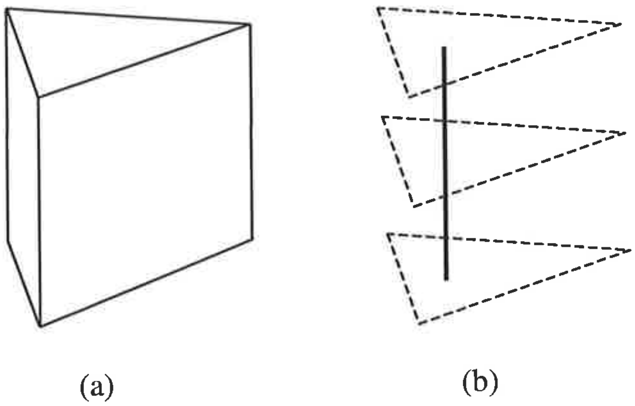


*Figure 4.1. An example of a generalized cone.* *The three-dimensional surface of the object depicted in (a) can be created by the process outlined in (b). In this procedure, a triangular cross-section (depicted at several stages along the trajectory by a dotted line) is moved smoothly along an axis (depicted by the solid line).*

process is guided by constraints supplied by the potential model. However, such a procedure is only one of several mechanisms suggested by Marr and does not appear to be necessary for objects seen in planar rotations only. Another possible explanation for orientation effects in the recognition of planar rotated images in the context of Marr's theory has also been discounted. Jolicoeur (1990) has suggested that such effects may be the result of the derivation of the axis being faster for upright objects. However, McMullen and Jolicoeur (1989) have demonstrated that prior information on axis orientation does not attenuate orientation effects, suggesting that the axis finding process is not responsible for the orientation effects.

There are several reasons why Marr's theory is attractive. It attempts to deal not just with planar rotations but also with rotations in depth of three-dimensional objects. Importantly, the theory can also account for axis foreshortening effects in psychology and neuropsychology because in such cases the information about the principal axis, which is primary for shape recognition, or another major axis, is limited or not available. Mathematically, Marr (1977) has demonstrated that it is difficult to derive information on a foreshortened axis from the occluding contours of an image. In studies on human performance there is also evidence that recognition of objects in such views is hindered and may require the use of an entirely different recognition strategy. For example, there is a disadvantage in naming (Humphrey & Jolicoeur, 1993) and matching objects seen from such perspectives (Lawson & Humphreys, 1996). In addition, several neuropsychological studies have shown that subjects with right hemisphere lesions often have a specific difficulty in matching images shown in foreshortened views that is not present when they are shown in prototypical views (Humphreys & Riddoch, 1984; Warrington and Taylor, 1973, 1978). Therefore, the limitation in dealing with objects shown in foreshortened views demonstrated in Marr's theory accurately reflects difficulties in human perception.

Another appealing aspect of Marr's theory is its sheer scope. It entails a wide variety of perceptual phenomena from lower level perceptual processes right up to our long-term memory of shapes. Most importantly, it allows for feedback between different stages of processing and the qualitatively different representations of shape they produce. Information from different representational levels can guide subsequent analyses at higher and lower levels giving the overall model great flexibility and computational power. The breadth and depth of Marr's account has had a profound effect on our conception of human visual processes and many later theories owe much to his pioneering work.

## 4.3 Biederman's (1987) Recognition-by-components theory

One such theory of shape recognition which also incorporates object-centred descriptions and volumetric primitives is the *Recognition-by-components (RBC) theory* of Biederman (1987). Like Marr's theory, this proposes shape representations that involve generalized cylinders. However, in RBC these are assigned to categories by their values on certain parameters. These categories are known as *geons* and these are defined by a discrete range of one or more parameters. There are 36 qualitatively different geons as defined along 4 dimensions. Firstly, the cross-sectional curvature can be straight (e.g., a square or rectangular base) or curved (e.g., a circular base). Secondly, the cross-section can be asymmetrical or be symmetrical in one of two ways: either reflectionally symmetrical or both rotationally and reflectionally symmetrical. Thirdly, the size of the cross-section can either be constant as it is moved along the axis, expanding or both expanding and contracting. The fourth qualitative parameter is the curvature of the axis which can be either straight (e.g., a pencil) or curved (e.g., an arch).

The 36 qualitatively different geons defined along these dimensions can be further subdivided in accordance with a quantitative dimension known as the aspect ratio. This is the ratio of the length of the axis that is swept to that of the largest dimension of the cross-sectional area and can take three values: approximately equal (e.g., a square box), axis greater (e.g., a tower) or cross-section greater (e.g., a piece of paper). Together with the 36 qualitatively different geons, the inclusion of the aspect ratio increases the total number of possible geons to 108.

Rather than an axis based description as in Marr's account, the objects are encoded in terms of the geons present and the spatial relations between them. There are a total of 108 qualitatively different relations between any two geons. There are three different elements to a spatial relation including relative size (e.g., 'is larger than'), relative position (e.g., 'on top of') and relative orientation (e.g., 'perpendicular to').

One particular advantage of RBC over Marr's theory is that it is custom built to handle part structure. In several experiments, Biederman and Cooper (1991b) have demonstrated some evidence that object recognition can be conducted, not on the lines and vertices of the image, but on the geons with which they are associated. In their first experiment, priming was equivalent for images displaying the same object, when different lines and vertices were visible, as long as there was still sufficient structural information about each component geon in the image. In other words, while the image was entirely different between the two presentations, there was presumably sufficient information to activate the same geon representations and recognition was successful. In the second experiment, the number of parts of an object that were shown was the same. However, for some objects the parts that were deleted constituted entire geons. In these cases, priming was significantly reduced presumably because the same geon representations could not be formed and therefore identical object representations could not be accessed.

RBC and Marr's theory have much in common beside object-centred representations and volumetric primitives. For example, RBC can also account for axis foreshortening effects noted previously in human performance. Rather than the result of a failure to derive an important axis as postulated by Marr's theory, these effects in RBC are the result of difficulties in deriving geons due to occlusion. The geons are derived from nonaccidental properties and not rare 'accidents' of viewpoint, as happens when an axis is foreshortened. However, while the exact source of these effects is different from that in Marr's theory, the resulting system performance is similar. Another similarity with Marr's theory is the use of top-down processing. Although mostly bottom-up, there is scope for feedback from higher representational levels when the available stimulus information is restricted, e.g., when images are noisy, obscured or degraded. In such cases, there can be feedback from geons to the features and / or feedback from categories to geons. Therefore, both theories are computationally flexible and can recognize images under restricted viewing conditions.

The two theories are also similar in their inability to account for normalization effects in the recognition of planar rotated objects. As Palmer (1999) has pointed out, the geon relations in RBC are not orientation invariant but are defined with respect to a gravitational standard (e.g., 'above' and 'below'). For some objects (for example a lamp with a base and shade), the relationship between the parts is simplest when the object is upright – i.e., the spatial relationship between the two geons can be defined as the 'shade' geon being *above* the 'base' geon. As such, when this object is presented at a non-upright view, the spatial relation will be more complicated – for example, at an orientation of $45^0$ the relation is *above and beside*. This will decrease the match between the model and image. However, this does not explain why objects appear to be rotated to the upright during recognition. Because the matching is orientation-invariant there is no need for an additional image transformation. Thus, as with Marr's theory, RBC cannot explain planar rotation effects in recognition performance.

However, there are some effects of size sensitivity that these models can account for. Empirical studies have revealed effects of size that are attributed to low level representation and not shape matching. We will discuss this phenomenon, known as the *zoom lens theory*, in detail in section 7.14. Both theories incorporate principles of early perceptual processing and could account for such effects.

In summary, both the accounts of Marr and Biederman are detailed explanations of visual object recognition whose fundamental prediction is that object representations are object-centred. This means that they do not involve any normalization for size or orientation. In later chapters, we will present empirical evidence from previous literature (chapters 5 and 7) and from our own experiments (chapters 6 and 9) demonstrating that human object recognition can be invariant to size and orientation. In the following section, we describe a simple neural network model that provides size and orientation invariant recognition in a different manner. Although not as thorough or powerful as the two approaches based on generalized cylinders we have described above, it demonstrates a qualitatively different approach to invariant recognition that reflects principles of neural mapping in the human visual system.

## 4.4 A Connectionist model of object recognition: Computational anatomy and autocorrelation

In the following sections, we will outline a model of object recognition that, like the theories of Marr and Biederman, proposes that identification is achieved in a manner that is invariant to a number of shape preserving transformations in the stimulus image. This account is inspired by neurophysiological data regarding the mapping of stimulus images in the visual system as well as psychological findings on the detection of forms in the visual environment. While the model is highly simplified and preliminary, it offers some insight into a radically different approach to the problem of object identification.

### 4.4.1 Connectionism

In this section, we will give a brief outline of the concept of Connectionism, which provides both an architecture for the proposed model via the use of neural networks but also a philosophy with respect to visuo-cognitive processing. Connectionist models 'mimic' certain functional and structural properties of the brain by employing a large number of simple, interconnected processing units operating in a parallel and interactive fashion (Rumelhart & McClelland, 1986; McClelland & Rumelhart, 1986). These processing units simulate, to varying degrees, the properties and activities of neurons in the brain. Connectionism represents a major break from the classical models of cognition. It shifts the focus of attention onto the ‘physical stuff’ responsible for our mental processing. Connectionism holds that our cognitive processes are the result of parallel distributed processes implemented by the ‘wetware’ in our heads, the human brain. These parallel distributed processes are analogue - that is, the causal laws associated with the material substrate govern the computational procedures.

For Connectionism, the physical locus of attention is the neurons - the billions of specialized nerve cells in the central nervous system which are extremely interconnected in networks and communicate with each other via electrochemical signals sent along axons, across synapses and onto the dendrites of the other neurons. As a result “neurons do not transmit large amounts of symbolic information” (Churchland, 1986, p.461). Accordingly, the units or 'artificial neurons' in these Connectionist models only send numbers to other units. Because all of the units, besides sending and receiving numbers or activation values from a large number of other units, also process simultaneously, Connectionist systems are also known as PDP (Parallel Distributed Processing) networks. Processing occurs, not through the explicit and serial application of explicitly stored bits of information, but rather by the combined and interactive effect of a large number of simple processing units acting in a parallel fashion, simultaneously applying all the ‘implicitly’ stored knowledge in the system.

Thus the Connectionist approach assumes a theory of cognition at the very basic level of conception - it is a very ‘bottom-up’ approach as opposed to the often very ‘top-down’ or ‘intellectualistic’ models that conform to the classical conception of cognition (Dennett, 1978). Despite this emphasis, Connectionist models often exhibit the very behaviours and tendencies, that are explicitly designed into ‘top-down’ theories, but in a far more flexible and lifelike manner. These ‘emergent properties' of the network occur through the interactions of units at a lower level in a purely local fashion in the same way that the properties of a wheel emerge through the interactions of the constituent molecules but where none of these molecules can be said to possess the property of 'wheelness'. There is no global 'executive' responsible for the progressive functioning of the individual units - each unit is simply processing on the basis of its interconnectivity pattern in the network (Horgan & Tienson, 1989). In addition to emergent properties, parallel distributed processing systems exhibit a flexibility impossible in classical models of cognition. They are capable of extremely lifelike learning behaviours, can spontaneously generalize to novel stimuli and exhibit what is known as *graceful degradation*. This means that damage to the system, by the gradual removal of processing units or 'neurons', results in a smooth and gradual decline in the performance of the system so characteristic of the human brain (Rumelhart & McClelland, 1986).

The approach we are taking is a refinement of the Connectionist philosophy. This involves the notion of *model-based neural networks*, which has been defined by Caelli, Squire and Wild (1993) as having the aim of "including a priori knowledge and desired invariances of the network, rather than requiring a training set that is sufficiently varied to cause such invariances to arise spontaneously" (p.617). In short, this requires more structured and modular networks incorporating domain specific knowledge and a functional decomposition of the task by the researcher. This places an emphasis on psychological theories and data. This approach has already proved successful in a similar domain. In one study, Caelli et al. (1993) have successfully applied the *model-based neural network* technique to a very simple invariance task involving the perception of horizontal and vertical lines.

In addition, the approach we are taking has a further advantage, not spelt out under the notion of model-based neural networks, because we are drawing on specific neurophysiological data. As Churchland (1986) has pointed out, one of the important features of Connectionist modeling is that it can 'return the favour' and guide further research into neurophysiology.

### 4.4.2 A preliminary stage of visual processing : The striate cortex

The first stage of the Connectionist model is inspired by neurophysiological research into stimulus representation in the early visual system. Talbot and Marshall (1941) were the first researchers to suggest that the manner in which visual stimuli were 'mapped' onto the striate cortex might provide some kind of neuronal representation, or internal code, of the visual field. But, according to Schwartz (1980a), this idea proved so problematic that for a long time it was ignored by the majority of researchers and theorists. For there seemed "little reason for the nervous system to retain the spatial form of a stimulus beyond the level of the retina" (Schwartz, 1980a, p.653). What sort of purpose could this serve? If the visual environment was initially encoded by the retina, what functional utility could be gained by simply re-representing this pattern at a higher processing level?

In addition, the concept of some kind of 'cortical movie-screen' (as this phenomenon has been termed by Somjen, 1972) seems to imply the existence of an additional system to interpret this representation - if you continue the 'movie-screen' analogy, you need someone to watch the movie as it is being played. This concept of a 'neural observer' is not only unfounded from a neurophysiological and psychological point of view, but it is a problematic notion in the philosophy of cognitive science. The idea of a separate functional entity in the mind, affectionately titled a 'homunculus', which interprets the output of another system, results in a infinite regress - it presupposes the existence of an even smaller homunculus to interpret the actions of the first homunculus, and so on. As Haugeland (1985) has pointed out, this amounts to falling into the "old trap of explaining intelligence [or cognition] by presupposing it" (p.113). Even if one were to somehow allow for the notion of a homunculus, Hubel and Wiesel (1979) speculate that it would have difficulty deriving information from the cortical map which deviates so wildly from its environmental and retinal origins. They claim that the visual world is represented in such a distorted manner that a homunculus "trying to glean information from the cortical projection would be puzzled indeed" (p.132).

However, an alternative point of view is that the cortical projection is in fact the result of a sophisticated and functional transformation of the original visual input. Schwartz (1980a, 1980b) first coined the term *computational anatomy* which refers to "the possibility that the anatomical structure of the retinotopic mapping may simplify certain aspects of perceptual coding" (Schwartz, 1980a, p.645). In other words, the mapping procedure may play an important role in the formatting of visual data for further processing. His research into the neurophysiology of the visual system has lead him to develop a model which quantifies the mapping of visual stimuli from retina to the two-dimensional plane of the visual cortex (Schwartz, 1976, 1977a, 1977b, 1977c).

This mapping involves the application of a complex logarithmic algorithm applicable on both a local and global scale. 'Global' refers to the overall cortical retinotopic mapping whereas 'local' refers to the spatial structure of *hypercolumns*. Hypercolumns have been defined by Hubel and Wiesel (1974) as the elementary functional unit of cortical architecture. Each hypercolumn covers an entire sequence of orientation tuned cortical columns and one left- and one right-eye ocular dominance columns. In this way, the global topographic map of the vertical projection onto the cortex may be considered to be made up of 'points', each of which represents a separate hypercolumnar unit. Schwartz found a close correspondence of geometrical form between the local and global scales. The resulting map is *conformal* - this means that the direction and magnification of local angles are preserved. It also maintains the attribute of connectivity but metric properties (area and perimeter) and other features like parallelism and perpendicularity relations are not carried over into the complex log representation.

Another feature of the complex logarithmic mapping is that it provides a pseudo-invariant processing of visual stimuli and it is this property that is utilized in our Connectionist model. The differences in the symmetries of size and planar rotation of an object are reduced to translations of an identical form in the mapped image when these motions are performed around the origin (0,0) of the stimulus image. Consider the representation of the original stimulus in Cartesian coordinates $(x, y)$. The image can also be described under a polar coordinate system, $(r, \phi)$, where $r$ represents the distance from the origin and $\phi$ represents the angular distance from the initial ray (e.g., the positive $x$ axis in the original representation). An example of a polar coordinate system is given in figure 4.2. The complex log mapping (or log polar

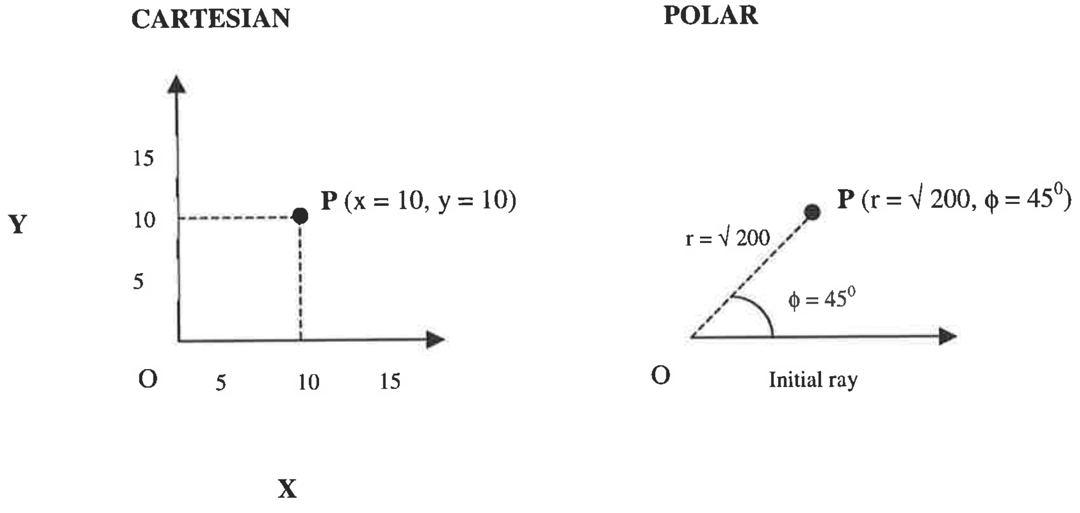


*Figure 4.2. A comparison of descriptions under Cartesian and polar coordinate systems.*

mapping as it is also known) is the locus (log ($r$), $\phi$) – a modification of the simple polar mapping such that the distance of a point from the origin is represented by a logarithmic function. Any change in size of the image is equivalent to multiplication of the radial coordinates by a constant, $k$. This produces a change in the complex log mapping of (log ($r$), $\phi$) → (log ($k$) + log ($r$), $\phi$). Therefore, the size dilation results in a linear shift by the amount log ($k$) in the complex log image, i.e., a horizontal translation as depicted in figure 4.3. Likewise, changes in planar rotation amount to linear shifts in the vertical direction by the amount of angular deviation and this property is also depicted in figure 4.3.

The precise function derived from the neurophysiological data by Schwartz (1980a) is $w = \log(z + a)$. The original formulation, where $a = 1$, is essentially the complex log mapping outside of the central $1^0$ of the visual field (Schwartz, 1983, p.833). Later research by Dow, Snyder, Vautin and Bauer (1981) provided foveal

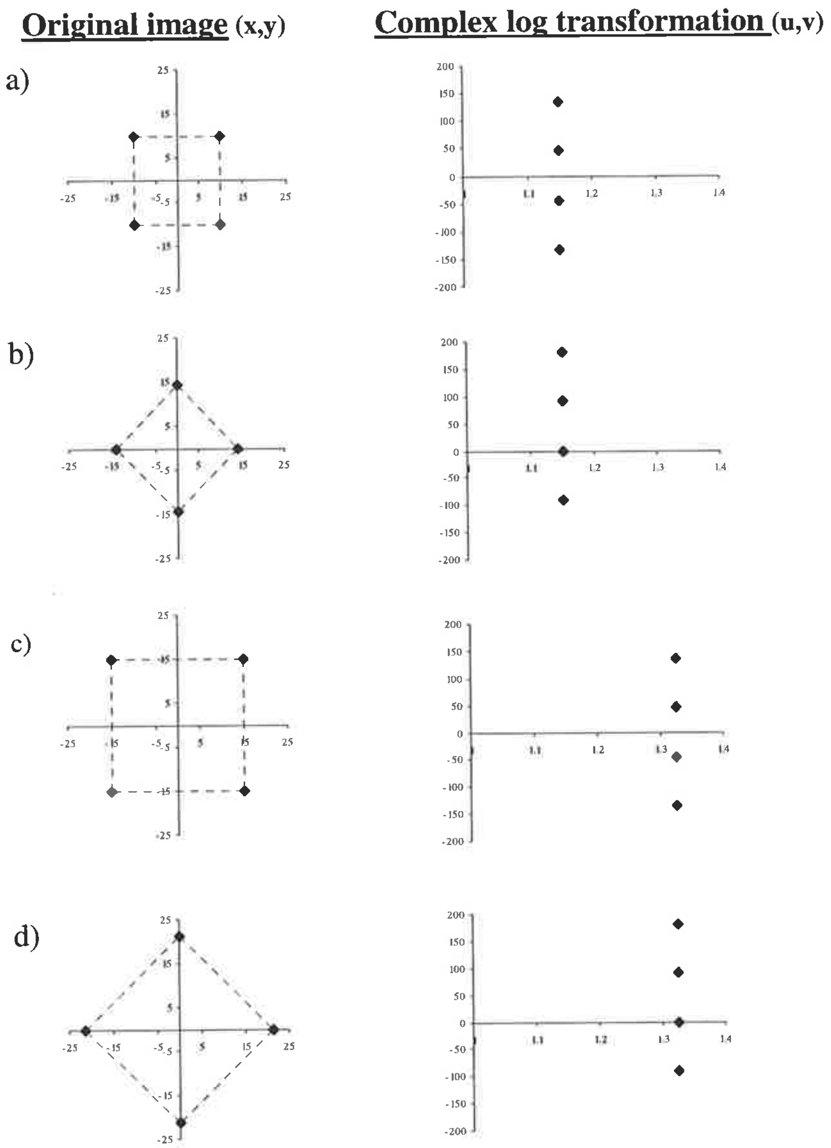


*Figure 4.3. An example of complex log mapping. The original image (left) and the image (right) following a complex log mapping. For the sake of simplicity, only the vertices are displayed. The transformed image is identical except for translation for b) planar rotations, c) scale changes and d) both planar rotations and scale changes when such transformations are performed around the centroid. Horizontal shifts represent changes in size while vertical shifts represent changes in planar orientation.*

magnification data demonstrating that the cortical topography is well approximated by the function log ($z$ + $0.30$). In other words, the complex log mapping is correct up to about $0.3^0$. For values of $z$ smaller than $a$, the map is linear and the magnification factor constant. This means that the geometric size scaling properties of the mapping hold to approximately 0.3 deg.

There are several other attractive aspects to Schwartz's account. The theory provides an account for feature extraction (i.e., the segmentation of colour and depth) according to which different 'maps' can be compared by means of the interconnecting columnar system. In addition, the author has demonstrated how the theory can account for particular visual illusions such as the Mackay complementary afterimage (Schwartz, 1980a, 1980b). However, it is the partial invariance to rotation and size symmetries of the complex log mapping that is central to our Connectionist model. As Schwartz has pointed out:

> *"it is certainly possible that a subsequent stage of shift invariant processing may provide a basis for geometrical constancies in vision." (Schwartz, 1980a, p.655)*

### 4.4.3 Other applications of computational anatomy and complex log mapping

Schwartz (1981b) has also demonstrated how complex log mapping and the notion of computational anatomy can be extended to the general problem of object recognition. This involves object recognition that is invariant not only to planar rotations and size changes but also to the more complicated problem of in-depth rotations. Schwartz's (1981b) model relies on a property that demonstrates projective invariance, i.e., a property of the two-dimensional projection of an object (such as the image on the retina) that remains unchanged despite in-depth rotations of the object.

The invariant Schwartz (1981b) used is known as the complex cross ratio (Ahlfors, 1966) and can be derived from the vector difference of successive cortical maps of a stimulus, i.e., different 'snapshots' of an object with different fixation points. In this manner, information derived from the comparison of different complex log mappings of the same object from different eye fixations can be used to derive the unchanging element of objective shape regardless of any shape preserving transformations applied to the stimulus image. In fact, the use of the complex cross ratio has two advantages over the conventional cross ratio which has been examined by Julesz (1971) and Johansson, von Hofsten and Jansson (1980). Firstly, it does not require that the three points from which the cross ratio is derived be collinear – a constraint which Schwartz (1981b) describes as "unreasonable and artificial" (p. 435). Secondly, in contrast to Schwartz's (1981b) own algorithm, none of the previous discussions on the possible use of the cross ratio in perceptual processing have provided a computational mechanism or a physiological basis for the use of this property in the visual system.

Several models have employed complex logarithmic mapping together with an initial spatial Fourier mapping to produce invariance to at least some of the properties of size, planar rotation and translation in pattern recognition (Brousil & Smith, 1967; Casasent and Psaltis, 1976; Cavanagh, 1978) However, Schwartz (1981a) has criticized such schemes for two reasons. Firstly, there are an infinite number of different stimuli (i.e., with objectively different shapes) that possess identical representations under such a model. Secondly, there is no area of the visual system, let alone one prior to the striate map, that "possesses anatomical or physiological properties which are consistent with the hypothesis of global Fourier mapping" (Schwartz, 1981a, p.456). Finally, Reitboeck and Altmann (1984) have also presented a model which employs complex log mapping. In contrast to our model, post-processing of the 'cortical mapping' is achieved, not by autocorrelation, but by the *R*-transform which is also translation invariant.

### 4.4.4 Post-processing of the cortical representation: Uttal's (1975) autocorrelation theory of form detection

The technique to be used for post-processing of the logarithmic transform in our model will be based on Uttal's (1975) autocorrelation theory of form detection. There are several attractive aspects to utilizing his algorithm in this context. Not only has this theory been used successfully to model human perceptual performance but it also lends itself to Connectionist modeling and, in many ways, is complementary to the work of Schwartz. In this section, we will outline some of these features of Uttal's theory and describe how the two mechanisms can be integrated.

### 4.4.5 Schwartz and Uttal: Complementary theories

In many ways the ideas of Schwartz and Uttal are complementary. Firstly, Schwartz (1980a) provided the inspiration for our model by suggesting that Uttal's technique, combined with his own complex logarithmic mapping, might provide the basis for recognition insensitive to various symmetries. Similarly, Uttal (1975) proposed that some kind of logarithmic pre-processing of stimuli might be a beneficial modification to his autocorrelation theory. He cites personal communication with Oyama who suggested that "spacing metrics based on logarithmic steps or even on constant cortical spacing might be much more effective in predicting the psychophysical data" (p.134).

Secondly, both authors explicitly or implicitly make reference to a hierarchical theory of visual information processing. In Uttal's (1975) classification, Schwartz's (1980a) complex logarithmic mapping theory fits nicely into what he calls *Stage 2* of perceptual processing. At this point, processing depends on geometrical relations at a local level and appears to involve only "interaction effects between neighbouring portions of the retina" (Uttal, 1975, p.6). Uttal's (1975) own autocorrelational mechanism occurs at the subsequent stage (*Stage 3*) and involves the detection of form, based on more global features of geometrical organization and the extraction of such patterns from background noise. Our Connectionist model is therefore consistent with Uttal's general scheme for visual processing.

### 4.4.6 Uttal's Research and Theory

Uttal's experimental focus in human visual form detection has involved the investigation of global factors of stimuli. He might be labeled a 'Neo-Gestaltist' with his dogma that "it is the parts of a stimulus whatever they may be, [which] are of less importance than the arrangement of those parts" (Uttal, 1975, p.3). His use of the *dot paradigm* in psychophysical experiments attests to this stance - the stimulus forms consist entirely of identical dots ordered in accordance with various global considerations. Uttal quantified the detection of these forms in experimental settings with stimuli consisting of both form and background noise (made up of the same dots as the target form but in a quasi-random configuration). With the application of the mathematical technique known as autocorrelation to the stimuli, Uttal was able to account for and predict a large amount of his experimental data.[3] Table 4.1 presents a brief summary of Uttal's (1975) psychophysical data.

Uttal's (1975) model accounts for all of these correlations although there is some discrepancy between the theory's predictions and the experimental results involving the degree of organization of line segments into forms. In such cases, there are differences in the Figure of Merit values that are not reflected in any change in detectability of the stimuli in the experiments. But Uttal (1975) claims that this is only a second-order discrepancy because the major orderings are reproduced by the model. In addition, he suggests that the deviations from psychophysical data may be due, not to a fundamental deficiency in the model itself, but to either an under- or oversensitivity of the model to various aspects of the stimuli or noise in the experimental data.

In summary, Uttal's (1975) theory provides a successful account of certain basic visual phenomena. In the Connectionist model we will present, the aim will be to extend this theory beyond its original application and formulation. Of particular importance among the many features of autocorrelation is its insensitivity to translation and rotation. Uttal (1975) himself claims that:

---

[3] Uttal devised a special measure, called the Figure of Merit (FoM), which was derived from the autocorrelational output and produced a single value for each stimulus to predict psychophysical data. But he conceded that it was an experimenter's tool and that "in the parallel processing environment of the brain this later operation may be unnecessary ... [because] one global state of the manifold of neural elements may be directly compared against another" (Uttal, 1975, p.100).

*Table 4.1*

***Summary of Uttal's (1975) results***

| **Experimental Variables** | **Relationship to Detectability** |
|---|---|
| Number of dots | Proportional |
| Interdot spacing | Inversely proportional |
| Deformed lines (deviations from linearity) | Inversely proportional |
| Regularity of dot spacing | Proportional |
| Polygon<br>- deleting dots from corners<br>- deleting dots from sides | <br>Proportional (marginal)<br>Proportional (greater) |
| Degree of organization of line segments into forms | Proportional |
| Corner Displacement of Polygons | Inversely proportional |
| Figural Goodness *(using Garner and Clement's (1963) measure)* | No effect |
| Translation and Rotation *(includes straight lines and the prototypical polygons - the triangle and the square)* | No effect |

> *"it is well known that in order to mimic human performance, an artificial pattern recognition system must possess a minimal sensitivity to translation and rotation" (p.124)*

Similarly, he also claims that "form recognition, in general, is not strongly influenced by ... magnification" and that any account must explain "how an observer can recognize objects that he [or she] has never seen before or has never seen before in a particular orientation" (Uttal, 1975, p.91-97). Our Connectionist model attempts to account for these phenomena by integrating the ideas of both Schwartz (1980a) and Uttal (1975).

### 4.4.7 Autocorrelation

The essence of Uttal's theory is the mathematical technique called *autocorrelation*. In this section, we will outline the formulation of the autocorrelation transform following a brief introduction to correlational functions in general. The generic correlational formula for two-dimensional space is:

$$\varnothing\ (x,y) = \int\int f(x,y) \cdot g\ (x,y)\ dx\ dy$$

where $\varnothing\ (x,y)$ is the general correlational transform of the function $f\ (x,y)$ with another function $g\ (x,y)$. In the present case, involving some kind of transformation of stimuli from the visual field, $f\ (x,y)$ may be considered the original stimulus pattern while $g\ (x,y)$ defines the particular type of correlation process. Other types of correlational transformations include cross-correlation and Fourier Analysis.

Autocorrelation is achieved when $g\ (x,y)$ defines a spatially shifted version of the input stimuli, i.e., when the input function $f\ (x,y)$ correlates with another function of the form $f\ (x + n\ \ \Delta x,\ y + m\ \ \Delta y)$ where $\Delta x$ and $\Delta y$ define the two-dimensional shifting operation and $n$ and $m$ are integers. Thus, the autocorrelation function can be defined as:

$$A_c = \int\int f(x,y) \cdot f(x + \Delta x,\ y + \Delta y)\ dx\ dy$$

The entire process can be visualized as a manipulation of the original pattern by comparing it with copies of itself that have been translated and then integrating all these various comparisons into a single representation. The results from the autocorrelation function can be plotted in $\Delta x$, $\Delta y$ space to form what is known as an *autocorrelogram*. This autocorrelogram, which is simply a graphical representation of the mathematical function, enhances the periodic or regular aspects of the stimuli. One outcome of this is that the autocorrelational technique is ideally suited to the extraction of a target signal (organized stimuli, e.g., a line of dots) from a background of noise (disorganized stimuli, e.g., quasi-random dots).

### 4.4.8 Autocorrelation: The transparency analogy

The *transparency analogy* is an intuitive way of understanding the essence of the autocorrelation procedure. Imagine that you have an image consisting of dots printed on an overhead transparency. Make an identical copy of this transparency and then superimpose this copy on the original. Start by lining up the two images perfectly (so that in fact all you can perceive is the original image) and then move the top sheet around by translations (UP, DOWN, LEFT, RIGHT and any combination of these directions) with respect to the bottom sheet. Do not perform any respective rotations or turn the sheet over. Notice how at different positions there are varying degrees of correspondence between the two images – i.e., the number of overlapping points changes (see figure 4.4).

Let us call the event where two points overlap a *correspondence*. If you add up the number of correspondences for each of these linear shifts and plot their summed

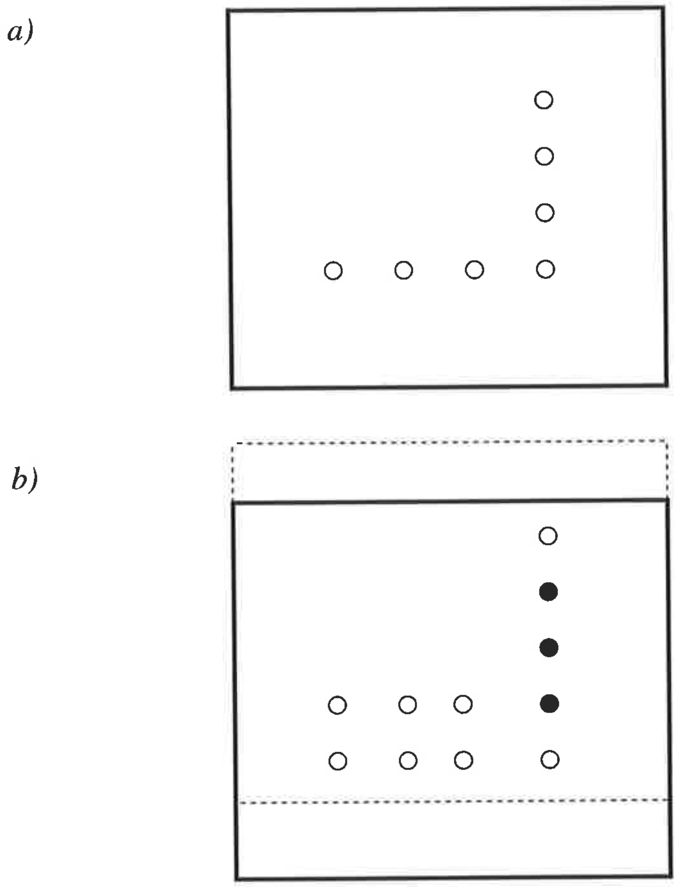


*Figure 4.4. The transparency analogy for autocorrelation. The first image, a), represents the original pattern of dots on a transparency while b) represents an identical copy superimposed on the original sheet and translated one unit "up" (i.e., shift (x,y) = (0,1)). In b) the symbol, ○, represents a single dot while the symbol, ●, displays the concurrence of two dots in the same Cartesian (x,y) position (overlapping points in the combined representation) otherwise known as a correspondence. Note that the edges of the original transparency are highlighted in bold while the dotted line signifies the border of the copied version.*

values at the positions on a new grid equivalent to the relative shift between the original and copied images, you would have started creating the autocorrelation matrix. The autocorrelation matrix is the two-dimensional grid, normally equivalent in size and geometric layout of the original grid, which displays the output of the autocorrelation procedure.

### 4.4.9 Other applications of autocorrelational techniques

Both its powerful mathematical properties and the way in which the function can be implemented in a manner consistent with neurophysiological principles have encouraged several psychological researchers to invoke the principles of autocorrelation. These include areas of human form recognition such as visual perception involving luminance at the retinal level (Engel, 1969), alphabetic character recognition (Engel, Dougherty and Jones, 1973) and even dot-mixing experiments similar to Uttal's (1975) work (i.e., Dodwell, 1971). Autocorrelation has also been used in the field of aural perception (Licklider, 1959; Cherry, 1961) and at the initial perceptual stage of a generalized model of memory (Anderson, 1968). Finally, autocorrelation has also been used successfully in the detection of symmetry. Tyler (1996) has demonstrated that the detection of symmetry may be achieved by the human visual system with the use of an autocorrelation mechanism. While models of symmetry detection like that of Palmer's (1982) require the use of third order and higher matching mechanisms, the use of a generalized autocorrelation mechanism (e.g., Tyler and Miller, 1994) reduces the processing required to a single, parallel step.

There is a good degree of overlap between the concepts of autocorrelation and Connectionism. As mentioned above Anderson (1968) used autocorrelation in a model of memory. In later papers, Anderson (1970, 1972, 1973) argued for the kind of parallel-distributed processing network that forms the basis of Connectionism and Anderson is currently an influential researcher in the field of Connectionism (see Anderson (1995)). Uttal (1975) himself acknowledges the link between autocorrelation and Connectionism. On the one hand, he is highly critical of many neural network theories which fall into the 'black box' category - in other words, those based solely on input and output considerations where “no proof of the uniqueness of any given solution to the problem of internal construction is possible” (Uttal, 1975, p.20). On the other hand, he is supportive of neural network theories that specify, to some degree, the internal architecture or mechanisms, i.e., networks that are based on the model-based approach discussed above. Our own Connectionist model falls into this latter category. In fact, the autocorrelation algorithm naturally lends itself to such neural network modeling. Because it is inherently spatially distributed and requires only unspecialized neurons and not exotic ‘feature-detector’ units it demonstrates a “reasonable correspondence to the known physiological properties of very simple nerve nets” (Uttal, 1975, p.135).

### 4.4.10 The first version

As mentioned previously, the Connectionist model consists of two major stages – the first consists of the log-polar mapping procedure while the second performs autocorrelation. In this section, the word *connection* refers to a single link between two computational units in the network. A *layer* consists of a set of similar units organized into a two-dimensional grid. The word *projection* refers to a set of one or more connections occurring between two layers where these connections are fundamentally similar.

### 4.4.11 Simulation Details

All simulations were created using version 1.11 of the PDP++ software package – a neural network simulator implemented in C++ and written by Randall C. O'Reilly, Chadley K. Dawson, and James L. McClelland (1998). They were performed on an IBM compatible PC using a Pentium II 200Mhz processor. The program was extended by the use of scripts written in CSS (PDP++'s own C++ language interpreter and script language) designed by the author and written by Mark Brown and the author.

### 4.4.12 Stage 1: Complex log mapping

The first layer, known as the *input layer*, of the network consists of a two-dimensional, 25 by 25 grid of 625 units, which represents the stimulus grid. It holds the original image of the object as it might project onto the retina. The stimulus pattern imposed on this layer is binary in its format (i.e., consists of only ones (1) and zeroes (0)). The units possess linear activation functions, which simply copy the excitation values they receive into their activation values.

The second layer, known as the *complex log layer*, is identical in shape to the first and consists of units of the same specification. The connections between these two layers provide an approximate polar log transformation of the original stimulus pattern in accordance with the equation $w = \log (z + 1)$. The projection from the stimulus grid contains one connection from each of the units on this layer.

The receiving unit's geometrical position on the log polar layer is equivalent to the output of the log polar transform on the sending unit's geometrical position, with two qualifications. Firstly, the Cartesian coordinates of the units in the first layer must be offset to include equal portions of all the quadrants on the one grid such that the origin of the stimulus grid (i.e., the unit with the *x,y* position of (0,0)) is aligned with the centre of the layer. The resulting representation is necessary for the log polar algorithm to operate in the manner it is hypothesized to work in the human visual system – the fixation point is the (0,0) point at the centre of the network's 'visual field'.

Secondly, the stimulus representation on the log polar grid is an approximation of the complex log mapping function because of the requirement that the destination coordinates take integer values. For example, the 'raw' output from the polar log transform of the coordinates *x,y* (11,7) is, to 9 decimal points, *u,v* (1.115224461, 57.52880771). The accuracy of this function can be extended indefinitely by increasing the decimal point accuracy of its representation. To ensure that each connection from the first layer projects to a single unit on the second layer, the destination coordinates of the projection must be rounded to the nearest integer. In addition, the projection function which determines the connections between the two layers also scales the values to ensure that the connectivity map extends to, but does not exceed, the boundary of the second layer.

### 4.4.13 Stage 2: Autocorrelation – Part A

The second stage of the Connectionist model implements the two-dimensional autocorrelation algorithm. This autocorrelation procedure is theoretically equivalent to, but slightly different in implementation to the transparency analogy given above. In part A of the algorithm, instead of having an original and a copied representation of an image, there is a single layer that detects the correspondences for each theoretical shift. The 'original' and 'copy' are, in effect, superimposed on this second layer, which is called the *shifted layer,* by two converging patterns of connectivity from the log polar grid. These shifted layers will be referred to by the size and direction of the relative shift implied by their operation (e.g., *Shift (0,1)*). Each unit in the complex log layer sends two connections to each shifted layer, both conveying a copy of its activation value (either '1' or '0'). One of these connections terminates at the unit on the shifted layer with exactly the same Cartesian (*x,y*) position. The other projects to the shifted layer unit whose position is equivalent to the sending unit's coordinates minus the offset (where this offset is equal to the relative shift). For example, consider the shift (1,1). The unit at (3,3) in the original or stimulus layer connects to the unit in the shifted layer at (2,2) as well as the unit at (3,3).

The shifted layer's units, having received these superimposed projection patterns, then calculate activation values using the values on the incoming connections as arguments in the activation function shared by all the units on this layer. Autocorrelation theory dictates that this stage of the model perform matrix multiplication (i.e., detects the correspondences mentioned above). However, neurophysiological evidence does not suggest that neurons behave as if their 'activation values' are achieved by the theoretical multiplication of incoming electrochemical signals. Correspondingly, there is very little use in neural networks of such a function. Instead, threshold linear units will be utilized in this project. In the following section, we will explain their theoretical origins, how they operate and how they are functionally equivalent to conventional matrix multiplication in the context of this particular network model.

### 4.4.14 Linear threshold units

One of the most striking aspects of the activity of single neurons is the *action potential* or *spike*. When the input to the cell, in the form of a depolarizing current, reaches a particular critical value, a sudden surge in voltage occurs across the membrane. This action potential is akin to a bolt of lightening, for no sooner has it has started, then it has disappeared (lasting no more than about half a millisecond) and the cell membrane potential returns to its normal resting voltage. Viewed on an oscilloscope, it always has the same shape regardless of any increase in current to the cell. McCulloch and Pitts (1943) were among the first to point out the similarity between this all-or-none aspect of the action potential and the logical functions (e.g., INCLUSIVE OR and AND).

In the present project, this 'logical' aspect of neurons is utilized to mimic the multiplication of inputs by a single unit. This is achieved by the use of *linear threshold units* in the shifted layers – the input to the unit is simply added together while the unit's activation level depends on whether this summed input is over a certain threshold. In a manner reminiscent of the action potential, when the threshold is exceeded the activation value takes a certain value (i.e., '1'). If this cut-off is not passed it takes on another value ('0'). If we set the threshold at '1' with the simulated 'action potential' at '1' and the resting value at '0', then the unit will behave as if it is multiplying the two binary inputs it receives. In other words, the threshold linear unit will effectively signal the detection / non-detection (i.e., a 'true' or 'false' value) of a single correspondence between the original and the copied patterns. If both inputs are '1', the unit will 'fire'. Given any other permissible input patterns, the unit will remain in its 'dormant' or 'quiescent' state' (i.e., not sending a spike).

Figure 4.5 demonstrates the behavioural profile for the full range of input values on the unit. As it clearly shows, for two binary inputs, the response profile of a linear threshold unit with a threshold of '1' and minimum and maximum activation values of '0' and '1' respectively, is equivalent to multiplication. Thus, a two-dimensional grid of these units (in a network layer, for example) will perform the equivalent of matrix multiplication. This profile also demonstrates how the unit is performing the logical AND function and can be viewed as a truth table. The unit is in fact virtually identical to a *McCulloch-Pitts neuron* – the only difference being that the time quantum for integration of inputs can be infinitesimally small, given that both inputs are received simultaneously.

In this model, all the shifted layers (except for *Shift (0,0)*) are truncated. In comparison to the stimulus grid, they are all smaller along at least one dimension. This is because the correspondences between the two patterns ('original' and 'copy') can occur only within the space where the two pattern grids overlap. In the transparency analogy, this truncation amounts to chopping off the edge(s) of either

a) (Matrix) multiplication

| **Input 1** | **Input 2** | **Multiplication Function (Input 1 x Input 2)** | | **Output** |
|---|---|---|---|---|
| 0 | 0 | 0.0 | = | 0 |
| 1 | 0 | 1.0 | = | 0 |
| 0 | 1 | 0.1 | = | 0 |
| 1 | 1 | 1.1 | = | 1 |

b) Linear threshold unit

| **Input 1** | **Input 2** | **Linear threshold Function (Threshold =1 Min = 0 Max = 1)** | | **Output** |
|---|---|---|---|---|
| 0 | 0 | $0+0\leq1$ | = | 0 |
| 1 | 0 | $1+0\leq1$ | = | 0 |
| 0 | 1 | $0+1\leq1$ | = | 0 |
| 1 | 1 | $1+1>1$ | = | 1 |

*Figure 4.5. Linear threshold units and matrix multiplication. The functional equivalence of, a), a linear threshold unit (threshold =1, min = 0, max =1) operating on two binary inputs to the application of, b), (matrix) multiplication. As mentioned in the text, the common input / output profile is identical to the AND function in logic.*

transparency where there is no overlap with the second sheet. Note that in the case of *Shift (0,0)* there is a perfect overlap between all the points. The shape and size of the shifted layer is calculated using the following formulas:

$$x_{\text{max (s)}} = x_{\text{max (o)}} - |x_{\text{shift}}|$$

$y_{\max(s)} = y_{\max(o)} - |y_{\text{shift}}|$

where $x_{max(s)}$ *is the maximum value of x on the shifted layer*

($y_{max(s)}$ *is the maximum value of y on the shifted layer)*

$x_{max(o)}$ *is the maximum value of x on the original stimulus layer*

($y_{max(o)}$ *is the maximum value of y on the original stimulus layer)*

$x_{shift}$ *is the relative shift associated with this shifted layer*

($y_{shift}$ *is the relative shift associated with this shifted layer)*

In other words, the shifted layer is truncated in each dimension by the magnitude of the difference between the original layer and the shift in this direction associated with this layer.

### 4.4.15 Stage 2: Autocorrelation - Part B

The final stage of the autocorrelation process is the integration of all the outputs from the shifted layers onto a single layer that displays the final output of the autocorrelation matrix, i.e., the autocorrelogram. This final layer is known as the *integration layer*. The autocorrelogram is achieved by creating a projection from each shifted layer onto the integration layer. This layer is identical in size and shape to the original stimulus layer. Each projection contains one connection from all the units on a shifted layer. All of these connections within a single projection terminate at the unit on the integration layer whose geometrical position corresponds to the relative shift implied by the shifted layer's operation. For example, all the units in the layer *Shift (2,2)* connect to the unit at the position (2,2) on the integration layer. Note that, as before, the origin (0,0) of the grid is in the very middle of the layer. The integrator layer's units are simple linear summators. As a result, they add up all the correspondences detected by the shift operation implied by the previous layers' calculations. A small scale version of this Connectionist implementation of autocorrelation is given in figure 4.6.

### 4.4.16 Results from the first model

Given the small size of the representational grid (i.e., a 25 by 25 matrix), the image resolution of the model is relatively low. As a result, images are displayed in vertices only and the transformations applied are limited to allow accurate display of shapes at different sizes and rotations. There were also limitations on shapes and transformations used because of the approximation of the complex log mapping. While this restricts the range of the examples used, the model nevertheless demonstrates the essential principles of the algorithm and a larger working model (i.e., differing only in scale) would provide a more realistic variety of images and their transformations.

While the combination of complex log transformation and autocorrelation performed perfectly for simple changes in magnitude, success in the recognition of images that had been rotated was variable. Some of these results are demonstrated in table 4.2. The problem is caused by a characteristic of the complex log mapping and contradicts Schwartz's (1980a) claim that a post-processing of the cortical representation by Uttal's (1975) shift-invariant algorithm, can provide the required invariance.

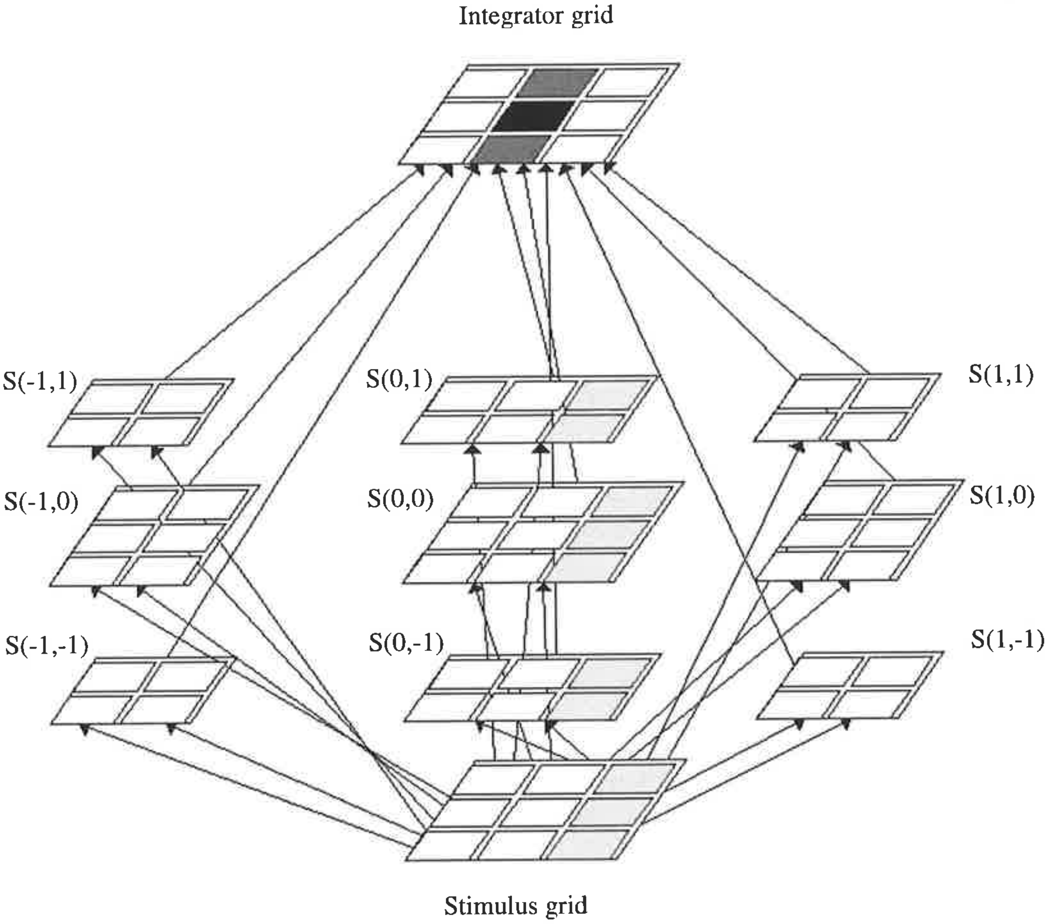


*Figure 4.6. The Connectionist implementation of the autocorrelation algorithm for a 9 x 9 grid. A unit's activity is depicted by different shades of grey scale (white = no activity, light grey = 1, dark grey = 2, black = 3). The projections are depicted by arrows indicating the flow of activation while the individual connections are not pictured. The shifted layers are labeled as "S(x,y)" where x,y represent the relative shift implied by an individual layer's operation.*

Consider the example in figure 4.7 where *a)* and *b)* represent a quadrilateral similar in size and shape rotated $90^0$ about the centre of gravity of the object. Directly below each of these is their respective complex log representations, *c)* and *d).* Autocorrelation of these latter images would not produce the same pattern – this

*Table 4.2.*

***Sample behaviour of the first model***

| Shape description | Geometrical positions of points | Correctly identified? |
|---|---|---|
| **Square** | | |
| Square | (1,1) (1,-1) (-1,-1) (-1,1) | YES |
| Square (magnified [x 2]) | (2,2) (2,-2) (-2,-2) (-2,2) | YES |
| Square (magnified [x 5]) | (5,5) (5,-5) (-5,-5) (-5,5) | YES |
| Square (magnified [x 1/√2], rotated [$45^0$,-$45^0$]) | (0,1) (1,0) (0,-1) (-1,0) | YES |
| Square (magnified [x √194], rotated [÷$21^0$]) | (4,9) (9,-4) (-4,-9) (-9,4) | YES |
| **Triangle** | (5,0) (-5,5) (-5,-5) | YES |
| Triangle (magnified [x 2]) | (10,0) (-10,10) (-10,-10) | YES |
| Triangle (magnified [x √2], rotated [$45^0$]) | (5,5) (0,-10) (-10,0) | NO |
| Triangle (magnified [x √2], rotated [–$45^0$]) | (5,-5) (0,-10) (-10,0) | YES |
| Triangle (rotated [$90^0$]) | (0,5) (-5,-5) (5,-5) | NO |
| Triangle (rotated [-$90^0$]) | (0,-5) (5,5) (-5,5) | YES |
| Triangle (rotated [$180^0$]) | (-5,0) (5,-5) (5,5) | NO |
| **Rectangle** | (5,2) (5,-2) (-5,-2) (-5,2) | YES |
| Rectangle (rotated [$90^0$]) | (2,5) (2,-5) (-2,-5) (-2,5) | NO |
| Rectangle (magnified [x 2]) | (10,4) (10,-4) (-10,-4) (-10,4) | YES |

process produces identical outputs only for patterns differing in "global" *x,y* displacements in Cartesian coordinates. In other words, the autocorrelation algorithm requires that all the points undergo the same linear displacement or translation in the *x,y* plane. The patterns represented by *c)* and *d)* can be conceived of as differing, not only in terms of such a "global" displacement but also in terms of a "local" displacement. All the points have been rotated $90^0$ about the pole. In the Cartesian representation, this amounts to shifting the points along the polar angle axis (*v*) by the same distance in the same direction for all but one point marked **P**.

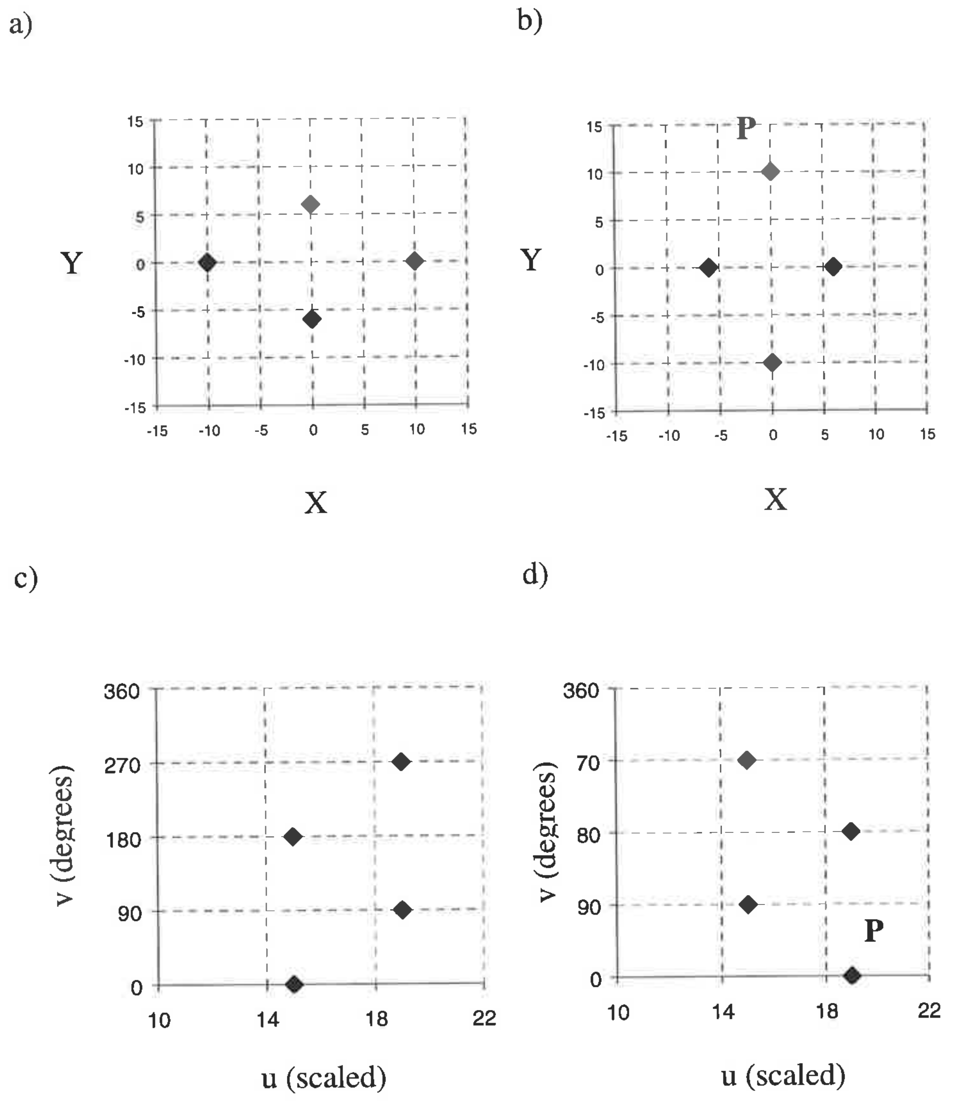


*Figure 4.7. Demonstration of the need for cyclical invariance in processing of the complex log representation. The original images, a) and b), consist of identical shapes differing only in orientation (i.e., a* $90^0$ *rotation) while the graphs directly below each original image, c) and d), represent their respective log polar transformations. The resulting images would produce different autocorrelation images.*

This discrepancy derives from the substitution of points between the two coordinate systems. The Cartesian representation of polar angles must be bounded (in this case $0^0 \leq \varnothing < 360^0$). Otherwise, the same point measured in polar coordinates can be represented (illegally) in two places on the Cartesian plane. Therefore the point, **P**, cannot be represented by $360^0$ in this case and is assigned the polar angle equivalent of $0^0$. In this process, it is shifted "down" and not "up" the $u$ axis as it is theoretically "pushed" over the upper boundary. As a result, **P** no longer shares the same translation in Cartesian coordinates on the $u,v$ plane as the other points and the 'shape' of the object representation has changed. This characteristic, which undermines the effectiveness of the autocorrelation algorithm, can be defined with respect to movement about the initial ray. Whenever a point is rotated through the initial ray (the positive $y$ axis in the above example) the representation of this angular displacement by means of a translation when substituted onto the Cartesian plane will differ from other points which, when rotated in the same way, do not pass through the initial ray.

Clearly then, Schwartz (1980a) was incorrect in claiming invariance to rotation and size symmetries would be possible through the use of "any form of subsequent visual processing which provided shift invariance" on the log polar representation he has specified (p.654). As has been demonstrated above, his suggestion that Uttal's (1975) autocorrelation technique could perform this function is incorrect. A subsequent processing stage needs to provide more than the type of shift invariance contained in Uttal's (1975) technique. What is needed is a function that is sensitive not only to periodic properties but cyclical ones as well.

In their paper on the topological transformations of the early human visual system, which incorporates Schwartz's ideas on retino-cortical mapping, Braccini, Gambardella, and Tagliasco (1982) make brief mention of additional requirements for invariance to rotational, and not just scale, symmetries in such models. The authors suggest two methods that allow for rotations around the origin to be reduced to simple shifts along the ∅ axis when the output is represented in the Cartesian space, (r, ∅). The first is to extend the range of ∅ beyond the 0-2π interval. However, this technique is problematic for two reasons. Firstly, it would allow the same point in the visual field to be represented more than once on the primary visual cortex, a view that contradicts Schwartz's (1980a) theory. Secondly, it does not guarantee the simple shifts required for shift invariance. The only way in which these simple shifts could always be obtained using this method is by arbitrarily changing the limit of the new Cartesian boundaries between different images. This technique would be difficult to justify given that there is no neurophysiological evidence supporting this notion and that it would require the awkward and costly method of searching through the large space of boundaries to identify which pair provides shape invariance. Braccini et al's (1982) second suggestion is to represent ∅ along the circular section of a cylinder with one of the generatrixes of the cylinder being *r*. In the context of cortical representation, this would require that the log polar mapping is not in effect represented in Cartesian space. However, this general notion of cyclical invariance will be utilized in our second model.

### 4.4.17 The second version

A variation of the first version was designed and implemented to solve the above problem. This involved modifying the second part of the model in a way that still involved autocorrelation but that also provided cyclically invariant processing of the stimulus image along the dimension representing angular change. The first two layers of the network, which produced the complex log transform, were unchanged.

However, the second part of the network was modified. This involved implementing an algorithm that performs autocorrelation of a one-dimensional representation of the stimulus. The first layer of the second part of the model is known as the *column summator* and consists of a single row of 25 units, i.e., one unit for each column of units in the complex log representation. Each unit receives a projection from one column of the previous layer which consists of one connection from each of the units in that column. For example, the first unit of this layer receives excitatory input from all of the units in the first column of the complex log representation. The units possess linear activation functions, which sum the excitation values received and copy this total into their activation values.

The next layer is known as the *column activation detector*. As its name suggests, its role is to signify the presence or absence of any activity (i.e., any '1' values) within each column of the complex log representation. Like the column summator, it consists of a single row of 25 units, each of which receives a single excitatory connection from the unit in the corresponding position in the *column summator* layer. However, the units do not possess simple linear activation functions but rather linear threshold functions. The threshold is set at '1' so that the units effectively detect the presence or absence of any activation within each column of the complex log representation. The unit signals detection by taking an activation value of '1' when the input exceeds the threshold and '0' when below.

The next 26 layers essentially perform a one-dimensional autocorrelation on the pattern contained in the *column activation detector* layer. The mechanism used is similar to that employed in the first model. The first 25 of these layers consist of shifted layers which are labeled according to the shift implied in their operations from *Shift (0)* through to *Shift (24)*. As with the first model, the ‘original’ and ‘copy’ of the pattern to be autocorrelated are, in effect, superimposed on each shifted layer by two converging patterns of connectivity from the *column activation detector* layer. Each unit in the *column activation detector* layer sends two connections to each shifted layer, both conveying a copy of its activation value (either ‘1’ or ‘0’). One of these connections terminates at the unit on the shifted layer with exactly the same *x* coordinate. The other projects to the shifted layer unit whose position is equivalent to the sending unit’s coordinates minus the offset (where this offset is equal to the relative ‘shift’). Again, like the two-dimensional mechanism for autocorrelation outlined in the first model, the shifted layers are truncated in such a way that they contain only the overlap between the ‘original’ and ‘copied’ projections. The number of units in any layer is given by the equation $(x_{max\,(s)} = x_{max\,(o)} - |x_{shift}|)$ where $x_{max\,(o)}$ is 25 (the maximum possible units in a shifted layer) and $|x_{shift}|$ is the size, in units, of the relative shift.

This final layer is known as the *output layer* and this consists of a row of 26 units. The first 25 of these represent the final output of the one-dimensional autocorrelation algorithm, i.e., the autocorrelogram. This is achieved by creating a projection from each shifted layer onto the output layer. Each of these terminates on a single unit on the output layer in such a way that the *x* coordinate of this single unit is the same value as the label of the shifted layer sending the projection, where the labeling of the units begins at zero. For example, the layer Shift (5), which compares an original copy of the image with a copy translated five units, sends a projection to Unit (5) on the output layer. Each projection contains an excitatory connection from all of the units in a shifted layer. All the units on the output layer are simple linear summators. As a result, they add up all the correspondences detected by the shift operation implied in the previous layers' calculations. The final unit on the output layer, Unit (26), is also a linear summator and receives a single projection from the column summator. This projection contains an excitatory connection from each of the units on the sending layer. In this way, the final unit on the output layer provides a tally of the total number of units active in the original stimulus representation. The final model is depicted in figure 4.8.

### 4.4.18 Results from the second version

The Connectionist model performed successfully in the categorization of a large number of shapes under different, shape-preserving transformations. Interestingly, testing of the performance of the model revealed an emergent property. Although the model was specifically designed to provide invariant recognition behaviour to shapes that are rotated and / or scaled around the center, testing revealed that the model was also invariant to mirror reflections and combinations of any or all three symmetries. In other words, performance was invariant to the set of these transformations. As suggested in chapters 5 and 7, there is some evidence for invariance to mirror reflected images as well as planar rotation and size. In addition, there is some indication that these transformations may be related psychologically in some way. Examples of the model's performance with respect to these properties are given in figures 4.9 and 4.10.

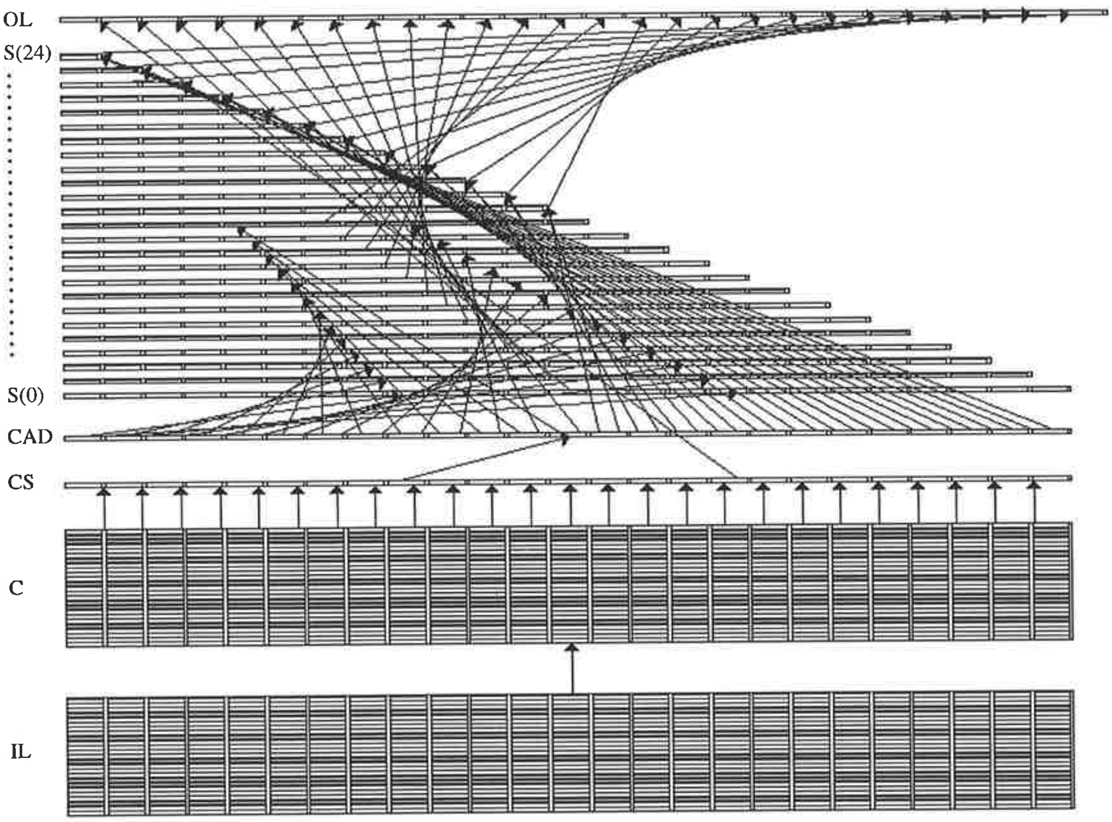


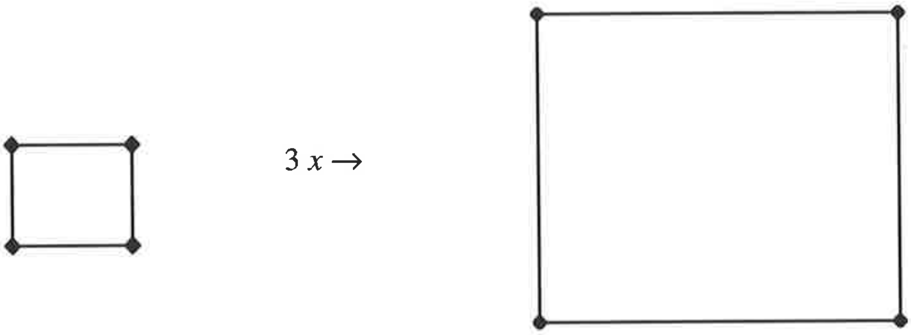
*Figure 4.8. The second Connectionist model. The projections are depicted by arrows indicating the flow of activation while the individual connections are not pictured. The labels for the different layers are as follows: IL = input layer, CL = complex log layer, CS = column summator, CAD = column activation detector, S (x) = shifted layer (where x is the relative shift implied in its operation) and OL = output layer. Note that only the first (S(0)) and last (S(24)) shifted layers are explicitly labeled in the diagram.*

Testing of the model also demonstrated the need for further refinements. One requirement is that images be presented in such a way that the center of gravity of the stimulus shape coincides with the center of the stimulus grid. In the examples given in figures 4.9 and 4.10, this requirement has not been met. The small scale of the model makes the exact centering of such objects problematic – the low resolution of the representation possible with a 25 by 25 grid is inadequate for precise centering of even such simple shapes. In addition, the model performs adequately for the objects given in the examples because only specific conditions are required for the performance of the model to break down. The basic model will fail to distinguish two objectively different shapes when several conditions are met regarding their representation in polar, or log polar, coordinates. Firstly, the values of $\phi$ must be equivalent for all the points in the two objects. Secondly, the length of the radii (whether measured in $r$ or $\log r$) must be equivalent for more than one point in at least one of the different objects. Thirdly, the difference between the two shapes amounts to a difference only in the number of points at a common radius length. An example of these conditions and how they lead to a shortfall in the recognition capabilities of the model is shown in figure 4.11.

However, the problem can be rectified by the additional requirement that the image is aligned so that its center of gravity coincides with the center of the representational grid (i.e., (0,0)). For example, if the two objects in figure 4.11 are aligned in such a manner, there are no points in common between the two images because the center of gravity is different, i.e., Object 1 is represented as {(4,4) (8,-8) (-4,-4) (-8,8)} whereas Object 2 is {(5.68,2.33) (9.68,-9.68) (-2.33,-2.33) (-2.33, 2.33)}. In other words, the position of Object 2, relative to Object 1, is translated in the positive direction along the $x$ axis and in the negative direction along the $y$ axis. This ensures that the conditions, mentioned previously as necessary for the lack of uniqueness in object representation, cannot be met and that the two objects are given unique descriptions by the model.

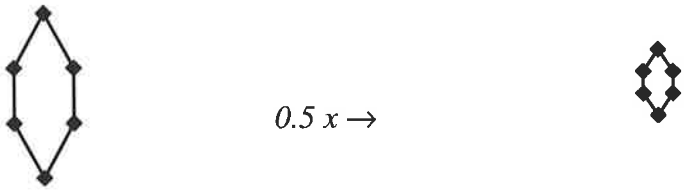
## SCALE

1)



| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **4** |

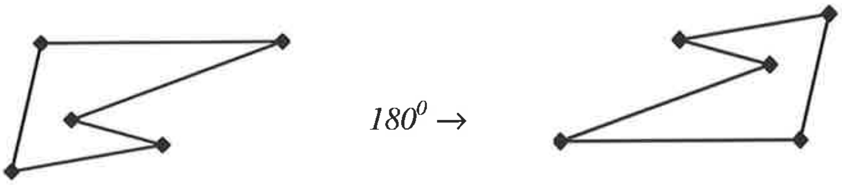
2)



| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2** | 0 | 0 | 0 | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **6** |

## ROTATION

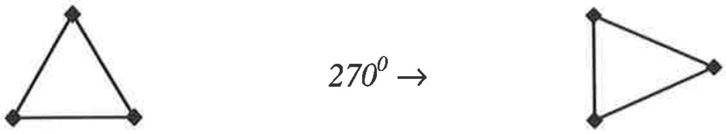
1)



| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **4** | 0 | 0 | **1** | **1** | **1** | 0 | **1** | 0 | **1** | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **5** |

*Figure 4.9. Samples of the network's performance involving size changes, rotations and combinations of the two transformations. Each example depicts the original image, the transformation applied and the resulting image. Underneath these images is the network output that is identical for both images.*

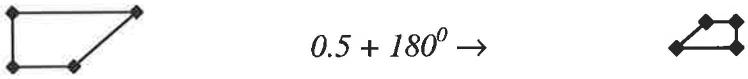
2)



| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2** | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **3** |

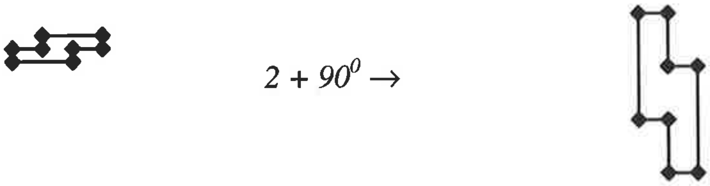
## ROTATION + SCALE

1)



| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2** | 0 | 0 | 0 | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **4** |

2)



| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **3** | 0 | 0 | **1** | 0 | 0 | **1** | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **8** |

*Figure 4.9. (con).*

Interestingly, the necessity of precise centering of the object prior to application of the complex log mapping would fit in nicely with meeting the additional challenge of translational invariance. Without such centering, for two images of the same object to be recognized as one, they would need to be presented in such a way that their centre of gravity was in the same position, relative to the center of the grid. In addition, if the two images differed in terms of size and / or planar rotation, the position of the centre of gravity with respect to the centre of the grid would need to be transformed in a similar manner. Such translations would paradoxically require knowledge of the object and the transformations applied to its stimulus image before recognition. The simplest solution is to align the object's centre of gravity with the centre of the grid before the complex log transformation is applied. Calculation of the centre of gravity $(x_0, y_0)$ does not require object knowledge nor any information of shape preserving transformations applied to the image and can be calculated via the formulas

$$x_o = \sum_{i=1}^{N} u_i d_i / \sum_{i=1}^{N} d_i$$

and

$$y_o = \sum_{i=1}^{N} v_i d_i / \sum_{i=1}^{N} d_i$$

where $(u_i, v_i)$ is the midpoint of the line segment between $(x_i, y_i)$ and $(x_{i+1}, y_{i+1})$, $d_i$ is the length of this line segment and $N$ is the total number of line segments in the shape.

## MIRROR REFLECTION

1)

*MR →*

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **5** | 0 | **2** | 0 | **1** | **1** | 0 | **1** | 0 | **2** | 0 | 0 | 0 | 0 | **1** | 0 | **1** | 0 | **1** | 0 | 0 | 0 | 0 | 0 | **7** |

2)

*MR →*

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **3** | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | **1** | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **5** |

## MIRROR REFLECTION + SCALE

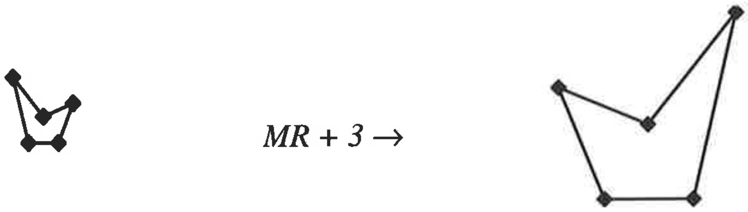
1)

*MR + 0.5 →*

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2** | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **4** |

*Figure 4.10. Samples of the network's performance involving size changes, rotations, mirror reflections and combinations of these transformations. Each example depicts the original image, the transformation applied and the resulting image. Underneath these images is the identical network output for both images.*

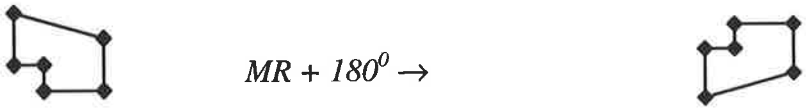
2)



| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |

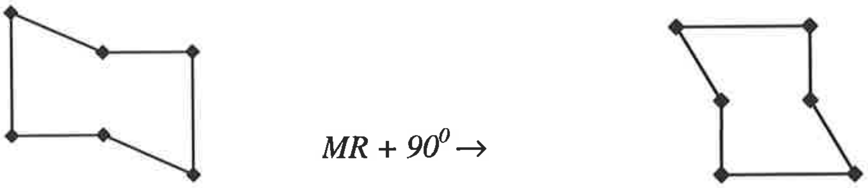
## MIRROR REFLECTION + ROTATION

1)



| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 0 | 1 | 2 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |

2)



| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |

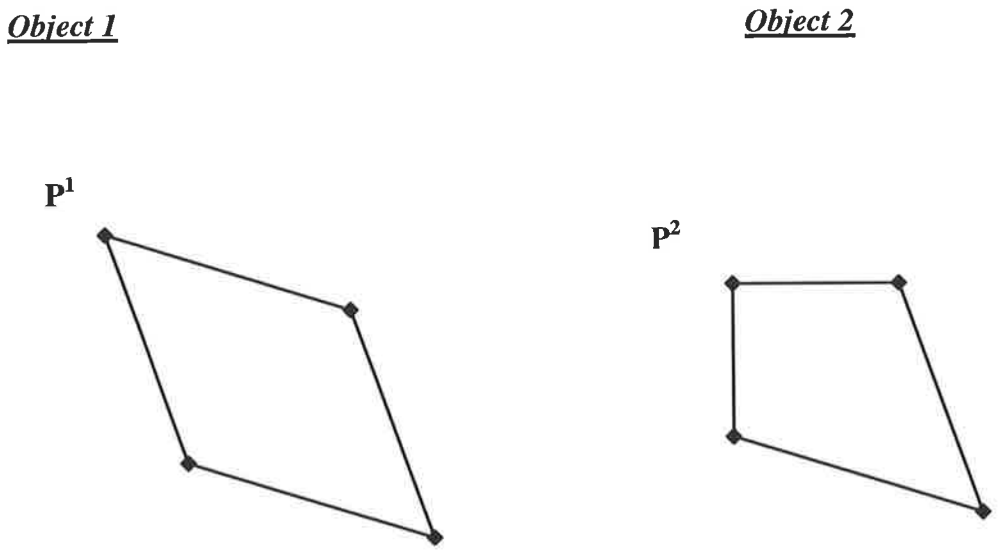
*Figure 4.10. (con).*

This centering procedure would guarantee proper translational invariance which would extend the symmetries accounted for by the model to the entire set of the similarity group of transformations – i.e., translation, size, planar rotation and mirror-inversion. In chapters 5 and 7, we will present evidence that human pattern recognition performance can be invariant to all of these properties. In addition, the centering would also provide the uniqueness of representation for objectively different shapes that could not otherwise be guaranteed.

### 4.4.19 Conclusion

Our approach represents a different perspective to other accounts that also provide recognition performance invariant to shape-preserving symmetries. Certainly, there are some similarities with the theories of Marr and Biederman outlined previously. With the requirement of centering the object on the stimulus grid, the model essentially involves object-centred coordinates. Likewise, the model also provides invariance to the planar rotations and size changes. However, unlike the models previously outlined, it does not provide invariance to rotations in depth. Whether this characteristic is more or less psychologically realistic is a matter for further research. There is some empirical evidence that the different rotations are compensated for at different stages of processing in the visual system and we will discuss this research in detail in section 5.14.

It may be that an additional mechanism or structure is used to deal with rotations out of the picture plane. In fact, as mentioned previously, Schwartz (1981b) has demonstrated how the complex log representation can also be utilized in a more



Object 1

Object 2

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2** | 0 | 0 | 0 | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **4** |

*Figure 4.11. An example of how the behaviour of the basic model can break down. In this case, the model cannot discriminate between the objectively different shapes of Object 1 and Object 2. The difference between the two objects is limited to variation in a single point, $P^1$ and $P^2$ between the two images.*

complex model to account for rotations in depth. In this case, the invariant measure used to provide shape constancy is the complex cross ratio. However, a complete working model of shape recognition would require further refinements than the basic algorithm outlined by Schwartz (1981b). For one thing, Edelman and Weinshall (1998) demonstrate that there are no three-dimensional operators, including the complex cross ratio, that can identify any possible object. Any working model relying exclusively on this invariant would require certain restrictions on the class of objects permitted. Without such a restriction, there would be cases where the model would not be able to discriminate between different objects.

It is currently 'de rigueur' for most visual models to incorporate some elements of neurophysiology in their design. However, the Connectionist model we have outlined utilizes aspects of brain anatomy in a novel way – it draws attention to the notion that the topography of stimulus mapping may provide an important function in visual processing (i.e., the idea of computational anatomy). This approach is relatively uncommon. However, we believe it may be important because the architecture of brain areas such as the visual cortex suggest "that a geometric mode of functioning may complement the combinatorial functions that are suggested by neural connectivity" (Schwartz, 1982, p. 164). Our model incorporates both of these approaches.

Perhaps the fundamental difference between the models of Marr and Biederman and the Connectionist model outlined above is that our model does not require the reconstruction of the viewed shape. The theory is based solely on bottom-up processing and reliance on this characteristic alone is unfeasible in a complete working model of human recognition. For one thing, higher level hypotheses regarding object identity often influence our recognition of objects, especially under reduced viewing conditions, such as poor visibility. Such a characteristic is fundamental to human perception and the current model would need to be extended, if possible, to account for this feature. Clearly, the model requires further development to provide a more complete account of visual processing. In fact, this Connectionist model might be better viewed as a mechanism that could be incorporated into other theories rather than a complete theory in its own right. However, as a simple working model the network provides a novel approach to shape recognition that complements more traditional approaches. In addition, it draws attention to several ideas, such as computational anatomy and the notion of the brain as a geometrical computer, which may play an important role in the human visual system.

# CHAPTER 5: The effects of planar rotation

*This chapter looks at the influence of rotation on object recognition. There is convincing evidence of normalization for orientation and also of the ability to form several representations of the same object at various orientations to use in such a normalization procedure. However, there is also evidence for recognition that is invariant to planar rotated stimuli. In the experimental literature, several factors appear to contribute to the use of such a mechanism. These include the type of foil images, the amount of practice, timing within the experimental paradigm, the context in which images are presented and stimulus complexity. Such orientation invariant recognition strategies are linked to specific stimuli and the tendency towards using them varies between individual subjects.*

## 5.1 Introduction

A large amount of research has investigated the effects of planar rotation on object recognition and the results from these studies are quite varied. This chapter attempts to bring together this body of research and to examine reasons for the similarities and differences between the various studies. Our first task will be to discuss how we define, in psychological terms, the orientation of an object. Secondly, we will examine the findings from tasks involving object recognition and contrast these results with those from mental rotation experiments. This will involve presenting evidence of orientation-invariant processing that is both direct and indirect and the conditions necessary for such processing to occur. Finally, we will describe in more detail the nature of this orientation-invariant processing as demonstrated in the experimental literature.

## 5.2 Frame of reference

Orientation is a relational property – it is always defined with respect to a particular standard. In psychological terms, this standard is called a *perceptual reference frame*. This frame is the result of a process or structure that is conceptualized as a grid which assigns direction (top, bottom, and sides) in perceptual space. The effects of orientation on recognition tasks are understood in relation to how the perceptual reference frame has been determined – i.e., the object's orientation is the relationship between the object and this frame. Takano (1989) has pointed out that the perceptual reference frame is influenced by a minimum of four factors. Environmental, retinal, bodily and gravitational direction all contribute to the designation of orientation information (Howard, 1982; Parker, Poston, & Gulledge, 1983; Rock, 1956, 1973; Rock & Heimer, 1957; Templeton, 1973). These factors all influence the 'up/down' and 'left/right' directions in our visual field and therefore the orientation of the image that we are viewing.

In the studies we will examine, and in the experiments we have conducted, the individual contributions of these four factors in determining the perceptual frame of reference cannot be ascertained. This is because the influence of these factors was not orthogonally manipulated (e.g., subjects were sitting upright and their heads were not tilted to one side). Our interest in not in the frame itself nor in what influences its designation but rather how objects, presented at different orientations within that frame, are recognized. Nevertheless, it is worth noting that one of the differences between naming tasks and handedness tasks appears to be how the reference frame is assigned. In the naming tasks, it is more closely aligned with retinal than with environmental directions, whereas in the handedness discrimination tasks, it is more closely aligned with environmental directions (McMullen & Jolicoeur, 1990).

## 5.3 Speed of normalization is faster outside of the handedness task

Outside of the handedness or mental rotation paradigm, a common finding in reaction time studies involving rotated images is that the normalization process, when it occurs, takes place at a faster rate than that of mental rotation. In Corballis, Zbrodoff and Roldan's (1976) experiment, subjects undertook the typical Cooper and Shepard (1973) task where the subject was required to verify the 'handedness' (i.e., indicate whether the figure is 'normal' or 'mirror reflected' about the *y* axis) of certain rotated alphanumeric characters (*G, J, R, & 2, 5, 7*). Corballis et al. (1978) replicated this experiment with the same apparatus and stimuli, but changed the task to one of recognition – subjects were now required to verify whether the image belonged to a set of previously presented images. Response times under this recognition paradigm were consistently shorter than those for the previously published verification task. For example, the average time to identify a character at $60^0$ tilt was 847ms but the mean time to verify a character from the same set was 890ms (Corballis et al., 1978, p.105). Due to this speed difference between tasks, Cooper and Shepard (1973) have suggested that, at least for rotated letters, mental rotation may not be the mechanism used for identification purposes.

Many studies using the identification paradigm have demonstrated a linear, proportional relationship between response time and angular deviation indicative of a rate of normalization faster than that found under the mental rotation paradigm. In Shepard and Metzler's (1988) summary the mean slopes ranged from 327 deg/s (degrees per second) to 621 deg/s. Many identification studies showing statistically significant effects of orientation on response latencies have not reported any rates or such rates are not easily estimated from the data provided (e.g., Corballis et al., 1976; Corballis et al., 1978; Humphrey & Lupker, 1993; Jolicoeur, Snow & Murray, 1987; Kubovy & Podgorny, 1981; Larsen, 1985; Murray, Jolicoeur, McMullen & Ingleton, 1993; Takano, 1989). When reported, speeds ranging from 2180 deg/s to 2640 deg/s have been demonstrated in a simultaneous presentation task (Simion, Bagnara, Roncata & Umilta, 1982) while rates of mental rotation of 900 deg/s, 1176 deg/s and 1286 deg/s have been reported for singular presentation tasks (Shinar & Owen, 1973; Jolicoeur, 1985; Sekuler & Nash, 1972).

Exceptions to this trend can be found in the studies by Tarr and Pinker (1989, 1990). Using similar stimuli for all their experiments, they found speeds of normalization for identification tasks (276 deg/s to 413 deg/s) to be well within those typically associated with mental rotation and, accordingly, similar to those found in their own verification task (Tarr & Pinker, 1989 [Experiment 1]) which utilized mirror reflected nontargets or foils (284 deg/s). However, Tarr and Pinker's (1989) study demonstrates other unique effects that are not common within the experimental literature. In section 5.8.1 we will discuss their findings in more detail. One interesting feature of their experiments - which may be the cause of the discrepancies

- is the complex and artificial nature of their stimuli, which may have induced a slower mental rotation procedure.

## 5.4 Task differences in rotation speed: The influence of orientation-invariant processing?

The simplest and most obvious explanation for the different overall rotation rates between identification and verification experiments is that the normalization mechanism is operating at different speeds between the two tasks. But an alternative explanation is possible, i.e., the reduction in overall rates may be due to the presence of orientation-invariant processing among some subjects. In other words, the rotation rate for normalization may be identical between tasks *for those subjects rotating the images*. But the 'rate' in the mean data may be increased because other subjects are employing an orientation-invariant strategy that yields a flat response profile. The overall 'rotation' slope derived from the aggregate data would then be intermediate between profiles for normalization and orientation invariant strategies. Its exact value would be dependent on the ratio of subjects using the two different methods.

For example, let us assume a rotation rate of 500 deg/s for the normalization strategy – a figure that is realistic in light of Shepard and Metzler's (1988) review of verification studies. If one half of the subjects normalized the image while the other half used an orientation-invariant strategy (with a slope of 0 deg/s), the overall rate would be 1000 deg/s. Note that this is similar to overall rotation rates found in identification studies (e.g., 1176 deg/s in Jolicoeur (1985)). This hypothesis is not unreasonable given the other sources of evidence for orientation-invariant processing given in the following sections. In fact, one study by McKone and Grenfell (1999), which will be discussed in section 5.11, has demonstrated evidence for individual differences in the choice of either normalization or orientation invariant recognition strategies.

## 5.5 The M-shape profile and other non-monotonic effects

Several experiments on naming and identification have demonstrated an M-shaped distribution of response times across orientations that contrasts with the monotonic distributions usually found in mental rotation experiments (Corballis et al., 1978 [Experiment 1]; Eley, 1982; Jolicoeur, 1985; Jolicoeur, 1988; Jolicoeur & Milliken, 1989; Jolicoeur et al., 1987; Pierret & Peronnet, 1994 [Experiment I]). Although still symmetrical about the $180^0$ point, response times show an unusual 'reversal' of direction close to this point. As the angular deviation from the upright increases in $60^0$ increments (as used in all the above studies), there is a corresponding linear increase in response times for all but the one of the measurements – at $180^0$ the mean latency is *shorter* than that for $120^0$ and $240^0$. This results in a local minimum or dip at the $180^0$ point where, under the normalization hypothesis, times should be longest. This gives the overall response profile an "M"-shape for angular deviations from the upright through $360^0$.

For example, although Eley's (1982) study did not report any statistical analyses with respect to this non-monotonicity, the average response time to letter-like forms oriented at $180^0$ was quicker than for those at $120^0$ and $240^0$ orientations in 3 of the 4 conditions. In addition, $180^0$ was the only orientation at which there was a change in the direction of the difference in response times between high and low familiarity groups within both conditions with respect to set-size (i.e., the low familiarity group responded, on average, quicker than the high familiarity group in the 20 symbol set condition only at $180^0$ orientations). Similarly, in comparing performance on the identification of rotated images and a task similar to mental rotation (involving a left / right directionality judgment), Jolicoeur (1988) discovered that the significant difference in slopes disappears when the results from the $180^0$ rotated objects are discarded. While the trend is non-monotonic for the assignment of direction task, the latency profile for the identification task is M-shaped even though the same stimuli were used in both experiments. Therefore, the M-shape effect is a feature that distinguishes performance on recognition tasks from results in handedness tasks. Jolicoeur (1992) has suggested that such non-monotonicity is the hallmark of "the operation of more than one underlying mechanism" (p.184). He suggests that the other recognition mechanism, which operates in addition to normalization in such tasks, is orientation invariant.

In addition to the M-shape effect, there is evidence of other, more serious breaches of non-monotonicity in the experimental literature. Lawson and Jolicoeur (1999) examined the response latencies to name line drawings of familiar objects that were rotated in the picture plane using brief presentation times and pattern masking. The majority of other studies have investigated RTs using angular deviations separated by $60^0$, i.e., $0^0$, $60^0$, $120^0$ and $240^0$ (e.g., Murray, 1995; Jolicoeur et al., 1987). However, Lawson and Jolicoeur (1999) employed a more fine grain analysis involving angular deviations separated by $30^0$ intervals. The response time profile demonstrated a breach in non-monotonicity more severe than simply an M-shape effect. Latencies at $30^0$, $90^0$, $150^0$ and $180^0$ were shorter than would have been predicted on the basis of response times to angular deviations at $0^0$, $60^0$ and $120^0$. Both the non-monotonicity evidence in Lawson and Jolicoeur's (1999) experiment and in those demonstrating the M-shape effect mentioned above, suggest the involvement of at least one recognition mechanism in addition to normalization. In the following sections, we will examine further evidence for shape recognition that is invariant to orientation.

## 5.6 The influence of mirror-reflected foils

In the classical mental rotation paradigm, the foil or distractor trials are mirror-reflected versions of the match images. But, are such experiments true tests of shape recognition? By our definition in section 1.3, a mirror-reflected variant of an image contains the same objective shape, i.e., shape is invariant under such a transformation. According to this account, the handedness task is not a test of shape recognition but reduces to the discrimination of a spatial property of a stimulus image that is independent of shape, i.e., left-right directionality. The identification of objective shape would be a useless strategy in such mental rotation tasks because it would provide a 'match' response to all the trials in the experiment, including foil trials. If we possess a recognition mechanism that can pick out the property of objective shape then it can be productive only in experiments free from mirror-reflected foils (MRFs). As mentioned previously, there are qualitative differences in response behaviours between the two tasks. The most salient of these differences is in the use of orientation-invariant recognition. Some type of normalization procedure is evident in both types of task. However, orientation-invariant behaviour is almost entirely restricted to experiments outside of the handedness paradigm.

### 5.6.1 Foil responses

We will first examine responses to the foil trials in matching experiments in singular, successive and simultaneous presentation paradigms. Little attention has been paid to foil trial responses in the experimental literature on object recognition with the exception being studies investigating the effects of size (see section 7.12.1). In fact, many researchers do not report analyses of data from foil trials because "the concept of normalization is most readily defined when the two patterns are brought into congruence" (Kubovy & Podgorny, 1981, p.20). However, where such data are reported, the effects of orientation in responses to foil trials appear to differ between tasks using MRFs and those employing different, objective shapes.

For example, as mentioned previously the RTs in Cooper's (1975) handedness task demonstrated a remarkably similar slope across angular deviations from the upright for both targets and non-targets – both types of images appear to be rotated to the upright for the purposes of recognition (see also Cohen & Kubovy, 1993 [Experiment 4]; Eley, 1982 [Experiment II]; Pierret & Peronnet, 1994 [Experiment II]). In Murray's (1999) first experiment, responses to foil images, which were entirely different objective shapes, also demonstrated normalization. However, no other matching experiments which involve foils that are entirely different objective shapes report any evidence for normalization for orientation in the foil responses

(Corballis et al., 1978 [Experiment III]; Humphrey & Lupker, 1993 [1]; Humphreys, 1984; Shinar & Owen, 1973; Simion et al., 1982; Pierret & Peronnet, 1994 [Experiment I]). This finding suggests that, when the matching task involves objectively different shapes and not simply mirror-reflected variations, orientation invariant processing is possible for the foil trials.

### 5.6.2 Match trials

While the type of distractor appears to influence processing on the foil trials themselves, it also appears to influence how the match images are recognized. In the studies listed above employing objectively different foil shapes, all but two also demonstrate evidence for either exclusive orientation-invariant processing throughout the experiment (Corballis et al., 1978 [Experiment III]) or both orientation-invariant and normalization procedures on match trials (Shinar & Owen, 1973; Simion et al., 1982). This pattern suggests that orientation-invariant processing for match as well as foil trials is possible only when the matching task does not involve MRFs.

The absence of mirror-reflected foils may also account for the demonstration of orientation-insensitivity in a variety of other experimental paradigms. In addition to matching tasks where the response is either 'match' or 'non-match', paradigms using more specific verbal, semantic labels (e.g., the name of the basic-level class of the depicted object) have often been used to investigate the effects of planar orientation

---

[1] The increase in RT for foil trials where the images are separated by a $120^0$ deviation over foil trials where the images are in the same orientation was only 16.5 ms. This result is unlikely to be due to normalization because it yields an unrealistically high rotation rate of 7273 deg/s.

changes on recognition. Whether they use simple naming paradigms (Braine, 1965; Corballis et al., 1978; Eley, 1982; Jolicoeur, 1985, 1988; Jolicoeur & Milliken, 1989; Jolicoeur et al., 1987; Maki, 1986; McMullen & Jolicoeur, 1992; Murray, 1995; Murray et al., 1993; White, 1980), a repetition priming task (McKone & Grenfell, 1999) or a word-picture verification task (Lawson & Jolicoeur, 1999; Leek, 1998) all such studies demonstrate evidence of orientation-invariant recognition strategies. By their nature, these tasks do not use MRFs.

The hypothesis that orientation-invariant processing is limited to studies free from such 'same shape' distractors has received further support from several comparison studies between identification and handedness tasks using either the exact same root stimuli or the same type of stimuli. Corballis et al. (1978) demonstrated orientation-independent processing for the set of alphanumeric stimuli – *G, J, R, 2, 5* and *7*. However, when the task involved discriminating between certain letters and their mirror-reflected cousins in the lowercase alphabet, e.g., *b* versus *d* and *p* versus *q*, a normalization procedure was used exclusively (Corballis & McLaren, 1984). These findings led Corballis (1988) to hypothesize that "mental rotation … seems to have a special function in the discrimination of mirror-images" (p.116).

Eley (1982) has also demonstrated the influence of MRFs in recognition strategies using, not letters, but letter-like stimuli. He compared a symbol identification task with a symbol verification or handedness task. The former required the explicit naming of the image (e.g., trigrams such as "SIF" and "MEC") while the latter required only a "normal" or "reflected" response from the subjects. Consequently, although both experiments utilized the same target images (simple, letter-like shapes at $0^0$, $120^0$, $180^0$, $240^0$ and $300^0$ rotations from the upright), the second experiment employed mirror-reflections as distractors or non-targets. RTs in the identification task revealed no significant differences across the different orientations. In contrast, the handedness task did reveal a significant variation in latencies associated with orientation for both standard and reflected images. Response times were consistently longer for non-target images and were directly proportional to the angular deviation from the upright. The use of orientation-invariant recognition in identification tasks and the lack thereof in handedness verification tasks has also been demonstrated for line drawings of common objects in both between- (Jolicoeur, 1985) and within-subject designs (Jolicoeur, 1988).

At first glance, Pierret and Peronnet's (1994) paper also appears to unequivocally support the notion that foil images must depict objectively different shapes to invoke orientation invariant processing on match trials. However, a closer examination of their data reveals that evidence of the role of orientation invariant processing in their experiment may be less direct than the authors have claimed. Pierret and Peronnet (1994) used similar polygonal stimuli in two same / different matching tasks - the first with objectively different shapes for foil images (Experiment I) and the second with mirror-image variants of the match images (Experiment II). However, while the findings of the latter experiment demonstrated the relatively uncontroversial mental rotation effect with slow normalization for both types of trials, the authors' claim that there are no normalization effects in the former task may not be justified.

Using a linear curve fitting procedure on the data presented in their paper, the slope of the matching RTs in Experiment I may be estimated at around 892 deg/s, omitting the mean latency score at $180^0$ where there is the aforementioned dip of the M–effect (or 1185 deg/s if the estimation includes this angle). As outlined in section 5.3, this is an acceptable rate of normalization in tasks outside of the handedness paradigm. For their second experiment, the rate of normalization may be estimated at 358 deg/s which is consistent with rotation rates from other mental rotation experiments. Our analysis suggests that, contrary to the authors' claim, normalization was present for match trials when foils were objectively different shapes. However, the flat slope on foil trials, the M-effect profile on match response latencies and the increase in 'speed' of normalization outside of the mental rotation experiment, are consistent with the differences between the influences of MRFs and entirely different shapes noted above. As we have previously discussed, such differences may be consistent with interference from orientation-invariant processing. However, the stronger claim that there is direct evidence of orientation invariant processing in their first experiment, where the foils are objectively different shapes, is not supported by our analysis.

The overwhelming majority of the experiments reviewed demonstrate that orientation-invariant recognition is limited to studies free from MRFs. Two studies appear to contradict this trend. Tarr and Pinker (1989) and Cohen and Kubovy (1993) both used mirror-reflected foils yet demonstrated, in addition to normalization, invariance to orientation in mean RTs. However, there are some exceptional features in both of these studies that suggest qualitatively different processing. These differences are best illustrated with respect to another factor influencing orientation-invariant processing, namely familiarity, which we will discuss in the next section.

Overall, it appears that the ability to process planar rotated images independently of orientation may go hand in hand with the ability to process the stimulus image in terms of objective shape, i.e., the notion of identity that is invariant to, among other transformations, mirror-reflections. It may be that the orientation-invariant processing is the product of a representational code or process that derives objective shape information. If this is the case, then we might expect that recognition might also be invariant to mirror-inversions. As Corballis (1988) has pointed out, "human children and, up to a point, human adults, seem to have a special difficulty with mirror-image discrimination" (p.119). He suggests that mirror-reflective invariance, primarily in the y or horizontal axis, may be advantageous in the natural visual world because many natural objects are symmetrical about such an axis. In fact, invariance to mirror-reflections in visual object priming has been demonstrated by Biederman and Cooper (1991a). Likewise, object naming is unaffected by mirror-inversions (Corballis & Beale, 1976; Rock, 1973). In fact, in chapter 7 we will examine evidence for invariance to another member of the similarity group, scale changes, and demonstrate some of the parallels with the effects of orientation. The notion of an objective shape recognition procedure that is invariant to the symmetries of the similarity group is a speculative one. What does seem clear is that there is sensitivity to the role of different transformations in the experiment and their interaction – in this case, orientation-invariant recognition is almost exclusively associated with tasks free from mirror-image distractors.

## 5.7 Familiarity / Practice

In form recognition tasks outside of the handedness paradigm, the use of normalization for planar rotations is uncontroversial and has been demonstrated with a variety of tasks and stimuli (Arnoult, 1954; Braine, 1965; Corballis et al., 1978; Humphrey & Lupker, 1993; Jolicoeur, 1985, 1988; Jolicoeur & Landau, 1984; Jolicoeur & Milliken, 1989; Jolicoeur et al., 1987; Lawson & Jolicoeur, 1998, 1999; Leek, 1998; Maki, 1986; McKone & Grenfell, 1999; McMullen & Jolicoeur, 1992; Murray, 1995, 1999; Murray et al., 1993; Posner & Mitchell, 1967; Rock, 1973, 1974; Shinar & Owen, 1973; Simion et al., 1982; Takano, 1989; Thompson, 1980).

However, many of these studies also demonstrate a significant attenuation of orientation effects with repeated testing of the same objects. This pattern has been exhibited with stimuli ranging from alphanumeric characters (Corballis et al., 1978; Jolicoeur et al., 1987; Simion et al., 1992), specialized, unfamiliar forms (Takano, 1989; Tarr & Pinker, 1989; Shinar & Owen, 1973), depictions of common, easily named objects (Braine, 1965; Maki, 1986) and line drawings of common objects taken from Snodgrass and Vanderwart (1980) (Jolicoeur, 1985, 1988; Jolicoeur & Milliken, 1989; Lawson & Jolicoeur, 1999; McKone and Grenfell, 1999; McMullen & Jolicoeur, 1992; Murray, 1995, 1999; Murray et al., 1993). All these studies are free from MRFs. This attenuation of orientation effects with practice we shall call a *'trend to invariance'*.

An example of such an effect is in Shinar and Owen's (1973) study which employed Sternberg's (1966) varied set procedure. Eight octagons of equal area were constructed using Attneave and Arnoult's (1956) Method I procedure and these were used throughout the experiment. In each trial, subjects were initially presented with between one and four of these shapes and were asked to memorize them. After subjects had indicated that the memorization process was complete via a foot pedal, there was a two second interval, followed by a single test pattern. This consisted of one of the eight octagons presented in either the upright position or rotated $30^0$, $60^0$ or $90^0$ clockwise. The subjects' task was to indicate whether or not the test pattern belonged to the memorized set by moving a lever on the armrest of the chair left or right. The reaction time data revealed that there were effects of orientation in only the practice block and the first of the three sessions. Shinar and Owen (1973) concluded that practice induced a qualitative change in the representation of the stimulus images and that the subjects "extracted from each pattern some information that was invariant across all orientations" (p.158).

This 'trend to invariance' appears to be object specific. In Jolicoeur's (1985) study, the orientation-invariant processing produced through practice was limited to practised objects – i.e., it did not transfer to new stimuli. In the first five blocks of Experiment III, subjects named black and white drawings of natural objects from Snodgrass and Vanderwart (1980) at various orientations – $0^0$, $60^0$, $120^0$, $180^0$, $240^0$ and $300^0$. The effects of orientation reduced with practice for all the images. However, unseen objects were used in the sixth block. For these unpractised forms, orientation effects were similar to those in the first block. Whatever change in processing has occurred to produce orientation independence it must be in some way stimulus specific. In other words, it is not a change in strategy for all the images presented in the task but is a representational code used only for the practised objects.

There is evidence that the 'trend to invariance' also involves a qualitative shift in processing. In Takano's (1989) second experiment, subjects were asked to discriminate between members of pairs of images. The root figures were highly abstract and artificial and consisted largely of horizontal and vertical lines. There were only two pairs. In the first, the foil figure differed from the match figure by only a mirror-reflection but in the second pair, the images were identical save for one noticeable element (a "hook" shape) that was mirror-reflected between the two images. As a result, this second pair consisted of objectively different shapes although they were easily confused as being mirror-reflected versions of the same shape.

On the whole, subjects discriminated between rotated versions of these figures using normalization at the start of the experiment. However, one half of the subjects were informed of the difference between the construction of the two pairs after the first session. For this group only, there were no significant effects of orientation in the remainder of the experiment for this second pair of figures. Therefore, once informed of the objectively different nature of these shapes, performance was orientation invariant. This suggests a qualitative change in processing and one that may be influenced by higher level control. The fact that 3 out of the 10 subjects reportedly noticed the difference in the uninformed group only after the first session may account for the need for practice in the 'trend to invariance' effects cited previously. In other words, practice may be necessary to make the qualitative shift in processing. This study also suggests that normalization may be used when the two images are highly confusable, i.e., when they are seen as having the same objective shape. When the distinction between the highly similar figures is pointed out, recognition proceeds via orientation invariant means.

As demonstrated by Takano's (1989) findings and those cited above, the attenuation in orientation effects is not generally associated with repeated performance on a handedness discrimination task. Cooper and Shepard (1973) and Bethell-Fox and Shepard (1988) failed to increase the rotation rate above 800 and 813 deg/s even after subjects performed 576 and 250 trials (on just two images) respectively. This post-hoc difference between the effects of practice on orientation sensitivity has also been demonstrated by Jolicoeur in both between- (Jolicoeur, 1985) and within-subject (Jolicoeur, 1988) designs. In other words, using the same match stimuli transformed in the same manner, the reduction in orientation effects was evident only in identification tasks and not handedness discrimination paradigms.

There is some evidence that the effects of orientation can reduce in a handedness task. However, this requires exceptional conditions and there is some proof that this may not in fact be attributable to an orientation invariant processing strategy. Cohen and Kubovy (1993) presented subjects with conventional handedness tasks, using both simultaneous and successive presentations and demonstrated a reduction in the effects of rotation with practice i.e., the hypothetical rates of rotation (2941 deg/s in Experiment II and 1818 deg/s in Experiment III) were deemed to be too fast to be normalization. However, this effect was available only when (a) just two polygons and their mirror-reflections were employed throughout the experiment and (b) when the experiment was conducted under time pressure. This involved the sounding of a beep after the subjects responded whenever the RT was judged too slow. Usually this was when the response time exceeded the .85 quartile of the previous block although this cut-off could be adjusted to prevent an increase in errors. Without such time pressure, there was no reduction of orientation effects. Significantly, the computer also beeped after errors – both incorrect responses and those that were deemed 'slow' received negative feedback. It appears that a transformation to match procedure may not be the only way in which handedness verification can be obtained. However, specific conditions are necessary for this phenomenon to coerce the subjects into adopting an alternative strategy to the preferred normalization mechanism.

Compared to studies free from MRFs that demonstrate orientation invariant processing, Cohen and Kubovy's (1993) experiment involved a substantially greater amount of practice to achieve comparable effects. They used only two root polygons presented in 1440 trials, i.e., 720 trials per root polygon. However, studies free from MRFs have demonstrated orientation-invariant processing in a much smaller number of trials. For example, Shinar and Owen (1973) failed to find any effects of orientation after 32 trials on each object presented in either match or foil conditions, Murray et al. (1993) found no orientation effects after each object was presented only 9 times and McKone and Grenfell (1999) found some subjects who demonstrated insensitivity to rotations even at the start of the experiment. Given Jolicoeur's (1985, 1988) findings (i.e., orientation invariance for identification and not handedness) and the exceptional circumstances in Cohen and Kubovy's (1993) tasks (i.e., time pressure equivalent to error feedback and a very large number of trials with only two root polygons), the attenuation of orientation effects is demonstrably more difficult to achieve, and may require exceptional conditions, in tasks with MRFs than those that are free from them. Cohen and Kubovy (1993) concede "an orientation-free, handedness specific representation is more difficult to achieve than an orientation-free, handedness-free representation" (p.379).

Cohen and Kubovy's (1993) experiments demonstrate that we can form a representation that is both handedness specific and orientation invariant but that this requires exceptional circumstances. In fact, Tarr and Pinker (1989) have shown that the attenuation of orientation effects within a handedness task can be achieved without orientation invariant processing. In their first experiment, Tarr and Pinker (1989) also found significantly reduced orientation effects on a handedness task, again with a large amount of practice (2304 trials) on only three root stimuli. This attenuation was shown to be the result, not of orientation invariant processing, but of a strategy that is essentially an extension of the normalization procedure. It is not clear why this strategy, known as *multiple representation + normalization*, was found in Tarr and Pinker's (1989) but not in Cohen and Kubovy's (1993) task. One possible reason is the lack of time pressure in Tarr and Pinker's (1989) task. The multiple representation + normalization account will be discussed in the next section where we will examine more closely the nature of the attenuation of orientation effects in various studies.

## 5.8 The nature of the diminished effect of rotation

### 5.8.1 The multiple representation + normalization account

What produces this 'trend to invariance' effect? The most obvious answer is that it is produced by orientation invariant processing. However, there is another mechanism, proposed by Tarr and Pinker (1989), that can account for this effect using a variation on the standard normalization procedure. In the following sections, we will address the question of how the effects of orientation may be reduced in these tasks. This will involve assessing evidence for the competing accounts and discussing what factors may influence the use of one strategy over another. Tarr and Pinker's (1989) account has been supported in one study but rejected by many others. We believe that this strategy and one based on orientation invariant processing occur in different experiments and we will discuss the factors that may influence the choice of one strategy over the other. In particular, we will discuss further research into orientation invariant recognition and end the chapter by looking at previous research investigating this effect using the experimental paradigm that we will employ in our first experiment.

To account for the apparent 'trend to invariance' within the general normalization theory, we need to assume that repeated presentations of a stimulus at certain orientations will result in the formation of multiple models of this stimulus at these different orientations. In other words, a separate model is formed for each practised orientation. Therefore, when recognition of this stimulus at a rehearsed orientation is subsequently tested, response times will be unaffected by the angular deviation from the upright - no normalization is required because the canonical representation of one of the models is already aligned to each probe that is tested. According to this *multiple representation + normalization* account, probing with a practised object at an unseen orientation will require a normalization procedure with respect to the nearest model, in terms of angular deviation. In this way, rather than being normalized with respect only to the upright, the image can be rotated to a number of other practised orientations for which there is another "standard" image.

The notion of multiple orientation-specific representations faces some conceptual hurdles. An implication of this account is that there must be some association between all these different models or templates containing the same object shown at different angles. In other words, at some stage in the flow of processing, the output from the different models must converge before the identification response. This ensures that, for example, all the different models of a *dog* seen from different angles are ultimately associated with the same semantic "dog" response. The higher level at which this association occurs could be either non-visual (i.e., involving the semantic label) or visual information (i.e., the geometrical properties). If the association occurs in a non-visual manner, there is consequently no mechanism operating on geometrical or physical properties of the models and the images they selectively match. The perceptual system is oblivious to any visual similarities between models and the images they selectively respond to regardless of the fact that the only difference between them is orientation. As an example, there is no reason why one could not include a model for a *triangle* in the associated multiple models for recognizing a *square* at different orientations. In other words, there is no reason a *triangle* could not be recognized as a *square* despite the obvious visual differences because the association is made at a non-visual level. This is because the link between visual image and semantic label is an arbitrary one.

This option seems intuitively unsatisfying. But the alternative model, whereby the association between the multiple models occurs at a visual level, does not fare much better. This account would not only be a poor fit to the data but it would be a cumbersome model that would violate *Occam's Razor*. For the perceptual system must be capable of 'matching' the multiple models themselves (or their output) based on visual information. This would occur at a level higher than that required for the first stage of normalization but would still be based on visual properties. Either this is achieved by *another* normalization mechanism, i.e., one that attempts to find a match between the model that fits the stimulus image and all the stored models, or some kind of orientation-invariant processor. Ironically, just such an orientation-invariant processor could have eliminated the need for the multiple normalization models in the first place, i.e., it renders the normalization stage superfluous. And, if such an additional mechanism is required, then it should be reflected in an *increase* in overall processing time with practice, given that it could not proceed until after the initial normalization stage is complete. This effect has not been demonstrated in experiments that otherwise find evidence for multiple representations (see Tarr and Pinker's (1989) findings outlined below). Such questions would need to be addressed by a complete account of multiple orientation specific representations.

Support for the multiple models theory has been provided in a series of experiments conducted by Tarr and Pinker (1989). In their second experiment, for example, having studied a sheet containing letter-like asymmetrical characters labeled

"KIP", "KEF" and "KOR", subjects were required to identify these stimuli at various orientations repeatedly presented within a total of either 1320 or 1680 practice trials (1/4 of which were distractor trials utilizing four distractor stimuli). For both conditions, there was a significant effect of orientation on response latencies suggestive of some kind of normalization mechanism. But there were strong practice effects as indicated by the significant effects of order and an interaction between order and orientation. For match stimuli at practised orientations, response times decreased as a function of angular disparity with each successive block.

Such results indicate that, after repeated presentations, the subjects (or an increasingly larger proportion of the subjects) no longer needed to normalize the stimuli into the upright position for identification if the form is presented at a practised orientation. But, for the surprise orientations introduced in the later blocks of the experiment, there was a high correlation between reaction time and nearest practised orientation (.88). In other words, there was a trend for response latencies to increase relative to distance from the nearest practised orientation. This suggests that subjects primarily recognized familiar forms at surprise orientations by rotating them into alignment with the practised orientation that was nearest in angular terms. Tarr and Pinker (1989) concluded that subjects form orientation specific representations for the sake of recognition and that they are "able to use these non-upright representations as targets for mental rotation of characters at new, never-before-seen orientations" (p.257).

At first glance, it might appear that the 'trend to invariance' demonstrated in many studies free from MRFs could be explained by the formation of such multiple, orientation specific models. However, there are several important features of Tarr and Pinker's (1989) study that differ from others demonstrating the 'trend to invariance' effect. These qualities are characteristic, not of tasks free from MRFs, but of mental rotation experiments. Firstly, initial normalization rates, e.g., in the range of 200 to 400 deg/s for Experiment II, are much slower than those in other identification tasks and more similar to those in their own mental rotation task employing the same match stimuli (Experiment I) and to those in other mental rotation experiments (see section 5.3). Secondly, Experiments III and IV of their paper confirm that the multiple representations are handedness-specific. Moreover, in Experiment I, a conventional mental rotation task, the effects of orientation are diminished but not eliminated (i.e., from 284 to 1000 deg/s) with a very large amount of practice (2304 trials on only three root stimuli) via the formation of these orientation- and handedness-specific representations. As discussed previously, the orientation invariant effects in other studies have been almost exclusively associated with tasks free from MRFs and this difference has been confirmed in between-group tasks. In addition, tasks without MRFs require much fewer repeated presentations of each stimulus object in order to achieve this effect.

Such findings suggest that the multiple representation + normalization procedure is an extension of the conventional normalization procedure. Rather than having one single representation (i.e., in the upright orientation), many different representations are formed in order to simplify the amount of normalization required (and thus overall processing time needed) to recognize tilted stimuli. Such a technique can be employed in tasks involving mirror-image discrimination, although only after a large amount of practice as in Experiment I. In other tasks, such as Experiment II, where mirror-image discrimination is not required, we propose that it is employed because of the confusability of the stimuli. As mentioned above, a similar role for normalization in discriminating confusable stimuli has also been demonstrated by Takano (1989). One of the most distinctive features of Tarr and Pinker's (1989) study is their unusual figures and this feature may explain why their findings differ from those in other studies.

In contrast to other studies involving the identification or matching of images, Tarr and Pinker (1989) are alone in choosing forms that are not only open but consist solely of perpendicular line segments. Each 'antenna'-like stimulus contains a clearly defined axis whose direction is also indicated by a small horizontal 'foot' that is not only the shortest, but also the only terminating line segment in the image. In addition to this axis, which defines the 'vertical' axis with its' 'stand'-like feature, there are two more elements making up an image. These two elements vary in rigid ways with respect to this axis. One, consisting of two line segments which meet at right angles at the ends of both lines, is only ever reflected in, and / or translated by a fixed amount along, the 'horizontal' axis. The second element is a line segment whose orientation within the stimulus array is related to the orientation of the same line segment in at least one other image by a $90^0$ rotation about one of its end points.

These design constraints result in stimuli that are not only highly artificial but also very confusable. Thus, as per Takano's (1989) experiment cited previously, we suggest that orientation invariant recognition, as found in other identification experiments, was not employed in Tarr and Pinker's (1989) tasks because of the confusability of the stimuli to be discriminated. Processing always relied on the normalization of one or more orientation- and handedness-specific representations. Perhaps orientation invariant processing relies on more abstract information than normalization and this more abstract information may be similar in all the highly confusable figures in their experiment. Whatever the reason, there is direct evidence for proper orientation invariant processing in other tasks where the figures are not as artificial and confusable as those in Tarr and Pinker's (1989) experiments.

### 5.8.2 Pure orientation invariant processing

One such study by Murray, Jolicoeur, McMullen, and Ingleton (1993) has explicitly tested, and failed to find evidence of, the multiple representation + normalization theory. Using line drawings of natural objects from Snodgrass and Vanderwart (1980) (e.g., an elephant, a windmill), the authors tested naming performance under planar rotation in three 120 trial blocks. The first two consisted of *training* blocks that were followed by a single *transfer* block. In training, subjects named objects twice at the same orientation where the possible orientations were $60^0$, $120^0$, $240^0$ and $300^0$. In the transfer phase, half the objects were presented in the same orientation as in the training phase (*match* transfer). The other half were displayed at a different, unseen orientation (*mismatch* transfer) from the same general set of 4 angular orientations listed above.

Analysis of the results revealed a significant effect of orientation on response latencies but also showed that the effect of orientation diminished significantly over the small number of blocks in the experiment. Importantly, there was no significant interaction between presentation condition (match versus mismatch) and orientation indicating that "the effect of practice did indeed transfer to new orientations" (Murray et al., 1993, p.607). The multiple orientation specific representation + normalization hypothesis would also predict that a $60^0$ disparity between training and transfer orientations of an object should yield shorter response latencies in the transfer phase than when there is a $120^0$ disparity. However, an ANOVA failed to support this difference and suggests that the practice effect was independent of the orientation of the object in training.

In Experiment 2, Murray et al. (1993) performed a similar experiment but utilized a different set of orientations (*surprise* orientations) in the two different phases to examine the possibility raised by Jolicoeur and Milliken (1989) that subjects may form representations of the objects at all the orientations experienced in the training blocks. The results replicated the general pattern found in the first experiment and thus also failed to support the multiple representations hypothesis with no significant interaction between presentation condition and orientation. In summary, Murray et al.'s (1993) study has demonstrated that training in the identification of images rotated in the picture plane not only significantly reduces the effect of orientation for practiced orientations, but that this orientation-invariant benefit transfers to never before seen orientations of the image. The implication is that there is a qualitative shift in processing style – from a normalization procedure to a proper orientation-invariant processing strategy, in which the orientation of the stimulus image no longer influences response latencies.

In addition to Murray et al.'s (1993) experiments, there are other studies that demonstrate orientation invariance that is not attributable to the formation of multiple orientation-specific representations. A study by Jolicoeur and Milliken (1989), which will be discussed in more detail in the next section, demonstrates orientation invariant behaviour in testing after practice with only upright versions. Such findings are not attributable to the self-motivated formation of multiple, orientation specific versions (Murray, 1995). Other studies demonstrate orientation insensitivity from the very beginning of the task (Eley, 1982; Leek, 1998; McKone & Grenfell, 1999; Takano, 1989; White, 1980). Such invariance from the very start of the experiment precludes the formation of multiple representations.

For example, in Leek's (1998) study, reaction times were independent of rotation for some objects even though each object was presented only once in the experiment. Similarly, in McKone and Grenfell's (1999) task the subjects showed flat orientation functions even in the first few blocks where "subjects had received no or little exposure to the objects in multiple orientations and, so, would have had no opportunity to learn view-specific representations" (p. 1597). Finally, the effects of response accuracy to orientation diminished in Braine's (1965) experiment where there was also no opportunity for the formation of multiple representations because subjects never viewed an object in the same orientation more than once. Like Murray et al.'s (1993) study, these experiments all demonstrate true orientation invariant recognition. Such findings contradict Jolicoeur's (1985) claim that that the "identification of patterns in novel orientations cannot be based only on orientation-invariant features" (p.301).

While it seems that true orientation invariance has been demonstrated in a larger number of cases, the multiple representation + normalization procedure exhibited in Tarr and Pinker's (1989) experiments may occupy an important niche role in our perceptual processing. For one thing, it may play an important role in accelerating the recognition of complex and confusable stimuli (see sections 5.8.1 and 5.12). In addition, it may facilitate processing in tasks where mirror-inverted discrimination is necessary. Such a role is realistic given that in Tarr and Pinker's (1989, Experiment I & II) study, the formation of orientation-specific, handedness-specific representations was confirmed not only for a handedness task but also for a recognition task using the same stimuli. In addition, the rates of normalization were similar in both and reflect those found in other conventional handedness or mental rotation studies.

## 5.9 Contextual influences on the use of orientation-invariant processing

The occurrence of this orientation invariant processing also seems to be influenced by contextual factors, as has been demonstrated by Jolicoeur and Milliken (1989). Firstly, the reduction in orientation effects does not occur when the root object is presented only in its upright position and in blocks where other stimuli are also presented in this manner. Secondly, the orientation-invariant effect with respect to a particular object can occur when it is always presented in training in its canonical orientation if, and only if, other objects in the same blocks are presented at varying orientations. In other words, Jolicoeur and Milliken (1989) have demonstrated that the context of disoriented objects in an identification task is sufficient to induce an orientation-invariant effect for an object trained in only one orientation. In addition, this effect is equivalent to that observed when the said object is also rotated within the same context (e.g., in Murray et al., 1993).

There are several implications of this important finding. Firstly, and most generally, it suggests that the orientation invariant processing may be very common in our visual world. Even if an object is never viewed in a manner beside its default orientation, the same may not be true for other objects encountered in similar contexts, and therefore an orientation invariant strategy may be used to recognize it. Secondly, orientation invariant representations are available without having experienced the object at any orientation other than upright – i.e., there is no perceptual learning of the form that requires the derivation of orientation invariant information across different versions. Whatever 'information' is needed for an orientation invariant representation is available from the stimulus presented at only one orientation. Theories that derive orientation invariant information from a single viewpoint such as Schwartz's (1980a) complex log mapping algorithm could provide such performance (see chapter 3). And thirdly, the significant effects of non-simultaneous visual context suggest that, not only is short-term memory involved (at least to 'recall', in some sense, that other stimuli are disoriented) but that some higher-level control or analysis is also necessary. It is not information about the particular object alone that determines whether it is represented in an orientation-invariant manner but also information on context. Perhaps a judgment is made that the object is likely to be seen in different orientations and an orientation-invariant representation is formed to facilitate such processing.

## 5.10 Background processing

There is evidence that practice at tasks that, like the traditional mental rotation experiment, involve a left / right orientation assignment, can facilitate the stimulus specific ‘trend to invariance’ on a later identification task (Jolicoeur, 1988). For those objects first seen in the handedness task, there was a comparative reduction in orientation effects on naming these objects in a later identification task. This transfer of training occurred without any reduction in orientation effects with practice within the handedness task itself. Jolicoeur (1988) suggests that there may be a shared visual representation of the objects between the tasks. It may be that familiarity with the object, regardless of task, facilitates the derivation of an orientation invariant representation. Such a representation would not be useful in the handedness task because it may also be invariant to mirror-inversions. However, familiarity with the object created during the handedness task may facilitate the formation of such an invariant representation in the naming task. In fact, such a representation may be formed in parallel with the representation used in the handedness task and remain dormant until the naming task begins.

On the other hand, one might also speculate about the formation of multiple, orientation and handedness specific representations within the left-right directionality determination task. As Tarr and Pinker’s (1989) experiments demonstrated, such representations are possible in handedness *and* naming tasks. The shared visual representation between tasks in Jolicoeur (1988) may in fact be multiple orientation specific templates. The fact that orientation effects have not yet attenuated by the end of the handedness task does not discount the use of multiple representations - Tarr and

Pinker (1989) still found a rotation rate with respect to the upright no faster than 1000 deg/s in a mental rotation task, despite evidence for multiple representations. Perhaps the representations are not sufficiently formed to be regularly used in the handedness tasks. However, in the naming task, such representations may be adequate for the task or become more full developed within the first block of testing. Such hypotheses would be worthy of further investigation.

## 5.11 Individual differences in the 'trend to invariance'

One particular study has demonstrated individual differences in the 'trend to invariance' effect. McKone and Grenfell's (1999) study showed that, while some subjects take time to develop an orientation invariant recognition strategy, others never use normalization and rely on an orientation invariant recognition strategy throughout the experiment. In a repetition priming experiment, subjects were required to name 54 different shapes again drawn from Snodgrass and Vanderwart (1980). The second of each pair of successive trials contained either (a) the same shape as the first in the same, upright orientation, (b) the same shape at a different orientation, (c) a different shape in the same, upright orientation or (d) a different shape at a different orientation.

In Experiment 1, there was no significant difference in priming across the different orientations in the final block. To examine the data further, McKone and Grenfell (1999) created two subgroups consisting of the five subjects who were most sensitive to orientation in the first 3 blocks and the five subjects who were least sensitive to orientation during the same period. While the former group demonstrated a ‘trend to invariance’ similar to that found in the overall analysis, there were no normalization effects at any point in the experiment for the latter group. Experiment 2 was identical to Experiment 1 except that there was twice the number of blocks and only the two most orientation sensitive subjects from the first experiment took part. By the end of the 24 blocks, there was no difference in priming between objects in the same or different orientations. Thus, even the subjects who were the most sensitive to orientation at the start of the experiment, responded to the differently oriented versions of the same shape equally as quickly after sufficient practice.

There are several important findings from this study. Firstly, normalization was not the default strategy for all subjects and some always used a recognition strategy unaffected by the disorientation of the shape. Secondly, this latter strategy is a pure orientation invariant one – it is not the formation of multiple copies of the shape at various orientations as espoused by Tarr and Pinker (1989). This is because there were no orientation effects at the very start of the experiment when subjects had had no opportunity to form such multiple representations (see section 5.8.2 for other examples). And thirdly, although there is some variation between subjects in the choice of strategy at the start of the experiment, the subject group as a whole eventually settled on the use of an orientation invariant recognition strategy. In other words, if subjects used an orientation invariant strategy at the start, they continued to do so and if they did not begin the experiment in such a manner, they eventually employed an orientation-invariant mechanism anyway. This finding suggests that there may be a benefit in using such a strategy, perhaps in terms of reducing the computational complexity of the task.

## 5.12 Stimulus Complexity

The choice of recognition strategy, whether orientation-sensitive or -insensitive, appears to be influenced by the complexity of the stimulus to be identified. Classic studies by Rock (1973, 1974) have demonstrated extreme effects of normalization in the form of error rates in the recognition of familiar faces. Most studies use the dependent measure of response latencies to demonstrate normalization (e.g., Leek, 1998; Murray, 1995). However, models of normalization such as Huttenlocher and Ullman's (1987) discussed in section 3.3, can easily account for a linear increase in both response latency and accuracy proportional to angular deviation by positing a limited capacity representational device. Similarly, different presentation conditions (e.g. presentation time and the presence / absence of a pattern mask) can invoke normalization effects in error rate (e.g., Jolicoeur & Landau, 1984; Lawson & Jolicoeur, 1998) or response times (e.g., Jolicoeur & Milliken, 1989; Murray et al., 1993). In Rock's experiments on the recognition of faces, the normalization effect is so extreme that even familiar faces (e.g., Marilyn Monroe) are often unrecognizable when inverted (i.e., after a $180^0$ rotation in the plane). The same effect has also been demonstrated using an image of the former British Prime Minister, Margaret Thatcher (Thompson, 1980). Faces are complex objects and discriminating between them is an important perceptual goal.

The use of normalization to recognize such complex shapes may also explain why the stimuli of Tarr and Pinker's (1989) always invoked a slow normalization procedure akin to conventional mental rotation even after repeated presentations, albeit one based on more than one canonical view. The authors suggest that the representations utilized were “concrete and pictorial, in the sense that they are specific to the local arrangement of the objects’ parts in the visual field at a particular viewing orientation” (Tarr & Pinker, 1989, p. 267). In contrast, less complex or confusable stimuli that depict familiar objects but are nonetheless unseen images have been associated with the deployment of an orientation invariant recognition mechanism, sometimes without any need for practice whatsoever (e.g., McKone & Grenfell, 1999; Murray et al, 1993). Similarly, even simple, unfamiliar quasi-random shapes can invoke an orientation-invariant mechanism (e.g., Takano, 1989; Shinar & Owen, 1973). Such post-hoc comparisons suggest that normalization may be required for the recognition of complex stimuli such as faces. But for simple shapes, such a transformation to match procedure may be only optional and the use of an orientation-invariant mechanism is also possible.

There is one caveat to this conclusion - we need to be cautious in interpreting results from studies involving the recognition of faces. Neuropsychological research suggests that the brain possesses a dedicated face processing system (De Haan, 2001) and that this employs more than one mechanism (Sergent & Signoret, 1993). In other words, face recognition and the recognition of other stimuli may take place in distinct areas of the brain. However, our argument regarding stimulus complexity does not rest entirely on findings from studies investigating the recognition of faces. In addition, the similarities between processing in studies involving faces and other stimuli suggest that there may be similar principles at work despite different localization of these mechanisms. The exact nature of the relationship between the recognition of faces and other complex stimuli is worthy of further investigation.

## 5.13 Do abstract representations underlie the orientation invariant effect?

How abstract are the representations used by this orientation insensitive processing? How closely do they relate to the pictorial representation used in the experiment? An innovative study by Leek (1998) has demonstrated that such an orientation-invariant code may be abstract and not intrinsically linked to the specific images used in the task. A ratings study was first conducted to determine two groups of objects to be used in the study based on how such objects are seen in the real world. Mono-oriented objects were defined as those objects that are predominately seen in one orientation (e.g., chair, car) and poly-oriented objects were defined as those that are frequently seen in different orientations (e.g., pencil, key). Subsequent testing using line drawings of planar rotated depictions of the common objects examined in the ratings study revealed significant effects of orientation in RTs for the mono-oriented class. In other words, perceptual experience at recognizing particular objects at various orientations in the real world seems to have instigated an orientation-free representation of such objects in a manner similar to within-task practice effects already demonstrated (e.g., Murray et al., 1993).

What is significant about this study is that the simple line drawing depictions used in the experiment proper differed in myriad ways from the real world objects that had presumably produced the practice effect. Therefore, what was derived for the purposes of orientation-invariant recognition in this case is an abstract representation that can be used for different depictions of the same ‘root’ object. In this study, the code for the presumably orientation-independent representation is therefore not pictorial but highly abstract and schematic. Such a finding also demonstrates that the orientation-invariant effect derived from practice within the tasks outlined previously may be a microcosm for real-world perceptual processing and not simply an experimental artifact. In other words, experiments investigating orientation-invariant representations may generalize to perceptual processing outside of the experimental laboratory.

## 5.14 Timing

The notion that an object is represented in multiple, qualitatively different ways within the brain is a common theme among theories of visual processing (Cavanagh, 1989; Ellis, Allport, Humphreys, & Collis, 1989; Marr, 1982; Marr & Nishihara, 1978). For example, in Marr's (Marr, 1982; Marr & Nishihara, 1978) theory of perceptual processing (section 4.2), the further along the 'stream of processing' the stimulus trace proceeds, the more abstract the stimulus representation becomes. In this way, the representations become more suited to the purpose of explicit identification with the resulting elimination of properties 'irrelevant' to this process (e.g., orientation and size).

Ellis et al. (1989) claim that accessing these qualitatively different representations can be achieved empirically by altering the timing characteristics in the experimental task. Specifically, the authors argue that, in a successive matching paradigm, varying the presentation time and interstimulus interval (ISI) alters the nature of the underlying representation that is accessed. Manipulating experimental variables such as rotation in depth, translation and scale in a successive presentation task involving black and white photographs of basic level, common man-made objects, Ellis et al., (1989) claim to have isolated two temporally distinct visual codes - VIEW and OBJECT (although see Lawson and Humphreys (1996)). The VIEW code is accessed when presentation times are less than 60ms. It is specific to viewing angle, is maskable, decays rapidly such that it is unavailable after a few seconds and may or may not be affected by size changes (it may in fact subsume two different codes – one that is size-specific and one that is size-invariant). But with presentation times longer than 60ms, the OBJECT code is accessed. This is independent of viewing angle and size, is no longer maskable and lasts longer (i.e., for at least a few seconds). In other words, the VIEW code is viewer-centred while the OBJECT code is object-centred.

However, two subsequent studies failed to find support for Ellis et al.'s (1989) theory. Using line drawings of common '3 dimensional' objects at planar orientations of $0^0$, $60^0$ and $120^0$, Humphrey and Lupker (1993) found no evidence that viewpoint specific representations disappear at either longer ISIs or longer presentation times (e.g., 500ms) and suggested that the long term representation of objects was always in the upright position (i.e., not in an orientation invariant form). Murray (1999) also employed a successive matching task in her first experiment in which the stimulus images were presented at various planar orientations. Similarly, there were no differences in sensitivity to orientation between the 250ms and 2s ISI conditions.

While the results of these two studies appear to contradict Ellis et al.'s (1989) theory, the discrepancies may stem from the different types of rotation used. Although Ellis et al. (1989) have couched their theory in general terms (i.e., with reference simply to 'orientation sensitivity'), their own experiments only used rotation in depth transformations. It is still a contentious issue as to whether the two types of rotations (in depth and in the picture plane) are dealt with in the same way by our perceptual system. A correspondence between the effects of the two types of transformations in object recognition is often assumed due to the apparent similarity between mental rotation for plane and in-depth rotated images (see Metzler & Shepard, 1974). But there is some evidence that neither plane rotated (Jolicoeur, Corballis & Lawson, 1998) nor depth-rotated images (Niall, 1997) are recognized via mental rotation.

Certainly, there are some promising post-hoc comparisons between different studies. There is evidence in recognition (Lawson, Humphreys & Watson, 1994; Newell & Findlay, 1997; Rock & DiVita, 1987) and priming tasks (Hayward & Tarr, 1997; Lawson & Humphreys, 1996, 1998b; Srinivas, 1995) of an advantage in naming objects rotated in depth from a particular, canonical view - a finding consistent with a normalization hypothesis. Further evidence for normalization from these canonical views (as exhibited in reaction time data) has been demonstrated by Niall (1997) in other identification tasks. In addition, there is also evidence for orientation-invariant recognition under such rotations (Biederman & Gerhardstein, 1993) and for a 'trend to invariance' to emerge after repeated practice (Lawson & Humphreys, 1998b). These findings parallel those from similar recognition experiments involving rotations in the picture plane.

However, despite such similarities, there is evidence of variations in the way the perceptual system processes changes in planar and depth rotations for several tasks, including recognition. Firstly, there is evidence that rotations in the picture plane may actually produce views that are judged less similar than those produced under equivalent (i.e., same angular deviation) rotations in depth (Langdon, Mayhew, & Frisby, 1991). Secondly, there is neuropsychological evidence demonstrating that our ability to recognize objects rotated in the plane and in depth are distinct. In a review of group studies involving patients with right-hemisphere lesions, Lawson and Humphreys (1998a) have concluded that such subjects, especially those with right-posterior lesions, exhibit a particular deficiency in object constancy. This involves a breakdown in recognition for objects from different perspectives despite an intact ability to recognize objects, even complex ones, seen from prototypical views – a finding demonstrated by the *unusual views* task which involves canonical and unconventional views rotated in depth. But for rotations in the picture plane, Layman and Greene (1988) found no such breakdown in the recognition of novel objects to accompany the failures in object recognition under rotations in depth. This result suggests some distinction, anatomically and psychologically, between the recognition of objects rotated in the two different manners.

Thirdly, in a study investigating the recognition of stimuli where both types of rotation were orthogonally manipulated, Lawson, Humphreys and Jolicoeur (2000) concluded that "different compensating mechanisms are required for plane and depth transformations and that the operations are carried out sequentially" (p.578). There was evidence consistent with normalization for both rotations in RTs and error rates. However, the effects of the two different rotations did not interact and practice reduced the error rates for plane rotations only.

In addition, the authors also demonstrated that the different transformations were compensated for by different stages in visual processing (Experiment III). A psychological refractory period (PRP) experiment was conducted consisting of two tasks. The first required a button press response to a high or low tone. Following a stimulus onset asynchrony (SOA) of 50 or 800ms, subjects were presented with a picture in either a plane disoriented or foreshortened view and were required to name the depicted object. Plane rotation effects were halved by the shorter stimulus onset asynchrony suggesting that compensation for this transformation was conducted by one or more early, parallel processes (i.e., prebottleneck) that occur simultaneously with processing on the tone task. However, the effect of depth rotation was additive with SOA presumably because the compensatory process or processes occur at a later, bottleneck or post-bottleneck stage – i.e., after the tone processing has been accomplished. In other words, compensation for the two types of rotation appear to be achieved by distinct processing stages given the results from both the PRP experiment and the conventional recognition tasks in the same paper via additive stages logic (Sternberg, 1969). Together with the findings with respect to similarity and neuropsychology noted above, there is good reason to believe in fundamental differences between the way the visual system compensates for the two types of rotation.

Such evidence suggests that we should be cautious about simply assuming a parallel between the effects of the two types of rotations. In light of this, Ellis et al.'s (1989) theory may not be directly applicable to rotations in the picture plane. However, there is evidence in the literature that one particular aspect of timing within experimental paradigms, namely presentation duration, may affect planar rotational sensitivity. A post-hoc comparison of studies suggests that longer presentation times may be more strongly associated with orientation invariant recognition and that, conversely, shorter presentation times may be more strongly associated with normalization.

Experiments using alphanumeric stimuli in unlimited presentation time conditions have shown a strong trend in the reduction of orientation effects on response latencies to non-significant levels (e.g., Corballis et al., 1978; Jolicoeur et al., 1987; Simion et al., 1982) or no orientation effects at all (e.g., White, 1980). However, Jolicoeur and Landau's (1984) study using similar stimuli but with brief presentation times (as low as 20ms) followed by a masking image, showed no reduction in orientation effects whatsoever. Similarly, the line drawings of Snodgrass and Vanderwart (1980) have proven successful in eliminating orientation effects when the identification paradigm displays the image until the subjects respond (e.g., Jolicoeur, 1985, 1988; Jolicoeur & Milliken, 1989; Murray et al., 1993). Effects of orientation are also attenuated when subjects practice unspeeded verification on these same images when they are briefly presented (up to 250ms) and masked (Lawson and Jolicoeur, 1999). This study employed the method of ascending limits in which the same object in the same orientation was repeatedly presented at steadily increasing presentation durations until the stimulus was identified correctly. However, when comparatively shorter presentation times are used (42 or 28 ms), again in combination with pattern masking, the same stimuli produced resilient rotation effects throughout the experiment in an unspeeded picture-word verification task (Lawson & Jolicoeur, 1998).

This trend suggests that orientation-invariant representations are not employed under conditions where processing is halted after brief exposures possibly because the orientation invariant representational code has yet to be accessed. The formation of such a representation may require a certain minimum duration of bottom-up processing of the stimulus image. Overall, there is some evidence suggesting that processing time, at least in terms of stimulus duration, may influence the degree of orientation sensitivity perhaps by allowing access to different representational codes which vary in their sensitivity to this transformation. We will investigate the influence of presentation duration on stimulus recognition under planar rotations in the first experiment.

## 5.15 Successive pair presentation paradigms

In our first experiment, we will look at the effects of planar orientation on shape matching with repeated practice on a limited number of shapes in an attempt to invoke the attenuation of the effects of rotation. We will also test for the presence of multiple orientation specific representations within the task. In this section we will review the results from previous experiments using successive pair presentation paradigms. As we will demonstrate, none of the previous studies using a successive pair matching paradigm has demonstrated a ‘trend to invariance’. In the broader category of successive pair paradigms, only one experiment has demonstrated a ‘trend to invariance’ that is not attributable to the formation of multiple orientation specific representations.

Only two experiments have investigated orientation effects within a successive pair (“same” / “different”) matching paradigm. In neither was there an opportunity for shape specific attenuation of orientation effects because each image was presented once only. We have already discussed Pierret and Peronnet's (1994) findings that revealed normalization for match trials only between the first and second item of the pair when foils were entirely different shapes (i.e., not MRFs). Murray's (1999) first experiment revealed that, on same shape trials, images were normalized along the shortest transformational path to match to either a) the orientation of the first image of the pair or b) the upright orientation. This experiment demonstrates a parallel with Koriat and Norman's (1988) findings in a normal / mirror-reversed task (see section 2.4) i.e., that images can be 'recognized' via a transformation to match procedure with one of two representations (upright or 'last seen').

Other experiments in the literature that have employed successive pair presentation paradigms have involved not matching but repetition priming (McKone & Grenfell, 1999) or a mixture of tasks in the same experiment, i.e., determining left / right directionality of the first stimulus followed by naming of the second stimulus (Murray, 1999 [Experiment II]). Both of these have demonstrated the attenuation of orientation effects with practice. However, only McKone and Grenfell's (1999) experiment which involves naming, has demonstrated evidence that these effects were not attributable to the formation of multiple orientation specific representations. This is because, for some subjects, there were no orientation effects near the beginning of the experiment where it is unlikely that they would have had the opportunity to form multiple representations.

Murray (1999 [Experiment II]) used different angular deviations in the final test block and found no corresponding effects of orientation. However, such results may still be attributable to the use of orientation-specific, handedness-specific representations as demonstrated by Tarr and Pinker (1989). The reason for this is that Murray (1999) mirror-inverted the images for use in the final test block. The most obvious interpretation may be that the subjects had formed orientation and mirror-inversion invariant representations. However, Tarr and Pinker (1989) demonstrated effects that could account for Murray's (1999) findings in the context of a multiple representational account and not one involving invariant recognition. In Tarr and Pinker's (1989) fourth experiment, subjects practiced naming the standard versions of images. However, in testing subjects were required to name both the standard and mirror-reversed images at various planar orientations. While subjects appeared to normalize the standard images at least at the start of the testing stage, there were no orientation effects at all on the reversed versions. In fact, the same pattern of results was present in Experiment III where subjects practised on both versions but only traced the standard versions.

Tarr and Pinker (1989) argue that these results are consistent with the formation and use of orientation and handedness specific representations. In the case of the reversed images, subjects were utilizing the image of the standard version that was the most familiar, and rotating it in depth along the shortest transformational path to match with its inverted cousin. The authors cite the work of Parsons (1987a, 1987b) who has demonstrated that the shortest path between a two-dimensional shape and its mirror-image, both of which are in the picture plane, always requires the same amount of rotation ($180^0$) about an axis in the picture plane regardless of planar orientational differences in the two images. Therefore, there is a possibility that subjects formed orientation and handedness specific representations in the first block of the experiment which were subsequently used to match the reversed versions in the final block. In other words, while the effects of orientation have diminished, this may not necessarily be due to orientation invariant processing.

There would also be an incentive to use orientation- and handedness specific representations in this experiment. Such representations may have been formed because the first pair of each trial required a left-right directionality assignment – a task which is closely related to mirror-image discrimination and which obviously involves information on handedness. As mentioned previously, such a task is similar to mental rotation in other ways because, unlike object naming, it is also influenced by real or illusionary rotational movement and does not show any attenuation of orientation effects with practice (Jolicoeur, 1985, 1988; Jolicoeur, Corballis and Lawson, 1998). In our first experiment on shape matching, there are no mirror-reflected versions of the root stimuli and we will also include tests of multiple representations.

# CHAPTER 6: Experiment I

*The first experiment examined the effects of orientation within a successive pair matching paradigm. Stimuli were quasi-random, unfilled heptagons. The type of foil images and practice characteristics of the experiment were those which the experimental review suggested would be associated with the use and development of an orientation invariant recognition strategy. This trend was confirmed in the results and, as hypothesized, there was no evidence that the attenuation of orientation effects was attributable to the formation of multiple orientation specific representations. In addition, there were no meaningful patterns of orientation sensitivity in the foil trials. However, contrary to predictions, there was no variation in the effects of orientation associated with different presentation durations (100ms vs. 500ms) of the first stimulus in each trial.*

## 6.1 Introduction

### 6.1.1 Aims

There were three main aims of Experiment I. The first was to examine whether a 'trend to invariance' would occur under conditions different from those reported in the literature. The first difference was the type of stimuli. Rather than line drawings of readily nameable objects (e.g., Jolicoeur, 1985), letters (e.g., Corballis et al., 1978) or abstract, purposefully constructed stimuli (e.g., Shinar & Owen, 1973; Tarr & Pinker, 1989) this experiment used quasi-randomly generated polygons. The second difference involved the way these stimuli will be presented. The experimental paradigm was a forced-choice, successive pair presentation one, in which subjects provided a *match* or *foil* response by pressing one of two keys on a special response board. It was similar to the handedness task so strongly associated with mental rotation with the notable exception that it employed entirely different shapes on foil trials rather than mirror reflected versions of the match shapes (MRFs). In the main part of our experiment, six differently shaped polygons were used repeatedly in various orientations.

We predicted that the effects of practice would attenuate the effects of rotation. As mentioned above, a 'trend to invariance' has not been established using this experimental paradigm. In addition, we also predicted that the effects of orientation would be limited to foil trials, given that the factors strongly associated with normalization on these trials (namely, confusability of match and foil images or use of the same objective shape for both) were not present in this experiment (see chapter 5).

An advantage in using this paradigm is that it eliminates interference from semantic and vocalization procedures. The shapes were random (or quasi-random) and thus relatively 'meaningless' and subjects responded simply by pressing one of two keys to indicate a *match* or *foil* response rather than giving a vocal response. This contrasts with the commonly used natural line drawings of Snodgrass and Vanderwart (1980) (e.g., depictions of a football helmet and a seahorse), which have been used in many similar studies involving the recognition of rotated shapes (e.g., Jolicoeur, 1985; Jolicoeur, 1988; Jolicoeur and Milliken, 1989; Murray et al., 1993). Such images have strong semantic associations and the responses in these experiments presumably required additional processing for the recollection of a verbal label. This additional processing is not visual but semantic. An example of the dissociability of the two different types of processes is the tip-of-the-tongue state (TOTS). Even when the name of an object eludes us, there may also be an experience of “some degree of partial knowledge, along with a feeling that one could recognize the ‘target’ ” (Brennen, Baguley, Bright & Bruce, 1990).

The final semantic and verbal component in naming tasks introduces variation that is not present when the shapes are random or quasi-random and when responses are not verbal but forced-choice key presses. Subjects will vary with respect to their familiarity with the basic-level object and this may be represented in response time variability. Similarly, the act of vocalization itself may be a source of variation. Thus, the use of quasi-random shapes versus depictions of real-world objects, the requirement of only *match* and *foil* responses versus basic-level labeling and the use of a key response as opposed to a verbal response should all be beneficial in reducing the amount of variation between and within subjects. Note that the type of variation that is being eliminated or minimized is associated with processes that are considered non-visual.

Cooper, Biederman and Hummel (1992) have argued that, unlike tasks involving naming, such successive matching tasks do not access the representations used for object recognition per se. However, as Tarr and Bülthoff (1995) have pointed out, there does not appear to be any experimental paradigm that accurately reflects visually perceiving and recognizing an object because such processing does not necessarily involve an overt response. Therefore, there is "no a priori reason to discount the results of one type of recognition over others" (p.1497).

The second aim of the experiment was to investigate more closely the nature of any 'trend to invariance'. In particular, we were concerned with whether a lack of variation in response times associated with different angular versions of the stimuli was the result either of a) the formation of multiple representations for each different angle (as demonstrated by Tarr & Pinker, 1989) or b) the use of a recognition strategy invariant to such orientation changes (as demonstrated by Murray et al., 1993). We predicted that it would be the latter on the basis that the foils are entirely different objective shapes. No successive pair matching experiment has examined the nature of any 'trend to invariance'. We have also argued that no previous experiment using any successive pair paradigm, allows us to distinguish between these two explanations (see section 5.15). The nature of any attenuation of orientation effects was tested by including test blocks incorporating different orientations to those used in the practice blocks. The design of the experiment also permitted testing for conventional normalization from the upright in these test blocks.

The third and final aim of the experiment was to examine how varying the presentation duration of the first stimulus might affect the degree of orientation sensitivity displayed by the subjects. As discussed in section 5.14, there is some evidence to suggest that timing may influence the choice of recognition strategy (i.e., whether normalization or an orientation invariant strategy is employed). It is hypothesized that a shorter presentation time of 100 ms may be more strongly associated with the use of normalization. This may be demonstrated either by the persistent use of a normalization strategy or by the need for a relatively longer period of time to employ an orientation-free mechanism (i.e., a slower 'trend to invariance'). Similarly, a 500 ms time may be more strongly associated with orientation independence in recognition either by the sole use of such an orientation-free recognition strategy or by a relatively earlier switch to this procedure.

### 6.1.2 Experimental Design Considerations

The angles of $0^0$ and $180^0$ were deliberately excluded from the set of angles used to rotate the second stimulus of the pair. This step was taken on the basis of previous evidence suggesting these angular versions may be associated with effects that a) may be independent of the recognition mechanism itself and that b) could act statistically to obscure a normalization pattern in the response data. Versions at $0^0$ were excluded on the grounds that several studies had exhibited an associated response bias. Both Maki (1986) and Braine (1965) have demonstrated a slight advantage in recognizing the upright shape. This effect persisted throughout the experiments despite the attenuation of orientation effects at all other angles. As mentioned previously, Jolicoeur and Besner (1987) also demonstrated an abnormal effect for images in the same scale and orientation. In this case, the effect occurred in foil trials and consisted of a disadvantage, namely comparatively longer response times for same-sized images. All of these findings might be subsumed by a similar effect – a bias towards 'same' when the image is depicted at the same size and in the same orientation as the match or stored image. These effects have not been replicated in other studies. However, their presence in this experiment could act to conceal the underlying recognition mechanism. In foil trials, it could obscure a normalization procedure and in match trials it could obscure evidence for orientation and size invariant recognition.

With respect to the exclusion of $180^0$ versions of the shapes, it has already been pointed out in section 5.5 that shapes at this orientation may also be responsible for a deviation from the pure normalization profile. Several studies have demonstrated that response times tend to be shorter at $180^0$ (relative to the normalization profile) and result in an M-shaped distribution of times across rotation angles (Corballis et al., 1978 [Experiment 1]; Eley, 1982; Jolicoeur, 1985, 1988; Jolicoeur & Milliken, 1989; Jolicoeur et al., 1987). For example, in Corballis et al.'s (1978) first experiment, the mean response time for recognition of images was shorter at $180^0$ than at $120^0$ or $240^0$. In section 5.5, it was suggested that this might be associated with interference from an orientation-free recognition mechanism. Whatever the exact cause of this M-shape effect and the identical image effect noted above, the angular deviations of $0^0$ and $180^0$ were excluded from the experiment. While interesting in their own right, the biases or influences evident at these rotations might obscure the influence of the overall recognition strategy in the response data.

The design also permitted a degree of pooling across equivalent measurements for the various experimental variables. This gives an increase in measurement accuracy and consistency. We pooled over shape type (match vs. foil), across blocks (pooling either 4 or 2 blocks into one unit) and also across the four angular deviations used within any one block into two rotation levels. The details of these summations will be included in the results section 6.3.

Finally, all comparisons were considered up to the .10 alpha level. This step was taken because a major research hypothesis predicted that there would be *no* differences in means associated with orientation transformations by the end of the experiment. In other words, our research hypothesis is the null hypothesis. By examining the comparisons up to the .10 level, we are therefore increasing the a priori probability of rejecting our research hypothesis, i.e., making every effort to allow this research hypothesis to fail. However, care was taken in the evaluation of effects significant only at the .10 level due to the increased likelihood of spurious or misleading findings.

## 6.2 Methodology

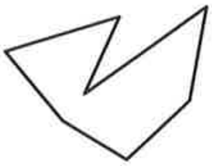
### 6.2.1 Subjects

The 18 subjects participating in this study came from a wide variety of backgrounds and education levels. They included students from the University of Adelaide and the University of South Australia studying a variety of undergraduate and postgraduate courses, as well as members of the general public from various professions. Subjects were either approached by the experimenter or else were recruited by means of posters (see appendix A) displayed on notice boards at both Universities and at various businesses and institutions (e.g., the Australian Bureau of Statistics). Ages ranged from 21 to 50 years with an average of 31.06. Seven of the subjects (38% of the sample) were female.

### 6.2.2 Stimuli

Twelve randomly (or quasi-randomly) generated heptagons (seven sided polygons) were used in the experiment. Each shape was created by (a) generating a list of 7 random *x, y* coordinates on the unit square and then (b) verifying that, by creating line segments between each successive pair, the list defined a closed polygonal image. Shapes were discarded if the resolution of the experimental apparatus was inadequate to display any two line segments of that shape as adequately separated in space. For example, shapes were rejected when the angle between any two line segments was so acute that part or all of the two lines appeared to coincide due to limitations in screen resolution.

All the shapes were displayed in contour form only (i.e., they were presented as a single, continuous line indicating the image perimeter). The size of each shape was adjusted to normalize for surface area. One shape was used as a sample in the information sheet and four were used in the practice stage of the experiment – two as *match* shapes and two as *foil* shapes. These shapes are depicted in figure 6.1. In the experiment proper, the remaining six shapes were used, half as *match* shapes and half as *foil* shapes. These shapes are presented in figure 6.2. Throughout the experiment, these were presented at different orientations and the manner in which this was done is outlined in the *Procedure* section.



*Practice sample*

*Practice match shapes*

*Practice foil shapes*

<u>*Figure 6.1. Sample and practice shapes.*</u> *This figure depicts the shape used for the information sheet (top) and those for the practice stage for match (middle) and foil (bottom) trials.*

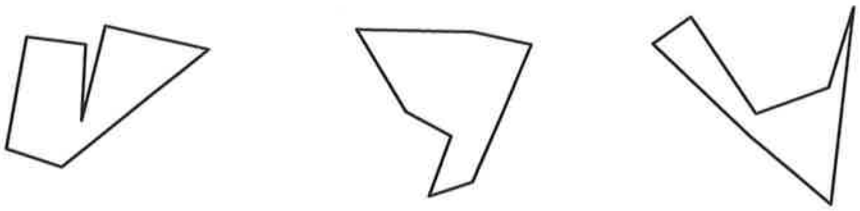
### 6.2.3 Apparatus

Experiments were conducted using a Toshiba T5200 SX portable, IBM-compatible computer. This featured a 386 processor and a mono gas plasma screen with a refresh rate of 16.67ms and a screen size of 11.5" (29.3 cm). Responses were entered via a specially designed response board with two buttons.

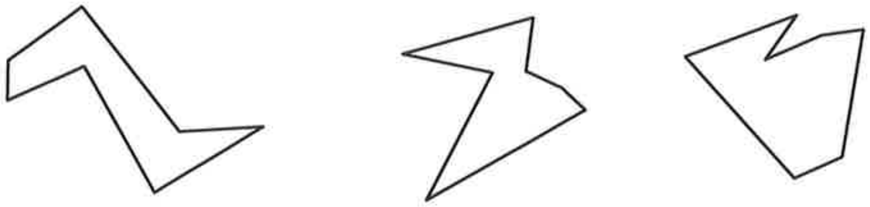
### 6.2.4 Procedure

The experiment involved a successive pair, forced-choice presentation task, where the subjects were assigned, in equal numbers, to one of two conditions. The only difference between the conditions was the presentation duration of the first stimulus in the trial – 100ms or 500ms. There were also two different types of blocks – TRAIN and TEST blocks. The only difference between these blocks was the angular deviations from upright at which the shapes were presented.

For the experiment proper, all the images used were derived from the shapes depicted in figure 6.2. Each image was presented so that its center of gravity (defined as the first moment of area) was aligned with the centre point of the screen. Each trial in the experiment proper commenced with a fixation cross presented for 500ms followed by a blank screen for 250ms. The first stimulus was always one of the four match shapes. These were presented in their original orientation (i.e., $0^0$ angular displacement from the upright) and were selected randomly with the restriction that all 4 match shapes occurred equally as often within each block and the same shape did not occur in more than two consecutive trials. As mentioned above, the duration of presentation for the first member of the stimulus pair varied between groups – i.e., 500ms for the first group and 100ms for the second. This first stimulus was followed by an interstimulus interval (ISI) of 500ms consisting of a blank screen. Then, the second stimulus was presented. This remained until the subject pressed either of the



*Experiment match shapes*



*Experiment foil shapes*

*Figure 6.2. Experiment shapes. This figure depicts the shapes used in the experiment proper in training and test blocks for both match (top) and foil (bottom) trials.*

two keys on the response board – one for a *match* response and one for a *foil* response.

The form of the second image depended on whether a) the trial was a *match* or *foil* trial and b) whether the trial occurred in a TRAIN or TEST block. On *match* trials, the second stimulus image possessed the same shape as the first but was presented at a different angle from the upright. On *foil* trials, the second image consisted of an entirely different shape from the first. These *foil* shapes are pictured in the bottom half of figure 6.2. Like the second shape in the *match* trials, this shape was also presented at different angles from the upright. For both types of trials, the angular deviation from the upright of the second stimulus was determined by whether the trial occurred in a TRAIN or TEST block. For TRAIN blocks the possible orientations were $60^0$, $120^0$, $240^0$ and $300^0$ while for the TEST blocks the orientations used were $90^0$, $270^0$, $210^0$ and $150^0$. These transformations were always conducted around the centroid of the shape. There were 20 blocks in total, each containing 24 trials (a total of 480 trials). Sixteen of the blocks were TRAIN blocks. Two of the TEST blocks occurred at the end of the first half of the experiment, while the remaining pair of TEST blocks occurred at the very end of the experiment (see figure 6.3). At the end of every second block, there was a self-timed break in the presentation of trials and the subjects recommenced the experiment by pressing any key on the keyboard.

There were an equal number of *match* and *foil* trials in each block. In addition, all shapes, angular deviations and combinations of the two appropriate to the type of block (TRAIN vs. TEST) occurred equally often within each block. Each subject was presented with the trials in the same order. This order was generated quasi-randomly within each block with the condition that the same shape never occurred as the first stimulus on more than two consecutive trials.

Subjects were given a copy of the information sheet to read which included images of the sample object presented at different planar rotations (see appendix B). They were then requested to sign a departmental consent form. The following verbal instructions were given at the start of the experiment.

○ ○ ○ ○ ○ ○ ○ ○ ● ● ○ ○ ○ ○ ○ ○ ○ ○ ● ●

KEY

○ = TRAIN BLOCK

● = TEST BLOCK

*Figure 6.3. The presentation order of the TRAIN and TEST blocks in Experiment I.*

> *In each trial in this experiment you will first see a cross on which to centre your gaze. Then this cross will disappear and you will see an image displayed on the screen. After a time, this image will disappear. Then, after a brief pause, a second image will appear. At this point you must decide as quickly as possible whether the second object is the same as the first by pressing either the red key on the right for "Yes" (i.e., "Yes, this is the same object") or the red key on the left for "No" (i.e., "No, this is a different object"). You should respond with the "Yes" key whenever the second image has the same shape as the first even if it has been rotated or turned. Try to respond as quickly as you can but without sacrificing accuracy, (i.e., try to avoid errors caused by trying to respond too fast). Press any key on the keyboard when you are ready to try some practice examples.*

Subjects then attempted 8 practice trials. These trials followed the same format as the main experiment outlined above with the exception that the stimuli used, both for *match* and *foil* trials, were unique to the practice stage. These practice objects are illustrated in figure 6.1. The angular variations for the second stimulus were also limited to the set of orientations used in TRAIN blocks within the main experiment (i.e., $60^0$, $120^0$, $240^0$, $300^0$). Having completed these practice trials, the computer displayed the number of correct responses and the subjects were given the option of repeating the practice trials or beginning the proper experiment. If the subjects decided to start the experiment the following verbal instructions were given:

> *In the experiment, there will be a pause after every 48 trials. When you are ready to recommence the experiment after a pause, simply press any key on the keyboard. Whenever you are ready to start the proper experiment, simply press any key on the keyboard and the trials will commence.*

## **6.3 Results**

In this *Results* section, the terms “highly”, “moderately” and “marginally”, when used in relation to the findings of a statistical test, refer to results achieving significance at the .01, .05 and .10 alpha levels respectively. Where Mauchly's Test of Sphericity is significant or close to significance at the .05 level, this result will be reported together with the relevant effect. In such cases, the degrees of freedom will be adjusted using the Huynh-Feldt correction.

### 6.3.1 Practice Trials

For the purposes of this analysis of practice trials, every four consecutive blocks from the raw experimental data were combined to form a larger block, known forthwith simply as ‘block’. This method provides a more reliable and consistent measurement of the subjects' performance, while still allowing the examination of variation over the course of the experiment. In addition, this pooling allows a better post-hoc comparison between practice blocks in the first and second experiments. In the first experiment, a block consisted of 96 trials while, in the second, a block consisted of 92 trials.

Before analysis, the data was filtered for excessively long reaction times. For trials where latencies were longer than two seconds, the responses were counted as guesses and, regardless of the accuracy of response choice, were treated as errors. Overall response time was fast at 383.65 ms (sd = 57.38) and accuracy good with an average percentage correct of 97% (sd = 1). Two omnibus, repeated measure ANOVAs were conducted on the practice trials - one for response times for correct responses and one for response accuracy, measuring the main and interaction effects of four factors. There was one between-subjects factor, called CONDITION, consisting of two levels comparing the effects of the two presentation conditions (i.e., 100 ms or 500 ms presentation times for the first stimulus). There were three within-subjects factors: MATCH vs. FOIL (2 levels), BLOCK (4 levels) and ROTATION (2 levels).

The first of these factors compared the responses of *matching trials* (where the second shape is the same as the first) to *foil trials* (where the second shape is different from the first). The second of these factors, BLOCK, compared response times across different blocks in the experiment. In the ANOVAs on practice trials, this is used to test for, among other things, practice and fatigue effects. For the ANOVAs on the test blocks and the train blocks which precede them, the BLOCK effect was also used to test for the multiple representation + normalization hypothesis. This is because the test orientations were all $30^0$ further away from the upright than the corresponding orientations in the practice blocks. Therefore, if subjects were recognizing the test versions by normalizing from multiple representations formed in the practice trials, there should be an overall increase in response latencies in the test block consistent with a transformation through $30^0$. In addition, the choice of angular deviations from the upright in the train and test blocks, as outlined below, still permitted testing for conventional normalization from the upright in both types of blocks.

The third factor in the ANOVAs, ROTATION, compared the effects of angular deviations from the upright of the stimuli. For this analysis, the different rotational deviations from the upright were combined into two levels. According to the normalization hypothesis, an image is always normalized by traveling the shortest rotational path, regardless of direction. This ensures that the angular deviation traveled is never larger than $180^0$ from the upright. Therefore, the time taken to recognize an image oriented $x^0$ from the upright would be the same as that for an image at $-x^0$.

On this basis, the reaction times for all training trials (match and foil) using rotations of $60^0$ and $300^0$ were combined to form a Level 1 reaction time while those for $120^0$ and $240^0$ were condensed into a Level 2 reaction time. For testing trials, the rotations of $90^0$ and $270^0$ were combined to form the Level 1 reaction time while those for $150^0$ and $210^0$ were similarly combined to form the Level 2 reaction time. In both training and testing, the difference between the two levels was always equivalent to a $60^0$ rotation under the normalization hypothesis. Thus, if subjects used this strategy, this $60^0$ difference would be reflected in similar differences in reaction times (i.e., the same size effect of ROTATION) regardless of whether it is a practice or test block. Therefore, the design allowed us to determine whether subjects normalized to the upright continuously throughout the experiment or whether training invoked a shift towards a different strategy. In addition, it allowed us to test whether the new strategy involved multiple representations or orientation invariant recognition.

An omnibus ANOVA was conducted on response time. The effects of practice varied between match and foil trials as evidenced by a highly significant BLOCK by TYPE interaction ($F(2.59,41.42) = 6.70, p < .01$); $W(5) = 4.77, p < .10$). As can be seen in figure 6.4, subjects took longer to respond to match trials only in the first block. At least at the start of the experiment, the rate of decrease in latencies for match trials was greater than that for foil trials and by the final two practice blocks, times for match trials were always shorter. A similar omnibus ANOVA was conducted using accuracy as the dependent measure. There was a moderate effect of TYPE, indicating that response choice was 2% more accurate on foil trials than on match trials ($F(1,16) = 5.07, p < .05$). The remaining effects with respect to response

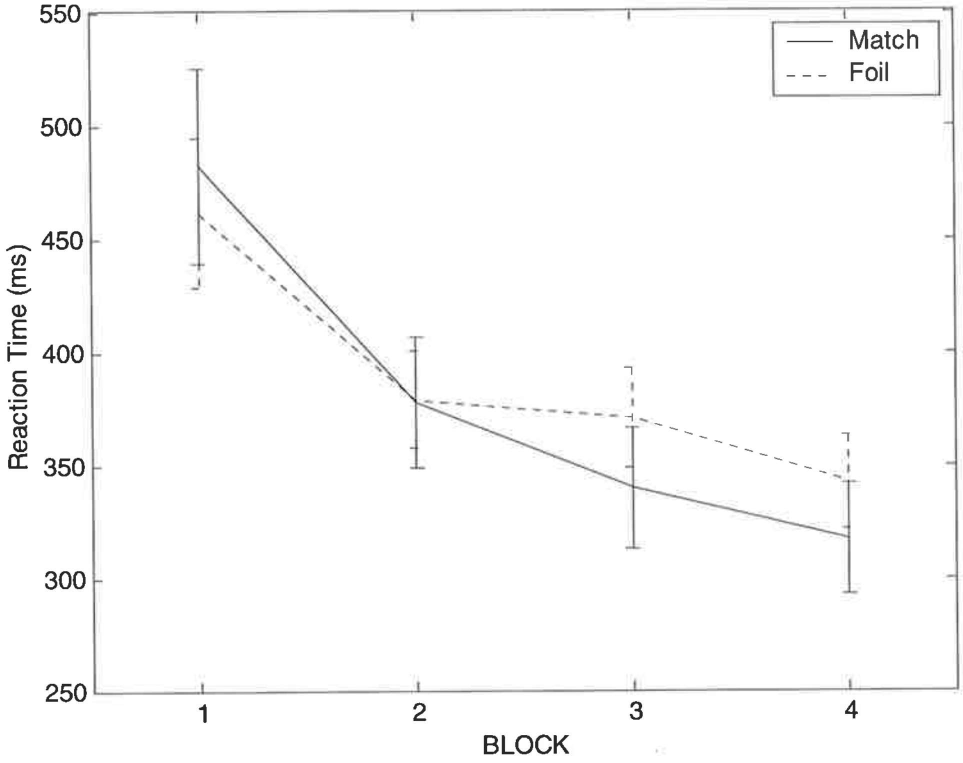
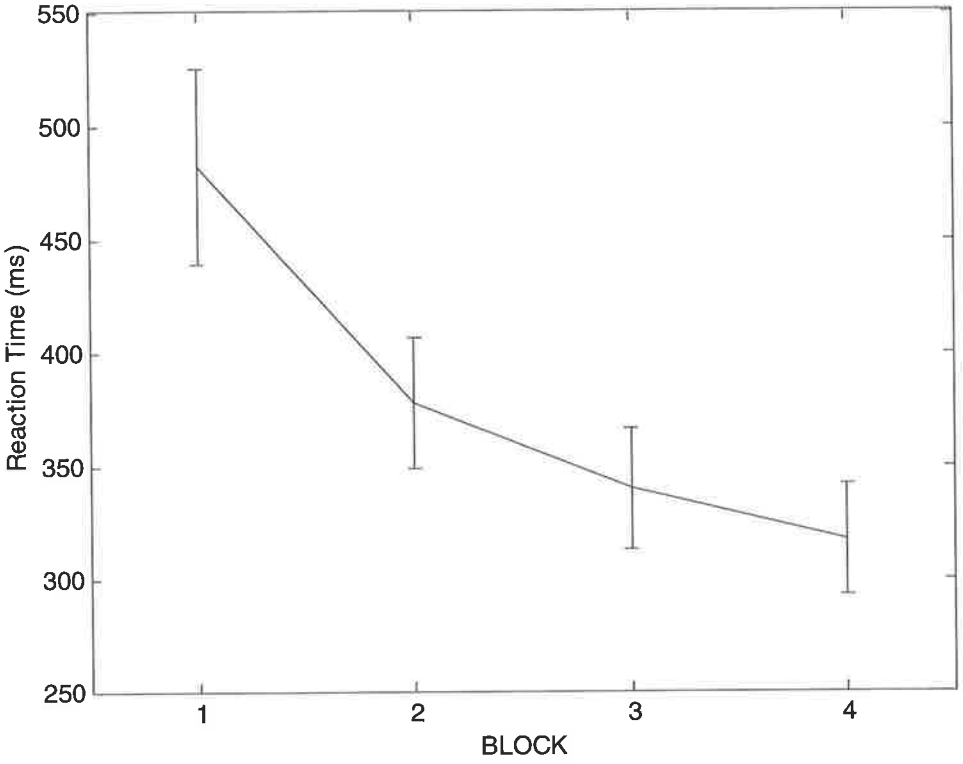


*Figure 6.4. Reaction times for match and foil trials across the practice blocks. The error bars represent the mean +/- one standard error.*

accuracy and reaction time are best illustrated via individual ANOVAs for match and foil trials.

### *6.3.1.1 Match trials*

The ANOVA on response latencies associated with correct responses to match trials revealed a highly significant BLOCK effect (F (3,48) = 22.79, p < .01) suggesting a decrease in response times as the experiment progressed. A trend analysis revealed highly significant linear (F (1,16) = 25.94, p < .001) and quadratic components (F (1.16) = 17.45, p < .01). This pattern is visible in figure 6.5. While the



*<u>Figure 6.5. Reaction times across practice blocks for matching trials only.</u> The error bars represent the mean +/- one standard error.*

trend is monotonic, with response times always decreasing from one block to the next, it is not simply linear and the rate of decrease diminishes as the experiment progresses.

There was a close to moderate sized effect of ROTATION (F (1,16) = 4.34, p < .10). The interaction between ROTATION and BLOCK was not significant. The effects of ROTATION across the different blocks are depicted in figure 6.6 in which the y-axis depicts the difference between the mean reaction times from the larger (rotation level 2) and smaller (rotation level 1) angular deviation conditions. In this

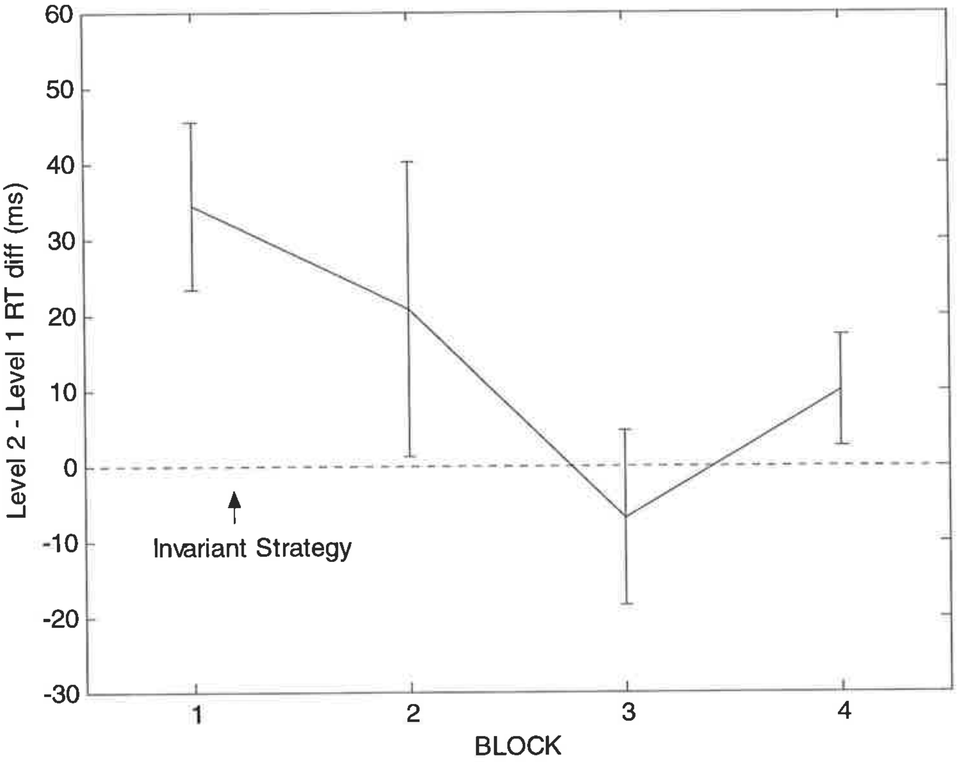


*Figure 6.6. The difference between reaction times to the two levels of rotation across the four practice blocks of Experiment I for match trials. Rotation level 1 represents "smaller" angular deviations of $60^0$, $300^0$, $90^0$, $270^0$ while rotation level 2 represents "larger" angular deviations of $120^0$, $240^0$, $150^0$, $210^0$. The dotted line depicts the hypothetical invariant recognition strategy while the error bars represent the experimental mean +/- one standard error.*

figure, and in many other figures in this chapter, the dotted line represents a hypothetical orientation invariant strategy for comparison with the actual data. Although there was no significant ROTATION by BLOCK interaction, this figure demonstrates that the effects of rotation appear strongest at the first block only. This interpretation is supported by planned comparisons on the effects of ROTATION in each block in which the only significant result was a highly significant effect of ROTATION in the first block (F (1,17) = 9.64, p < .01). The direction of this difference was consistent with a normalization hypothesis (i.e., a mean difference of

34.28 ms). The *hypothetical rate of normalization* - the rate at which subjects would be normalizing one or other of the images to achieve the response times that they did - was 1751 deg/s. In the corresponding repeated measures ANOVA examining the dependent variable of accuracy, there were no significant effects even at the .10 alpha level.

There was no significant CONDITION by ROTATION interaction in the ANOVA and this was further explored by examination of the relevant means and standard errors. As can be seen in figure 6.7, there was no meaningful pattern in the effects of ROTATION between the two conditions across the different blocks. The overall mean difference between the Level 2 – Level 1 rotation score in the two conditions was only 1.07 ms.

#### *6.3.1.2 Foil Trials*

For the foil trials, the response times for correct responses and accuracy measures were similarly examined in separate repeated measures ANOVAs. In the ANOVA on response latencies, there was a highly significant effect of BLOCK (F (3,48) = 15.75, $p < .001$). As can be seen in figure 6.8, response times generally improve with practice but the rate of improvement seems to vary inconsistently as the experiment progresses. This interpretation is supported by the trend analysis. As with the match trials, there was a highly significant linear component (F (1,16) = 18.61, $p < .001$) and a moderately significant quadratic (F (1,16) = 8.34, $p < .05$). Yet in the case of these foil trials, there was also an additional moderate cubic component describing the added nonlinearity and complexity in the trend (F (1,16) = 8.62, $p < .05$).

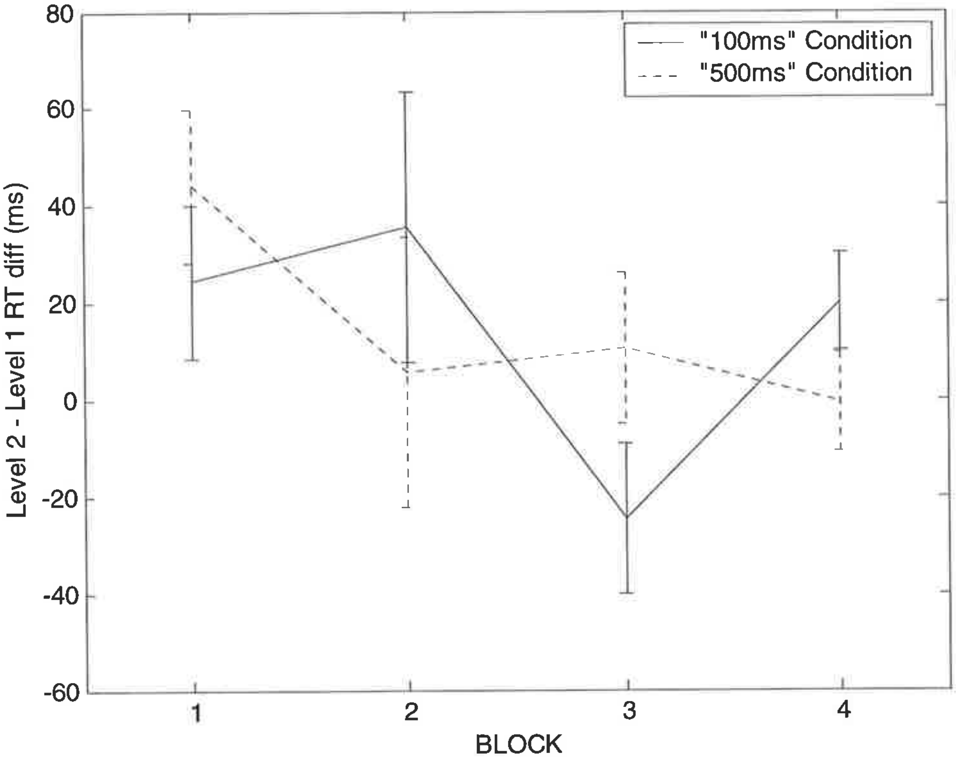


*Figure 6.7. The difference between reaction times to the two levels of rotation across the four practice blocks of Experiment I for match trials split by condition. Rotation level 1 represents "smaller" angular deviations of 60$^0$, 300$^0$, 90$^0$, 270$^0$ while rotation level 2 represents "larger" angular deviations of 120$^0$, 240$^0$, 150$^0$, 210$^0$. The error bars represent the mean +/- one standard error.*

Unlike the match trials, the effect of ROTATION was not significant but there was a comparatively smaller BLOCK by ROTATION interaction that was only marginally significant (F (3,48) = 2.44, p < .10). To examine this effect more closely, planned comparisons were conducted on the effects of ROTATION in each block. The only significant finding was a moderate effect of ROTATION in the second block (F (1,17) = 6.74, p < .05). The direction of the effect was consistent with normalization and the *hypothetical rate of normalization* was 2465 deg/s. These

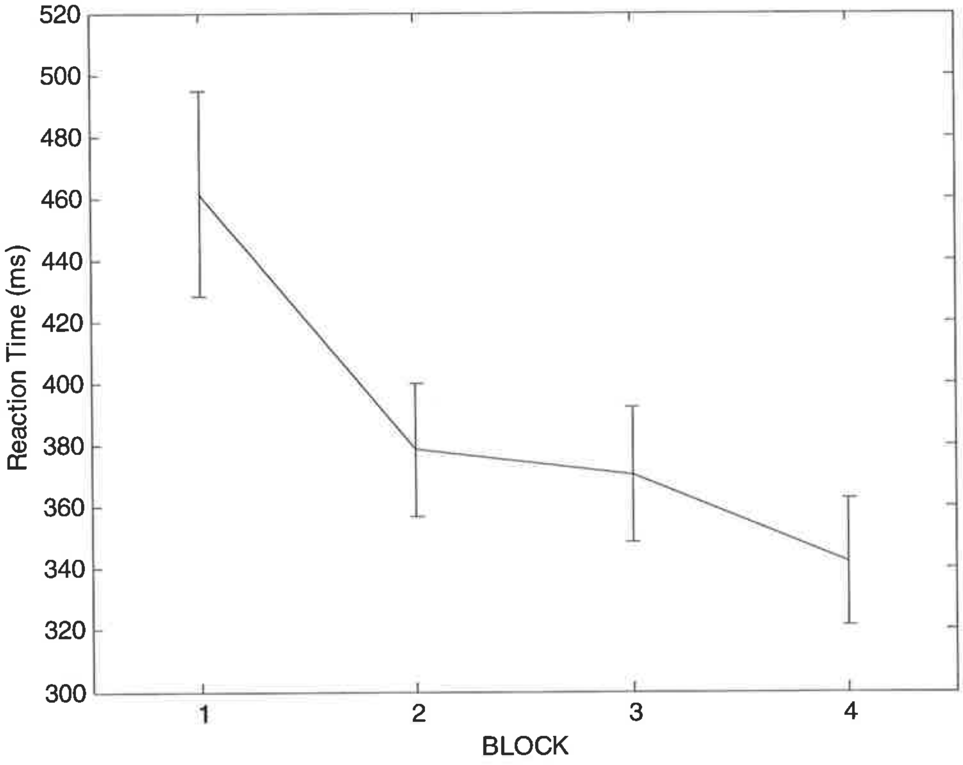


*Figure 6.8. Reaction times across the practice blocks for foil trials. The error bars represent the mean +/- one standard error.*

effects are visible in figure 6.9 and this chart also demonstrates the rather complicated pattern described by the BLOCK by ROTATION interaction.

An ANOVA on accuracy of response for these foil trials yielded only effects associated with CONDITION and these were mostly marginal. Overall, there was a marginal main effect of CONDITION (F (1,16) = 3.42, p < .10) and a marginal interaction between CONDITION and BLOCK (F (3,48) = 2.34, p < .10). The chart in figure 6.10 demonstrates the trend of ROTATION effects over the different blocks. Subjects in the longer presentation time condition were more accurate than those in the shorter presentation time condition for the first half of the experiment. However,

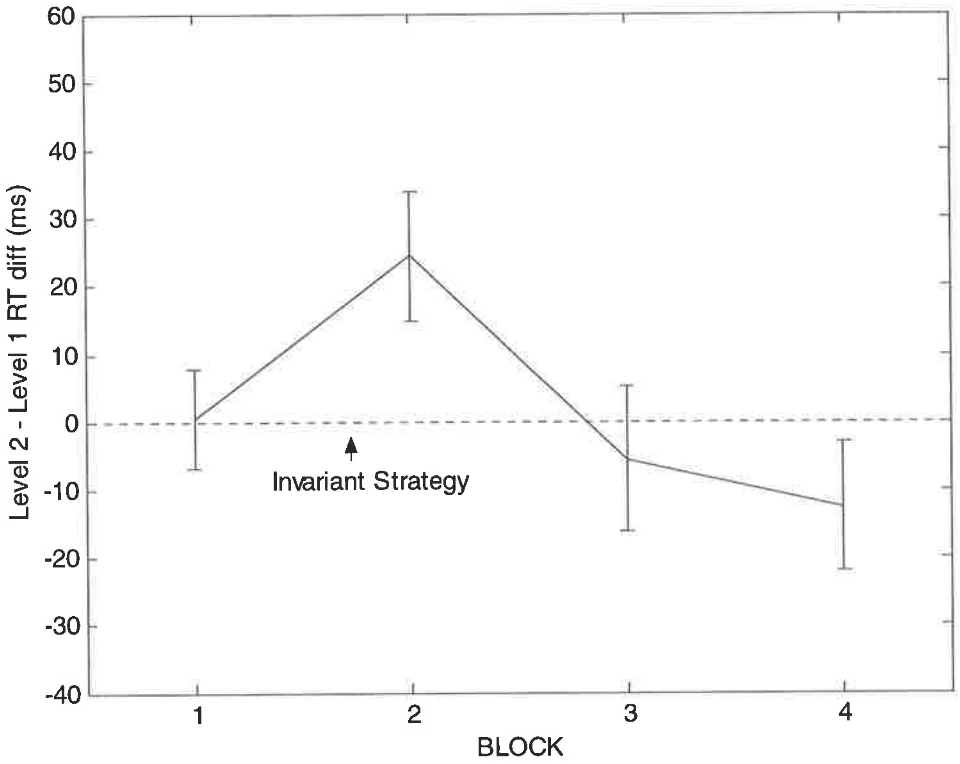


*Figure 6.9. The difference between reaction times to the two levels of rotation across the four practice blocks of Experiment I for foil trials. Rotation level 1 represents "smaller" angular deviations of $60^0$, $300^0$, $90^0$, $270^0$ while rotation level 2 represents "larger" angular deviations of $120^0$, $240^0$, $150^0$, $210^0$. The dotted line depicts the hypothetical invariant recognition strategy while the error bars represent the experimental mean +/- one standard error.*

in the second half of the experiment, the response accuracy measures for the two different conditions appear to converge and there is no advantage in either condition. However, all but one of these effects in response accuracy were marginally significant and none of the CONDITION effects represented a mean difference of more than 3%.

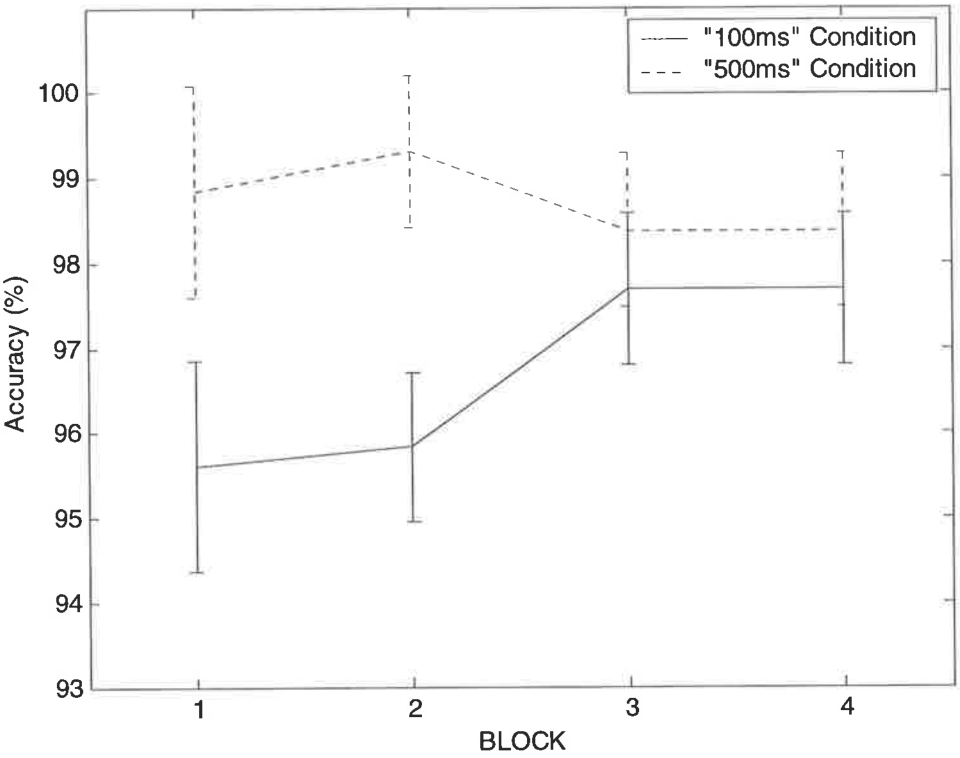


*Figure 6.10. Response accuracy to foil trials for the two conditions across the practice blocks. The error bars represent the mean +/- one standard error.*

### 6.3.2 Train vs. Test trials

As with the practice trials, the data from the individual blocks were combined to allow us to test for the presence of multiple representations. Blocks 7 and 8 from the raw data were combined to form *Train block 1* while 9 and 10 were combined to form *Test block 1*. Similarly, Blocks 17, 18 and 19, 20 were combined to form *Train block 2* and *Test block 2* respectively. This ensures that analysis is based on at least two measurements of each shape and orientation combination to provide a more reliable measure of the subjects' performance. The means and standard errors for the

*Test vs. Train 1* and *Test vs. Train 2* comparisons are contained in appendices C and D respectively.

In neither of the two *Test vs. Train* ANOVAs on response times for match trials was there any significant increase in latencies in the *Test* block in comparison to the previous *Train* block, (i.e., no significant BLOCK effect). In fact, in the *Train vs. Test 1* ANOVA, there was a highly significant *decrease* of 32.80 ms (SE = 10.13) in response latencies (F (1,16) = 10.48, p < .01). The confidence interval for the population mean difference, adjusted for multiple comparisons, was $\{-54.28 < u_d < -11.32\}$. In other words, we can be (at least) 95 % confident that there was no mean increase in response times for the first *Test* block. In the *Train vs. Test 2* ANOVA, there was a non-significant mean increase in response time of 8.84 ms (SE = 13.62). Assuming a multiple representation + normalization strategy, the *hypothetical rate of normalization* based on this mean difference would be very high - 3393 deg/s. However, the 95% confidence interval for the mean difference was very broad at $\{-20.03 < u_d < 37.70\}$ and subsumed population means for no difference in response time. But it also includes the population mean of 17.13 ms which is consistent with a *hypothetical rate of normalization* for multiple representations of 1751 deg/s - the rate found at the start of the experiment presumably associated with normalization from a single upright version.

The only other significant effects in response time were BLOCK by ROTATION (F (1,16) = 3.17) and BLOCK by ROTATION by CONDITION (F (1,16) = 3.16) effects in the *Test vs. Train 2* ANOVA that were only just marginally significant (p < .10) and uninterpretable. Similarly, there were two inexplicable interactions in response accuracy for match trials in ANOVAs comparing *Train* and *Test* trials. The first was a marginal interaction effect of BLOCK by CONDITION in *Train vs. Test 1* ANOVA (F (1,16) = 3.31, p < .10). The second was a moderately significant ROTATION by CONDITION interaction in the *Test vs. Train 2* ANOVA (F (1,16) = 6.84, p < .05). However, none of the mean differences between different rotation levels in the same condition or the same rotation level under the two conditions amounted to more than a 2% change in response accuracy.

Similar analyses were also carried out on responses to foil trials to examine differences in *Train* and *Test* blocks. As with the match trials, there were no significant increases in response times in either *Test* block. For the *Train vs. Test 1* ANOVA on response latencies, there was a highly significant BLOCK by ROTATION interaction (F (1,16) = 11.93, p < .01). This appears to be the result of a difference in rotation levels in the *Train 1* block, which equates to a relatively slow hypothetical rate of normalization of 1111 deg/s, that is not present in the *Test 1* block. However, this result may be spurious. Similarly, there was an inexplicable three-way interaction between BLOCK, ROTATION and CONDITION that was probably the result of noise (F (1,16) = 4.92, p < .05).

For the accuracy of responses to foil trials in the *Train vs. Test 1* ANOVA, there was a moderately sized interaction between BLOCK and ROTATION (F (1,16) = 6.52, p < .05). However, none of the differences between levels in the two blocks was greater than 3%. For the *Train vs. Test 2* analyses, the only significant effect occurred in the ANOVA on accuracy. This was a moderately sized interaction between ROTATION and CONDITION (F (1,16) = 6.84, p < .05). However, the mean difference in accuracy between both the same rotation level in the different conditions and different rotation levels in the same condition never amounted to more than 3%.

## 6.4 Discussion

### 6.4.1 The 'trend to invariance'

The predicted 'trend to invariance' hypothesis was supported by this study. The response time data was consistent with the notion that, on match trials, the stimuli were initially recognized by a process of normalization. Yet with practice, the normalization procedure was replaced by a method of recognition invariant to the orientation of the stimulus image. This 'trend to invariance' replicates the general findings of previous literature as outlined in section 5.7. This is the only experiment to demonstrate this pattern using random (or, strictly speaking, quasi-random) shapes. The orientation-invariant effect was apparent for images that were not purposefully constructed and were free from obvious semantic associations. It is also the only experiment to demonstrate this effect using a successive pair matching paradigm. In the broader context of the experimental literature, this experiment demonstrates that the 'trend to invariance' effect generalizes to different stimuli and experimental contexts.

While the analysis demonstrated minor effects of angular deviation in the foil trials, we believe that these are better attributed to measurement noise than to a normalization procedure. In section 5.6.1 we discussed the use of strategy on foil trials in previous literature and on this basis we predicted that there would be no effects of orientation on foil trials in our task. Normalization is usually only apparent when the foil images are the same objective shapes as, or are otherwise highly confusable with, the match images. Neither of these two factors were present in our task. It is possible that normalization was used occasionally on distractor trials in this experiment. However, the rotation effects in foil trials are more complex than simply normalization or a 'trend to invariance'. The variation in the effects of rotation was marginal, complex and noisy. There was neither an overall effect of normalization nor a clear trend to invariance to parallel the results from match trials. In fact, the only individual block that demonstrated a significant normalization pattern was, oddly, the second block where the mean difference was comparatively small and the *hypothetical rate of normalization* extremely fast (2465 deg/s). Such an unusual pattern suggests, not the operation of a recognition mechanism but rather the interference of measurement noise. The question of how foil images are processed will be further addressed in the second experiment.

### 6.4.2 Multiple Representation + Normalization Hypothesis

We attribute the 'trend to invariance' in this experiment to be the product of an orientation invariant recognition strategy. There was no evidence of normalization from representations formed within the experiment (i.e., the multiple representation + normalization hypothesis suggested by Tarr and Pinker's (1989) data). As mentioned previously, this theory predicts that subjects would form several copies of the root image at the different orientations presented to them during the practice blocks of the experiment. These multiple images could later be utilized as templates that could be rotated in their own right to match the same shapes at new, unseen orientations. Yet, while neither the significance tests nor the mean differences supported this notion, the confidence interval for the second test of this theory suggested that there was a lack of statistical power. In other words, the test was not rigorous enough, and the failure to find any evidence for the multiple representation + normalization account is therefore less convincing.

One of the causes of this lack of power was the noise in the data. This was demonstrated by the number of inexplicable interaction effects in reaction time and accuracy among the variables examined in the same ANOVAs. But there are two reasons to be confident in the orientation invariant hypothesis in this instance. Firstly, the failure to find evidence for the multiple representation + normalization account in the first test block *was* convincing (i.e., post hoc, this test *was* adequately powerful). A priori, it is implausible for both orientation invariance *and* multiple representations to be utilized in the same task and there is no precedent in the literature for such an occurrence. Secondly, there was no evidence for the multiple representation account in the raw data. In summary, there is no evidence for the multiple representation account in this experiment although a more rigorous test is necessary to reject it more convincingly in this context. To examine the issue further, tests for this hypothesis will also be included in the second experiment and both the statistical power and the measurement consistency will be improved.

### 6.4.3 The effects of stimulus duration

While there were some effects of the different experimental conditions, none of these was of the type predicted – namely, that normalization would be more strongly associated with the shorter presentation time of 100 ms on match trials. In other words, the evidence of normalization and the decrease of this effect with practice did not vary between the two groups. Varying the presentation duration did affect subjects' performance, but not in this particular manner. For example, there was an effect of condition for the foil trials in the practice blocks of the experiment. It appears that, with the luxury of the longer presentation time, there was a short-term benefit in the accuracy of response. However, by the second half of the experiment this advantage disappeared. This suggests that a longer presentation time improved accuracy in the initial stages of the experiment but that practice alleviated the difference between how subjects responded under the two conditions.

The only evidence, albeit unconvincing, that the actual normalization pattern may have differed between conditions was found in the final train block. However, not only did this involve a stronger trend in normalization under the longer 500 ms condition (i.e., the opposite to the predicted pattern) but it was associated only with foil trials. This finding is one of many interactions uncovered in the *Train vs. Test* ANOVAs – the vast majority of which were inexplicable and small in magnitude. Such findings are best attributed to noise. Thus, there was no convincing evidence that altering the presentation time affected the use of normalization or, more specifically, that longer times were associated with a stronger trend towards invariant recognition.

There are two possible explanations for the failure to find the predicted effects of stimulus duration on orientation sensitivity. Firstly, the two conditions may not have been appropriate to invoke different processing styles. For example, the difference of 400 ms in the presentation durations may not have been large enough.

Greater differences may be necessary to access different representational codes in this paradigm. In addition to stimulus duration, backward pattern masking, as used in both Lawson and Jolicoeur's (1998) and Jolicoeur and Landau's (1984) experiments, may also be necessary to invoke resilient rotation effects. The second reason why the experiment may have failed to detect evidence of different representational codes is insufficient statistical power – the experiment may have lacked adequate power to detect such differences. The two groups consisted of 9 subjects each and a larger N may be necessary to find the predicted differences, especially if the effect size is small.

# CHAPTER 7: The effects of size

*Studies investigating the effects of size have demonstrated some important similarities with those examining the effects of orientation. Both normalization for size and size invariant recognition have been demonstrated. There is also some evidence that size insensitive processing is influenced by similar factors to orientation-insensitive processing – namely practice, stimulus complexity, timing variables, and the nature of foil images. Such similarities may indicate fundamental similarities in processing invariant to these two transformations and may even reflect the operation of a single orientation and size invariant recognition mechanism. The lack of evidence for other size effects or phenomena, that might parallel those for orientation, may be due to the relatively smaller number of such studies investigating scale effects and the failure to replicate comparable experimental conditions. However, there is evidence for effects unique to scale changes that are associated with early, pre-recognition stages of visual processing which can also influence performance on recognition tasks.*

## 7.1 Introduction

The effect of object size on recognition has received far less attention than the influence of orientation. However, the significance of such research is no less important to our understanding of human visual perception. Such a bias may be the result of an intuitive belief that size does not have a significant effect on recognition ability – as long as the image is not so small that its detail cannot be ascertained or so big that its boundaries are outside of the visual field, we can identify the shape. A similar view has been espoused by Rock (1973, 1974) who claimed that, while orientation has a significant impact on our perception and recognition of shape, size is irrelevant to identification. Certainly, there are some fundamental differences in the effects of the two variables, and we will examine these in the latter part of this chapter. Yet both transformations are independent of objective shape and the ability to extract information regarding both orientation and size is fundamental to interacting with the environment. In fact, there are also some important similarities in the experimental literature between the effects of these two variables in recognition tasks.

This chapter is similar in format to chapter 5 which investigated the effects of orientation. Firstly, we will discuss normalization of size and present evidence for this process from recognition tasks and other tasks involving mental representation. This will include a discussion of *size constancy* – the ability to perceive the real world size of an object despite changes in the size of the image that projects on our retina. Following this we will summarize some neurophysiological and neuropsychological findings on how size information is processed during recognition. The next section of the chapter will look at evidence of size-invariant processing in recognition tasks, especially in priming experiments. The format of this part of the chapter is similar to the chapter on rotation with corresponding sections on the influence of paradigm, timing, training and familiarity, stimulus complexity and same shape foils. In contrast, the final part of the chapter discusses important differences between the effects of rotation and size. This includes a description of the *zoom lens theory* of attentional deployment and how this theory can account for several findings in the literature that are normally attributed to normalization.

## 7.2 Normalization of size

What role does size play in the mental representation of shape? If there is a representation level at which size is a quintessential element of an image's representation, then a normalization process, similar in principle to that used in the identification of rotated objects, may be necessary to align the image with the representation in memory. Such a normalization operation would be demonstrated by an increase in response times and / or errors for matching or identifying images that is monotonically related to the magnitude of the size disparity. Accordingly, there is no lack of experimental evidence supporting the normalization hypothesis for size differences, with just such a pattern of results for matching responses (Besner, 1983; Besner & Coltheart, 1975, 1976; Bundesen & Larsen, 1975; Cave & Kosslyn, 1989; Corcoran & Besner, 1975; Howard & Kerst, 1978; Jolicoeur, 1987; Jolicoeur & Besner, 1987; Larsen & Bundesen, 1978; Larsen, McIlhagga & Bundesen, 1999; Milliken & Jolicoeur, 1992; Posner & Mitchell, 1967; Santee & Egeth, 1980; Yonelinas & Jacoby, 1995).

## 7.3 Tasks other than recognition demonstrating analogue size effects in an object's representation

In addition to conventional studies of identification and matching, other experimental paradigms have also demonstrated that size can be an important, if not quintessential, aspect of our memorial representation of shapes. It has also been shown that familiarity decreases with a 4:1 size discrepancy between study and test phases in a manner similar to that for overt identification responses for randomly generated geometric shapes (Yonelinas & Jacoby, 1995). Other more diverse paradigms have also exhibited an effect of size on response behaviours. When subjects are asked to compare the size of objects in memory the results suggest an imaginary perceptual comparison – reaction time is an inverse linear function of the log of the estimated difference in the objects' sizes (Moyer, 1973; Paivio, 1975). In addition, when the stimuli presented were pictures, the reaction time was shortened when the relative size disparity of the two pictures was in agreement with the size difference of the objects in real life (Paivio, 1975 [Experiment II]). The effect was also reversed when the task was to decide which image was further away (Paivio, 1975 [Experiment III]).

These findings support the notion that the comparisons made by the subjects on these tasks are based on "visual analog representations in long-term memory" (Paivio, 1975, p.645). In this case, size is presumably an integral part of the image stored in memory. Interestingly, these results mirror those found in actual perceptual comparisons where the time to judge an image as larger is related to the magnitude of the physical difference (Welford, 1960; Woodsworth & Schlosberg, 1954). This suggests that our long-term representation of objects can possess a concrete or pictorial quality not unlike that of direct perception, at least with respect to relative size information. On the other hand, Holyoak (1977) has demonstrated that internal size comparisons can occur without the need for imagery when subjects are instructed not to use such a method. In addition, the accompanying reaction time profile did not reflect real life size ratios. In other words, in addition to the more pictorial comparison process, there is evidence for perceptual / cognitive processes in which size is not stored in an analogue manner.

## 7.4 Size constancy

It is important at this point to distinguish between the perceptual phenomena of *size invariance* and *size constancy* (otherwise known as *size-distance invariance*). The former refers to the ability to identify an object independently of changes in its retinal size. The latter refers to the capacity to recognize the physical or real size of an object independently of viewing distance. For example, objects seen from a considerable distance often appear the same size when they are viewed from nearby. Over varying distances, apparent size remains remarkably constant (within sensory limitations) despite changes in the retinal size of the object's image (e.g., Brosgole, PlaHovinsak, Roig, & Notaro, 1985; Frank, 1925; Rock, 1975; Ross & Plug, 1998). In other words, the perceived size remains constant. The phenomenon of size constancy is described by the following equation:

$$\text{Perceived size} = \text{Visual angle} \times \text{distance}$$

Put simply, perceived size is a function both of the size of the retinal image and the judged distance of the object from the observer.

## 7.5 Perceived, not retinal, size may underlie size-normalization effects

It is the perceived size of an object that appears to be normalized in recognition experiments like those outlined in section 7.2. In the conventional recognition experiments (e.g., Jolicoeur, 1987; Kolers, Duchnicky, & Sandstroem, 1985), the contribution of perceived size and that of pure retinal size are confounded. This is because there is no change in the viewing distance – the only factor that can differentiate between the two scale measures because it affects retinal but not perceived size.

To tease apart the contribution of these two variables, Milliken and Jolicoeur (1992) dissociated the two factors in a recognition memory experiment involving study and test phases. Overall, the results suggested that perceived size, and not retinal size, was responsible for the size normalization effect in these experiments, and the authors propose that the same may also be true for other experiments demonstrating size sensitivity. Milliken and Jolicoeur (1992) conclude that the "perceived size of a shape is a functional part of its internal representation" (p.94). The object's internalized form or representation is not that of a strict template of the incoming image from the eye but one that involves size information in a way that takes into account other factors, i.e., viewing distance. In turn, this perceived size can be normalized to provide a match for the purposes of identification.

Similarly, Karwoski (1933) has argued that memory for size is not based on retinal or absolute information. Using a paradigm that involved judging the size of letters, memory of size information was evident only for relative size between different stimuli and not the individual object's absolute spatial scale. However, under reduced (e.g., monocular) viewing conditions there is some evidence for size-congruency based on retinal size (Brosgole, PlaHovinsak, Roig & Notaro, 1985; Holway and Boring, 1941). One possible explanation for these findings is the *equidistance tendency* (Gogel, 1965). According to this principle, subjects assume that all the objects are situated the same distance away and make appropriate decisions on this basis.

These findings provide an interesting parallel between the functional role of size and rotation effects. In reviewing the effects of rotation, we discussed the influence of a frame of reference for orientation (see section 5.2). The influence of orientation in perceptual tasks is not necessarily determined by retinal information alone. The canonical orientation that is ascribed to an object may not be the retinal upright but can also be influenced by the environmental upright. The influence of perceived size may reflect a similar frame of reference – it is not the size measured in purely retinal coordinates that is encoded in the object representation but also the size with respect to environmental considerations.

Thus, the object representations in long-term memory, demonstrated in Milliken and Jolicoeur's (1992) experiment, were presumably size-specific. However, the size was based not only on retinal scale at the time of encoding, but also on information regarding the distance of an object from the observer. Therefore, both the orientation and size of object representations in long-term memory appear to be based on a judgment that takes into account both retinal and environmental dimensions. This may be the result of a common frame of reference that specifies representational coordinates for shape representation according to judgments based on the subject's physical relationship to the objects in the visual scene, both in terms of relative orientation and distance.

Is there then any influence of purely retinal size on shape recognition tasks? We believe that there is, but that this reflects an attentional mechanism that involves stimulus encoding. The influence of size at this level may be independent of any influence of size in the higher level of representation used for recognition. We will discuss this mechanism in more detail in section 7.13, because it represents a fundamental difference between the effects of orientation and scale on perceptual tasks.

## 7.6 Some neurophysiological considerations

### 7.6.1 The neuropsychology of size

As with orientation, there is neuropsychological evidence, presented in Ashbridge and Perrett's (1998) review, that size may be treated in a qualitatively different manner by different systems in the brain. Again, as with planar rotated images, there is some support for a ventral / dorsal distinction, although the full picture of results is more complicated. Milner et al. (1991) give the example of a patient with damage to her ventral system (areas 18, 19 and the parasaggital occipitoparietal region). Despite the damage, she could still utilize size information in reaching and grabbing for objects. Yet, on a task involving the matching of simultaneously presented stimuli (Goodale, Milner, Jacobson, & Carey, 1991), this subject was severely impaired and could only indicate the size with her hand by reaching to grasp an object.

Investigations into the effects of various brain lesions in monkeys have similarly reinforced the general dorsal / ventral distinction. For example, IT lesions in monkeys trained on discrimination tasks, did not inhibit generalization across size compared to normals despite an increase in time necessary to learn the original discriminations (Holmes & Gross, 1984). As with orientation-specificity, the AIT (anterior inferotemporal) cortex has been implicated in size-invariant encoding (Weiskrantz & Saunders, 1984). However, size-specificity was associated with area V4 and PIT (posterior inferotemporal) cortex. Such findings suggest that, even though size-invariant processing is associated with the ventral stream, this stream is still involved in the analysis of size information.

A study by Larsen, Bundesen, Kyllingsbæk, Paulson and Law (2000) has implicated the role of the dorsal stream in size normalization on a recognition task. In their task, subjects were asked to provide match responses whenever any two images of a sequence depicted the same shape in one of three conditions. In the first condition, all the images were small while in the second, all the images were large. In the third condition, the stimuli alternated between small and large. Consistent with a normalization procedure, RTs were longer in the third condition. In addition, positron emission tomography (PET) implicated areas in the dorsal pathway (e.g., the posterior parietal cortex and the occipital, parietal and temporal transition zone) only in the third condition, i.e., when the images were different sizes and presumably normalized to match. In other words, there is evidence that normalization for size in a matching task involves the dorsal stream.

### 7.6.2 Single cell studies and size

Ashbridge and Perrett (1998) have suggested that, with respect to single-cell response selectivity, there is evidence for size invariant processing within the IT cortex. Citing studies by Ito, Tamura, Fujita and Tanaka (1995) and Lueschow, Miller and Desimone (1994), these authors argue that approximately 20% of cells sensitive to features of complex objects generalize over size changes in the temporal cortex (see also Tanaka, Saito, Fukada, & Moriya, 1991). However, it should be noted that, in addition to such size-insensitive cells, there is a larger number of cells *selective* for size in these areas. As suggested by Lueschow et al. (1994), there may be different systems or populations of cells that are either sensitive or invariant to the quality of size within the IT cortex itself. Such a finding suggests that, although a useful approximation, the ventral / dorsal distinction of "what" and "where" processing styles may be an over-simplification. Ashbridge and Perrett (1998) have argued that the stimulus dimensions of size and orientation are analyzed in both visual streams although the information is used in different ways.

## 7.7 Size-invariant processing and the priming debate

The priming paradigm has proved successful in demonstrating size-invariant effects in object naming. Using this methodology, size changes are not systematically associated with changes in error rates or reaction times and this profile has been demonstrated both for grey-scale pictures of real-world objects (Fiser & Biederman, 1995) and simple line drawings of common objects or animals with readily available basic-level names (Biederman & Cooper, 1992). In the latter paper, the authors report two experiments with brief and successive presentations of the stimuli (50 and 100 ms) in two blocks. The first block, otherwise known as the priming block, consisted of the 24 root stimuli that could be presented at one of two sizes. The second, primed block consisted of half of the images from the first block, some of which were shown at a different size from their first presentation.

The other half of the images in the second block consisted of objects that were semantically related to, and had the same basic level name as, the other 50% of the objects in the first block but were visually dissimilar. For example, instead of a depiction of a grand piano, subjects could be presented with an image of an upright piano. However, both images were to be named “piano”. These different shape exemplars were included in order to examine whether the priming was occurring at a visual or semantic level. In general terms, the results replicate the finding from other picture naming tasks that an image, viewed on one occasion (in the priming block), is more quickly and accurately perceived on a different occasion (in the primed block) (Bartram, 1974; Schacter, Dulaney, & Merikle, 1990). More importantly, there was no effect of size disparities on object naming between blocks on either accuracy or response latency measures. In addition, the mean RTs for different exemplars was significantly longer than that for same exemplar trials. Biederman and Cooper (1992) claim that this is evidence that the priming is visual and not semantic.

Biederman and Cooper (1992 [Experiment II]) also replicated the old / new discrimination task of Jolicoeur (1987 [Experiment Ic]) with the same presentation conditions and stimuli as used in their own priming experiments. Biederman and Cooper’s (1992) study differed from Jolicoeur’s (1987) by using shorter presentation times (100ms versus 6s and 4s), using only one presentation of each root stimulus in the learning block (versus 2) and using foils that are same-name different-exemplars (versus different pictures altogether). An important difference between this task and the priming experiments is that, instead of naming the object in the second block, subjects were asked to make "old/new" judgments – "old" if the object was one that was presented and named in the first block and "new" otherwise. In contrast to the priming experiments, this task produced a significant effect of size change on reaction time and error rate. In other words, the same stimuli appear to invoke size-specific or size-congruent effects – the latter occurring when the second presentation involves episodic discrimination and not explicit identification.

How can these contrasting results be reconciled? On the one hand, Jolicoeur (1987) himself has pointed out that the old / new discrimination task may not "address the problem of recognition per se, but rather address the problem of the recognition of episodic traces" (p. 541). On the other hand, questions have also been raised as to the relevance of Biederman and Cooper's (1992) findings. Ashbridge and Perrett (1998) claim that it is unclear whether the priming tasks actually accessed visual representations and not simply "semantic or naming codes that are size and view independent" (p. 198). A similar point was also raised by Jolicoeur and Humphrey (1998) when they commented that Biederman and Cooper's (1992) results may represent "differences in semantic priming (possible), purely differences in visual priming (unlikely), or some unknown mixture of the two (most likely)" (p. 94). In essence, the suggestion is that the size-invariance demonstrated in the priming tasks is not associated with a visual code but rather a higher-level name or semantic code.

But to what extent is this criticism of Biederman and Cooper's (1992) study valid? Prima facie, it is unclear how the influence of the visual processing stage can simply be bypassed even if the priming occurs only at the higher semantic level. As Biederman and Cooper (1992) have claimed, there is no evidence that the demands of naming "are not additive with the effects of perceptual variables" (p. 131). If upon first naming the picture of a 'piano' there is a priming of the name 'piano', how is it that a successive presentation of the same visual image, at a different size, cannot involve the visual representation system (size-specific or size-invariant), regardless of the facilitation of the name response at a higher level? Even if priming occurs at both semantic and visual levels, the sensitivity to the attribute of size in the visual analysis would still be reflected in the trend in response latencies and hence the degree of priming.

In fact, priming experiments involving images with parts missing *have* displayed convincing evidence of visual priming within a naming task (Biederman and Cooper, 1991b). In the first block, subjects were asked to name line drawings of common objects with a readily available basic-level name (e.g., cars and phones). All these images were incomplete – half of the component parts of the images had been removed. These parts are known as *geons* and the exact definition of these is given in section 4.3. In the second block, three types of images were presented – those that were *identical*, those that were *complementary* (i.e., that consisted of the remaining 50% of parts of the image in the first block which had been removed) and those that were *same-name different-exemplars*. In terms of semantic representations, it is arguable that there would be more similarity between *identical* and *complementary* images (which involve exactly the same semantic code) than between either *identical* and *complementary* and the *different-exemplar* images (where the *different-exemplar* images could access different semantic codes). However, the results do not reflect such a semantic account with shorter RTs and lower error rates for *identical* images but no significant differences between the response profile for c*omplementary* and *different-exemplar* images. Consequently, it seems reasonable to conclude that priming was visual and not semantic in this experiment.

In deference to Jolicoeur and Humphrey (1998), it may be questioned whether all the priming in Biederman and Cooper's (1992) experiments was visual. To determine whether priming was visual or semantic, *same-name different-exemplar* trials were employed and responses to these trials was compared to performance on trials using *same-exemplars*. According to Biederman and Cooper (1992), the fact that the *different-exemplar* trials were associated with longer RTs than the *same-exemplar* trials, both same or different size, demonstrated that visual priming was behind the similar response patterns for *same-* and *different-size same-exemplar* trials. The problem is that the *different-exemplars* do not necessarily share the same semantic representations as the root examples despite belonging to the same basic-level category. As an example, consider a grand piano and an upright piano – both labeled as "piano" in the task. The semantic representations associated with the two objects are not identical. As Jolicoeur and Humphrey (1998) point out, the two pianos are used in different contexts, are put to different uses and have different timbres. Clearly, if semantic priming did occur, it would not be equivalent in the way assumed by Biederman and Cooper (1992). But this constraint does not detract from the conclusion, reached by the authors, that the visual processing exhibited in this experiment was size-invariant. As mentioned before, it is difficult to envisage how the effects of the visual representation system can be entirely bypassed, regardless of where priming occurs.

## 7.8 Task differences

Perhaps the most parsimonious resolution of the debate over the significance of Biederman and Cooper's (1992) versus Jolicoeur's (1987) findings is that size sensitivity is a function of task requirements. As noted earlier, Tarr and Bülthoff (1995) have argued that there is no a priori reason to nominate any particular experimental paradigm as singularly representative of true 'recognition'. It may be that different representational codes or processes are accessed by the different tasks and that these vary in size sensitivity.

Further empirical support for this notion can be found in Santee and Egeth's (1980) study, where size sensitivity varies with task differences with the very same stimuli. Using a speeded classification paradigm, involving the sorting of cards into 'squares' and 'circles' primed at three different sizes, identification behaviour was not significantly influenced by size (Experiment I). The same stimuli, presented tachistoscopically, and where the stimuli required a naming response ("square" or "circle"), produced a similar pattern of results (Experiment II). In addition, there was no evidence of order effects that would suggest that subjects were normalizing internalized representations from the preceding trial in order to match the stimulus in the current trial. But changing to a simultaneous paradigm ("same" – "different" categorization task) *did* produce effects of size – namely, that the magnitude of size disparity between the two concurrently presented stimuli was linearly related to increasing reaction times.

Such variations in size sensitivity with task demands suggest both that size is not an inseparable component of an object's representation and that whether size adjustment is involved in the visual processing or not is influenced by what the perceptual system is trying to achieve. In other words, size sensitivity may be a function of task requirements and the perceptual system may have the ability to use more than one representational code for the same object, i.e., at least one that is size-specific and at least one that is size-independent

## **7.9 Timing**

Another factor, demonstrated by Santee and Egeth (1980) to influence size sensitivity, is the timing of experimental variables. In their third experiment, results were indicative of both size-invariant and size-sensitive recognition. The stimuli were those used in the previous experiments, but the paradigm was one of successive presentation requiring a match / foil response. Although the first stimulus was always presented for 200 ms, the interstimulus interval (ISI) could take a number of values (0, 100, 300, 500, 1000 and 2000 ms). Overall, there was an interaction between the effect of size disparity and ISIs such that the shorter the blank gap between stimuli, the more the response time profile reflected a normalization to match procedure.

At the 0 and 100 ms ISIs, there was a significant monotonic increase in mean RT associated with increasing size disparity, i.e., a pattern indicative of normalization.

However, a reliable normalization effect was not found at 300 and 500 ms ISIs although the mean time for the zero-disparity case appeared markedly lower. The faster responses to the same-size comparison sequences were, the authors argue, the result of a fast template or 'physical' match like that which occurs in the same-size sequence at 0 and 100 ms ISIs. But whenever a size disparity is involved, a size-invariant representation is employed (hence similar RTs for both levels of size-disparity). And at the longest ISIs (1000 and 2000 ms), there was no effect of size disparity at all, which reflected the consistent application of size-invariant recognition for all scale disparities. Thus, it seems that increasing the interval between stimuli to be matched invokes a qualitative change in the processing style applied. While brief ISIs were associated with normalization, there was a gradual trend towards size-invariant processing on both match and foil trials as the ISIs increase. In other words, there is evidence not only for size-specific and size-invariant representation styles between tasks but also within tasks.

The influence of the time difference between successive presentations on size sensitivity has also been noted by Cooper, Biederman and Hummel (1992) in a broader context. In simultaneous or sequential matching tasks (e.g., Bundesen & Larsen, 1975; Howard & Kerst, 1978), the time difference between first and second presentations is either small (less than 2 seconds) or zero. Such studies generally show some size sensitivity. However, in the priming experiments of Biederman and Cooper (1992, Experiment I & III), where size invariant processing is evident, the time between first and second presentations of same-shaped stimuli was between five and ten minutes. Therefore, Cooper et al. (1992) suggest that different visual representations are accessed as a function of time – only at longer intervals is the size-invariant code used in "object recognition" accessed (p. 198). For one thing, the use of long periods of time between presentations of the same shape "precludes the use of a short term visual buffer for stimulus matching" (Cooper et al., 1992, p. 198).

Cooper et al.'s (1992) post hoc observations, and the findings of Santee and Egeth (1980), demonstrate that qualitatively different representational codes or processes are available and that timing variables influence which of these codes or processes is utilized. Importantly, the influence of timing on size sensitivity parallels some of the findings with respect to orientation sensitivity outlined in the previous chapter (see section 5.14). For example, Santee and Egeth's (1980) claim that increasing ISI reduces the influence of each transformation in matching experiments parallels similar findings for orientation and size in Ellis et al. (1989). Such findings are consistent with the idea that higher representational codes are both size and orientation independent.

## 7.10 Training and familiarity

The imprint of size-invariant recognition has also been discovered in experiments where differently sized depictions of the same root stimuli are repeatedly presented. This paradigm is not unlike that with which Murray et al. (1993) have confirmed orientation invariant effects in recognition. Stefanova (1970), cited by Glezer et al. (1974), showed line drawings of common objects (e.g., mushrooms, scissors, teapot, cat, airplane) repeatedly and for brief durations tachistoscopically and asked subjects to name them. This process was repeated until subjects successfully named all objects. At this stage, subjects were tested with a larger set of images consisting of a) all the original images, b) the original images magnified (200%) and c) the original images reduced (25%). Stefanova (1970) found no difference both between responses to all the different versions in the testing sessions, and between these responses and those to the original versions in training. This suggests that the subjects had formed, not 'retinal' or 'concrete' representations involving specific size information, but size-invariant representations. As with the effects of orientation (e.g., Murray et al., 1993), the effects of size were attenuated with practice. Again, as with Murray et al.'s (1993) experiment, the effects cannot be associated with the formation of multiple representations. Such representations could not have been formed in Stefanova's (1970) study because the test phase presented the shapes at scales not previously utilized in the experiment.

In an experiment involving memory for size and not object recognition, a comparable attenuation in size effects was also discovered (Karwoski, 1933). In this task, the recollection of relative size differences reduced significantly with repeated testing in a series. The author concluded that "any condition which produces monotony or attention indifference tends to lessen memory for size" (Karwoski, 1933, p.539). Both these requirements are fulfilled in Stefanova's (1970) study – namely repeated presentations and responses involving knowledge of objective shape and not image size. The attenuation of size effects with familiarity appears similar to the comparable phenomenon with respect to orientation. In other words, there is evidence of a 'trend to invariance' for the attribute of scale similar to that for rotation outlined previously (see section 5.7). However, it is not known whether this reduction in scale effects is stimulus specific, as it is for the orientation effects demonstrated by

Jolicoeur (1985), or whether it generalizes to unfamiliar, unpracticed shapes (see section 5.7).

The same set of stimuli employed in Stefanova's (1970) study has also been used to demonstrate that object knowledge and image size knowledge are dissociable (Leushina, 1967; Stefanova, 1970 (both cited by Glezer et al., 1974)). In these experiments, again using the tachistoscopic presentation technique, the images were presented in four different sizes and subjects were asked to provide the name and estimate the size of the drawings. The results demonstrated that the ability to recognize and estimate size were independent – subjects could recognize the objects when size was estimated incorrectly and could recognize the size when the object was erroneously reported. In other words, these experiments demonstrate that size and shape evaluation can occur in separate channels. The authors argue that size is not involved in pattern recognition processing, i.e., identification is size invariant. But there is a caveat to such a conclusion from these results. The fact that the information on size may not be encoded or be consciously accessible in one form does not necessarily mean that it is not still utilized for identification in another form by the perceptual system.

## 7.11 Stimulus complexity

Kolers, Duchnicky and Sandstroem (1985) compared the identification of faces and words using a continuous recognition paradigm where subjects were asked to indicate whether or not they had seen an image before. While accuracy of response declined with increasing size disparities between first and second presentations for the recognition of faces, no such effect was found in the identification of words. The authors repeated the experiment with the variation that, instead of indicating only that a stimulus was presented for a second time, subjects were also asked to respond whether or not the second presentation was the same size as the first. The results in terms of recognition were largely similar to those for the first experiment. Meanwhile, the findings with regards to the encoding of size information ("same size" or "different size"), revealed that the initial scale of presentation of both faces and words was encoded and that the accuracy in retrieving this information did not differ significantly between the two stimulus types. As a result, even though information with respect to size was still encoded for words, it was not used for the purposes of recognition. This echoes the findings of Leushina (1967) and Stefanova (1970) that size and identity information can be processed separately – in other words, that recognition can proceed independently of size compensation.

One explanation for Kolers et al.'s (1985) findings is that size sensitivity in identification tasks is related to stimulus complexity – i.e., complex shapes, such as faces, are more likely to require normalization for size than simpler shapes, such as words. Note that a similar relationship was proposed in section 5.12 with respect to orientation sensitivity. In fact, Kolers et al.'s (1985) findings are remarkably similar to those of Rock (1973) and Thompson (1980), who investigated the effects of planar orientation on the recognition of faces – namely, a breakdown in performance associated with the degree of transformation applied to the face depiction. While most studies demonstrating transformation to match effects do so by means of response latencies, the effects of normalization in all these studies using face stimuli are expressed in error rates. In other words, the recognition procedure does not simply take longer at larger transformations but breaks down entirely. This parallel between the identification of the same type of image under both planar rotation and scale transformations suggests that there is more than a passing resemblance in the use of normalization and invariant recognition procedures between the two transformations.

## 7.12 Same shape foils and the necessity for normalization

### 7.12.1 Foil responses

As with studies involving the recognition of rotated shapes, there is an association between the type of foil used and the how the foil is processed. Several studies exhibit normalization effects for foil trials (Bundesen & Larsen, 1975; Howard & Kerst, 1978; Larsen & Bundesen, 1978). Interestingly, all these studies use foil images that are the same shape as the images, differing only by either a left-right mirror reflection (i.e., a MRF) or a $180^0$ rotation in the plane. In the majority of studies, when the matching or new/old discrimination tasks involve foils that are completely different shapes, latencies on the distractor trials do not reveal any normalization for size (Besner & Coltheart, 1975, 1976; Biederman & Cooper, 1992 [Experiment II]; Corcoran & Besner, 1975; Milliken & Jolicoeur, 1992 [Experiment I]; Santee & Egeth, 1980 [Experiments II & III]).

This post-hoc comparison between different studies has also received empirical support from Besner's (1983) study. Using line drawings of geometrical figures, the author compared performance on two matching tasks that were identical except for the type of foils. In Experiment I, distractor trials used images that were identical in shape to those used in the match trials except for a $180^0$ rotation in the plane. In Experiment II, however, the foils were entirely different shapes. The results supported the hypothesis that normalization was necessary only for the similar shape foils. Interestingly, there was still a slight effect of size in different shaped foil trials. We propose that this effect is in fact the product of an additional size-specific mechanism that is not directly related to object knowledge or recognition. We will discuss this effect in detail in section 7.14.

There is evidence that normalization is sometimes used for different shaped foils (Jolicoeur & Besner, 1987; Larsen, McIlhagga & Bundesen, 1999). In both these studies, same and different shapes are highly confusable. In the former study, all the images were closed polygons composed of only straight, horizontal and vertical line segments. In the latter, the root stimuli used were open forms (random polygons) constructed of between six and ten sides. However, the images used as foils were not independently generated shapes. The foil image was the same image as the match image except that it was subject to minor, random perturbations to all of the *x*, *y* coordinates. The resulting foil images were highly similar to the original shapes and the pairs were highly confusable. The apparent confusability of same and different stimuli was confirmed using binomial tests on error rates. This demonstrated a positive response bias with a significantly larger number of false alarms (responding “same” when the shapes were different) than misses (responding “different” when shapes were the same). Normalization may be required to differentiate between match and foil images when the two images are highly similar – either because they share the same objective shape or are otherwise highly confusable. A similar role for normalization for rotation was also discussed in chapter 5 (see sections 5.7 and 5.12).

Whatever role confusability may play in determining processing style, the overall pattern in the experimental literature is clear. Size-invariant processing for foil images is possible only when the images are entirely different shapes from those in match trials.

On the other hand, Jolicoeur and Besner (1987) have proposed an alternative explanation for the common finding of size invariant processing on foil trials. Results from their first experiment demonstrated a normalization profile for the foil latencies, with the exception of an abnormally long mean response time for the same-sized shapes (a scale ratio of 1:1) in comparison to the next smallest size difference (1:1.5). By using blocks in which only one size discrepancy is possible, the effect disappeared. The authors claim that this abnormal increase was the result of a bias caused by the "unusual added sameness evidence" of the similar scale images (Jolicoeur & Besner, 1987, p. 481). They suggest that the action of this bias, together with the small size ratios used in experiments antecedent to theirs (no larger than 1: 1.67), may have obscured the normalization profile in the data.

However, Jolicoeur's later study used a sufficiently large range of scale ratios (up to 1.1: 2) and yet failed to find any normalization effects in the foil data (Milliken & Jolicoeur, 1992). In fact, there is no evidence for, nor mention of, a response bias at the 1: 1 size ratio and the authors instead settle on an explanation for foil processing based on an exhaustive search procedure (p. 86). There is similarly no evidence or suggestion of such a bias in any comparable study involving the effects of size on identification. For example, in Besner and Coltheart's (1976) first experiment, where Jolicoeur and Besner (1987) suggest that the bias effect may have obscured the normalization pattern for foils, the mean response time for the same size condition was 86 ms *less* than that for the next smallest size ratio of 1: 1.6. This difference is incompatible with the operation of a “same” response bias (Jolicoeur and Besner’s (1987) study demonstrated a 56 ms *increase* for the same scale condition versus a comparable next smallest size ratio of 1: 1.6). Finally, the bias account does not explain the findings of Besner (1983) because his experiments found alternatively size-invariant and sensitive effects even though there was no change in the size ratios used (the only difference being the type of foil employed in the two tasks). In summary, it appears that the bias account may not always explain the failure to find normalization effects on foil trials.

### 7.12.2 Match responses

In addition to the influence of same shape foil images in the processing of foils, there is also evidence of a similar influence on match trials as well. That is, size-invariant processing is possible for the match trials only when the distractors are different shapes. All the studies in which there is evidence of size-invariant recognition can be characterized by the lack of similar shaped foils (i.e., foil shapes that were not rotated or mirror-reflected versions of the match shapes). In Santee and Egeth’s (1980) third experiment, the foils were entirely different shapes while in none of the other studies that also demonstrated an absence of size effects on performance in match trials, did the experiments use foil images at all (Biederman & Cooper, 1992 [Experiments I & III]; Farell & Pelli, 1993; Fiser & Biederman, 1995; Kolers, Duchnicky & Sandstroem, 1985; Stefanova, 1970). This suggests that the use of a size-invariant recognition mechanism for match trials may be possible only in the absence of same shape distractors.

There is an important similarity here between the findings from recognition studies involving either planar rotational or size symmetries. It appears that the presence of the same objective shape (i.e., forms differentiated from match images only by a shape preserving transformation) in match and foil trials, may deter the use of a recognition strategy that is independent of size or rotation. Such a finding is consistent with the hypothesis that the invariant strategy used in both experiments may be the same – namely, a general shape recognition mechanism that is simultaneously invariant to symmetries from the similarity group such as size, planar rotation and mirror-reflection. Many theorists have proposed just such a mechanism and we have discussed some of these models in chapter 4. The use of such a mechanism would be inhibited when the foils are the same shapes as those in the match images because it could not discriminate between the two types of trials. The system would consistently provide a "match" response to *both* match and foil trials and would therefore be inappropriate for the task at hand. Instead of picking out objective shape, the perceptual system would be required to use an alternative method (e.g., normalization) that could discriminate on the basis of shape irrelevant properties. To examine this hypothesis further, in the next chapter we will examine other examples where such a mechanism might be evident – namely, studies where the objects to be identified are both rotated and scaled.

## 7.13 Important differences between the effects of orientation and scale

Thus far, we have detailed some of the similarities between rotation and scale influences in matching and other recognition tasks. In this section, we will examine points of difference between the effects of these two transformations in the experimental literature. Firstly, we will look at some of the phenomena, associated with orientation effects, for which there is no evidence in the literature examining size. The failure to find these effects may be due to the lack of corresponding research investigating the effects of size. It may be that these size effects have not been found because paradigms or experimental constraints equivalent to those employed in orientation-effect studies, have not been used. Secondly, we will look at a mechanism related to retinal size – namely, evidence for an attentional mechanism for scale – for which there is no evidence for a similar process in terms of orientation and where the discrepancy is less likely to be due to the lack of comparable research.

The specific nature of orientation effects has received a good deal of attention, especially by Pierre Jolicoeur and colleagues (see chapter 5). In particular, the use of recognition strategies invariant to orientation has been linked to specific practiced objects and where this practice involves either transformations of the object in question or at least transformations of the objects associated with it. However, these findings relating to stimulus specificity and conditions associated with the attenuation of the transformational effects, have not been demonstrated with respect to size. Similarly, there has been no explicit examination of the presence of multiple, size-specific representations (although Stefanova's (1970) results are inconsistent with this notion). Are there fundamental similarities between the effects of rotation and size in recognition tasks? The lack of comparable findings may simply be because of the greater interest in, and larger number of published papers on, the effects of orientation in perceptual tasks involving recognition processes. As Kolers et al. (1985) claim, size "is not a well-studied influence on perception" (p. 726). It may be that comparable studies, investigating size effects under similar conditions used for orientation changes, may provide similar results. Further investigation is clearly needed to examine whether the two transformations yield similar results.

At the same time, there is convincing evidence of a fundamental difference in the effects of the two symmetries on perceptual processing. Unlike orientation, there is evidence for two fundamentally different size-specific mechanisms in the psychological literature (Cave & Kosslyn, 1989; Larsen & Bundesen, 1978). The first mechanism is *mental image transformation.* This involves the scaling of one of a pair of object representations in order to bring them into possible alignment followed by a test for identity between the two forms. This is the transformation of size that we have discussed up to this point in this chapter – namely, normalization for size for the purposes of shape recognition and matching. This is directly comparable to normalization for orientation discussed in chapter 5. Both these normalization strategies could be accounted for by a mechanism such as Huttenlocher and Ullman's (Huttenlocher & Ullman, 1987; Ullman, 1989) alignment procedure outlined in chapter 3.

However, there is also evidence for a second size-specific mechanism, known as *perceptual scale transformation* that acts in parallel with mental image scaling. Rather than manipulating the stimulus image in a direct manner, this involves an analogue transformation of “the unit of the perceptual reference system … in proportion with the changing stimulus size” (Larsen & Bundesen, 1978, p.2). It does not involve a shape matching process per se but is evident whenever two stimulus images of differing retinal sizes are compared in naming, matching or parsing tasks. Because it depends on gross visual area subsumed by an object and not object specific knowledge, Cave and Kosslyn (1989) claim that it precedes higher level processing and that it is a “selection mechanism that chooses inputs on the basis of size” (p.158).

The two transformation procedures are also distinguishable by their different time courses. Larsen and Bundesen (1978) found that, while the time taken for the image scaling procedure increases linearly with the linear size ratio difference, latencies associated with the perceptual scaling mechanism are proportional to the logarithm of the linear size ratio. Cave and Kosslyn (1989) propose that a linear trend fits the data equally well for the perceptual scale procedure but that the rate is much faster in comparison with mental image transformation. In summary, not only is there a qualitative difference in the two types of transformations (i.e., image versus scale), there is evidence of a difference in the functions of latency and size ratio between the two mechanisms.

Both Larsen and Bundesen (1978) and Cave and Kosslyn (1989) provide psychometric evidence for these two size-specific mechanisms. For example, Larsen and Bundesen (1978, Experiment II) found that, in letter discrimination tasks, reaction time was a logarithmic function of the linear size ratio between current and previous stimuli. In this experiment, consecutive stimuli never depicted the same letter. These results demonstrated the perceptual scale transformation phenomenon empirically and subjective reports supported this notion. Many subjects reported that, after the presentation of a letter at a certain size, they were "perceptually prepared for letters in the same format" (Larsen & Bundesen, 1978, p. 6). When the following letter appeared at a different size they had to "adjust themselves to that format in order to recognize the letter" (Larsen & Bundesen, 1978, p. 6).

In their third experiment however, same shapes could appear in consecutive trials. In such cases there was evidence of both perceptual scale and mental image transformation – the latter evidenced by a non-logarithmic RT function with respect to size ratio. Similar results were achieved by Cave and Kosslyn (1989). In their second experiment, for example, the time to identify the single target shape was a 'slow' function of the size when that stimulus was the target shape but was a 'fast' function when it was a different shape altogether. The use of two different mechanisms can also explain the findings of Besner (1983). In his first matching experiments, foils were MRFs and latencies for both trial types demonstrated a similar rate of change across size ratios. In contrast, Experiment II used entirely different shapes as foils. For match trials, the relationship between size ratio and response latencies was similar to the rates in Experiment I for both response types, presumably because all used an image specific transformation or alignment mechanism. However, the transformation rate for foils was significantly shallower, which is consistent with the use of the perceptual scale mechanism hypothesized by Cave and Kosslyn (1989) and Larsen and Bundesen (1978).

We suggest that the shape-specific mechanism of mental image transformation may reflect a normalization or alignment procedure that is similar to, or the same as, that which accounts for shape-specific orientation effects discussed in chapter 5. In other words, this would be a system or systems that tests for a match between forms via image transformation. In contrast, the perceptual scale mechanism may reflect a lower-level attentional mechanism that is not involved in recognition per se. As Biederman and Cooper (1992) claim:

> *What conceivable purpose would be served by referring new inputs, for the purposes of recognition, to some mean size – or any size for that matter – of the objects that we have recently experienced, especially when the size will almost never match the current size? (p. 131)*

We propose that the effects of the 'perceptual scale' may be attributable to an attentional mechanism that is scale-specific. Evidence for this phenomenon, known in other literature as the *zoom lens theory,* comes not only from Cave and Kosslyn's (1989) and Larsen and Bundesen's (1978) studies but also from other experiments on visual attention and neurophysiological mechanisms of perception (both empirically demonstrated and hypothetical).

## 7.14 Zoom lens theory

Biederman and Cooper (1992) cite two examples of research, demonstrating selective attention to global and local scales in visual processing, that can account for perceptual scaling (Martin, 1979; Pomerantz, 1983). Other later studies, not cited by Biederman and Cooper (1992), have developed such notions of attention to scale into what is known as the *zoom lens theory* (Eriksen & St. James, 1986; Eriksen & Yeh, 1985). This model proposes that attention is like a zoom lens on a camera that is adjusted to a particular scale or scope at any one time. There is no shortage of support in the experimental literature for this account and the theory has also received support from eye movement data (Zelinsky, Rao, Hayhoe & Ballard, 1997), as well as a single case neuropsychological study (Halligan & Marshall, 1993).

Shulman and Wilson (1987) have also demonstrated evidence for a zoom lens by employing spatial gratings. They found that identifying global features in a particular image facilitates subsequent responses to low-spatial-frequency gratings (indicative of large scale processing), while identifying local features in the same image facilitates responses to high-spatial-frequency gratings (indicative of small scale processing). These experiments demonstrate: 1) that attention can occur at different spatial scales; 2) that, at any one time, it is set to a specific scale; and 3) that it takes time to adjust the 'zoom lens' between different scales. This suggests that the zoom lens operates in an analogue fashion, much like adjusting the lens on a camera, i.e., to move between two different scales may require moving through the intermediate scales. All of these properties are shared with the so-called perceptual scale effect and we contend that they are one and the same mechanism. In other words, non-shape-specific scale effects, exhibited in the studies such as Cave and Kosslyn (1989) and Larsen and Bundesen (1978), are due to the zoom lens effect. This zoom lens is an attentional, scale-specific mechanism and not a recognition procedure per se. In other words, the effects of size attributable to the 'perceptual scale' may be the result of a time consuming adjustment of a scale-specific encoding mechanism.

Why would such an attentional mechanism exert an influence in object recognition experiments involving single objects? Why isn't the attended area simply set such that it encompasses the largest possible stimulus image used in the experiment? We propose that subjects constantly adjust the 'lens' to maximize performance on the task. One of the predictions of the zoom lens theory – namely, that performance is an inverse function of the size of the attended area – has been confirmed by Barriopedro and Botella (1998). Given this relationship, it would be a good, if not a necessary, strategy for Ss to adjust the attentional scale for each stimulus image in a recognition or matching task so that the attended area, which encloses the entire image, is as small as possible. This guarantees optimal perceptual resolution at any one time over the different retinal sizes presented. Eriksen and Yeh (1985) similarly suggested that the zoom lens always attempts to enclose the smallest possible area by trying to "contract and concentrate resources" (p. 595). Note that the smallest attended area will always be proportional to the retinal size of the image. As a result, the time it takes to adjust the scale in an analogue manner will always be proportional to the size difference between successive stimuli when, as is commonly the case in the experiments reviewed, subjects remain at approximately the same distance from the screen or projector throughout the experiment.

Another reason why subjects may optimize performance on matching or identification tasks by constantly adjusting the zoom lens is to control the amount of information associated with shape and structure in the image. The scale specificity of informational content of a retinal image is the product of the bottom-up nature of perceptual analysis. Obviously, we are not capable of perceiving stimulus images in infinite detail. It has long been acknowledged that the neurons in the visual cortex respond selectively to small regions of the entire visual field known as *receptive fields* (Hubel & Wiesel, 1979). Theories such as Marr's (Marr, 1982; Marr & Nishihara, 1978) which posit that visual representations, at least in the lower levels of perceptual analysis, can be derived in a largely data driven manner, presume that higher visual representations are also concocted in a similar manner by simple 'units' that provide summary information over neurons in the visual cortex. These 'primitives' or elementary units of a representation are also derived from larger receptive fields located across the entire visual area. Each primitive represents a summary of basic information about local surface orientation and distance, relative to the viewer, of the object from its own particular area of the visual field. It is the size of these primitives, i.e., the relative size of the area in the visual field from which it is derived, which "acts as a very strong determiner of the kind of information made explicit by a representation, the information made available ... and the information that is discarded" (Marr & Nishihara, 1978, p.276).

The detail of the shape structure that can be derived from these primitives depends in turn on the size of the area in the visual field that they subsume. Consider for example the perception of a human hand. For primitives based on very small areas of the visual field, the representation of the stimulus would be highly detailed, e.g., with depictions of details such as hair and pores. To extract the shape of the arm from such a finely detailed representation, such detail is normally not needed and would require "a rather sophisticated analysis" (Marr & Nishihara, 1978, p.275). In such a case, our ability to see the 'forest' may be hindered by the 'trees' – we are seeking information on overall structure, and not subtle nuances in the retinal image.

In contrast, if the primitives are based on very much larger areas of the visual field, the description of shape may be so coarse that elements of the structure, such as fingers and hands, may be omitted entirely.

These examples are extreme cases of scale specificity. However, they demonstrate that the scale at which visual processing occurs, which is presumably fixed at any one time, determines the amount of information available to make shape judgments. On the one hand, there may be too much superfluous detail (which complicates the task) or, on the other hand, too little information to properly determine the stimulus shape. To compensate for the change in the amount of shape information available caused by scale differences in images projecting on to the retina, an adjustment of the size of the shape primitives may be necessary. This would equalize the amount of information obtained regarding shape and may be achieved at a level prior to shape recognition.

In addition to the image parsing, shape matching and attention experiments listed above, there is also neurophysiological evidence for the zoom effect in the visual system. Smith and Marg (1975) have demonstrated that the receptive fields for some neurons in the visual cortex increased at near fixation for macaque monkeys (see also Marg and Adams (1970) for similar findings for humans). This would reduce the amount of image information obtained from the highly detailed stimulus image present at larger retinal sizes. In turn, this would simplify the representation of overall shape (as opposed to the fine details of the object structure that are available at nearer fixations). In other words, an internal scale transformation may be required to ensure that the level of detail is not too high (making the derivation of shape attributes more difficult) or too low (failing to encode quintessential elements of the shape itself). Note that the preliminary scaling is not done to align two images for the purposes of recognition. Rather, it is performed in order to maintain a consistency of detail and information in the stimulus image that will facilitate later recognition mechanisms, regardless of whether these are sensitive or independent of orientation and size transformations. In other words, the attentional or otherwise low-level perceptual mechanism may be completely independent of the transformational sensitivity of the resulting representation because it is concerned only with gross informational content and stimulus detail.

The presence of two independent size-specific mechanisms makes the task of interpreting results between and within studies difficult. This task is made even more difficult due to differing (and sometimes low) power levels between the experiments reviewed. An attention-based zoom effect should be present in all experiments and in all conditions where scale varies. For example, it would be present in addition to shape-specific size normalization in matching trials of matching tasks. The separate effects of the two mechanisms would be additive in producing an overall 'normalization' profile in response latencies. However, when no normalization for size is necessary for same shape matching, or when the two pairs of shapes in the matching task are entirely different, the only size specific effect would be that of the attentional mechanism. This would be reflected in RTs as a faster overall rate of transformation than in the previous cases where both size sensitive operations were in action. The combined influence of both size-specific mechanisms can account for a number of findings in various studies – not only those involving scale specific attentional phenomena but also matching studies (e.g., Besner, 1983; Bundesen &

Larsen, 1975; Larsen & Bundesen, 1978; Cave & Kosslyn, 1989; Larsen, McIlhagga & Bundesen, 1999).

The role of an attentional lens may also explain other anomalies in the experimental literature. For example, in Larsen et al.'s (1999) study, performance on a matching task was compared between simultaneous and successive presentation conditions. The overall slope of size effects was steeper under the simultaneous paradigm. In addition, only in this condition were the functions affected by the complexity of the stimuli as measured by number of line segments (where increasing complexity is associated with higher intercepts). The authors provided a good simulation of the response behaviours with a random walk model. This model assumed that, in the successive paradigm, Ss encode a subpattern of the first image. This encoding is then normalized to the size of the second image to test for a match. Meanwhile, in the simultaneous paradigm, the process of encoding, transforming and testing occurs several times between the two images.

However, these results could also be explained in terms of the two size-specific mechanisms discussed above. In the simultaneous paradigm, size-specific normalization is dominant while, in the successive paradigm, the only mechanism is that of the attentional zoom lens. This would account for the difference in overall rates because there is a faster rate of scaling for the zoom mechanism. It would also account for the different effects of complexity. Only the shape-specific normalization is affected by complexity because it involves a relatively detailed shape analysis whereas the zoom lens is affected only by gross image area which is independent of stimulus complexity. Such an account would mesh well with the influence of timing on size sensitivity, as outlined in section 7.9. That is, size sensitivity is linearly related to gap size in such a way that shape-specific normalization is less prevalent with larger gaps between same shape images. In the successive paradigm, the ISI was 500 ms, which is the point at which Santee and Egeth (1980) failed to find consistent normalization for scale. In contrast, there is no ISI whatsoever in the simultaneous paradigm, although an ‘effective’ ISI may be the time to shift eye gaze between the two images. According to Santee and Egeth’s (1980) account, normalization should be dominant under this condition only and this is consistent with our account of Larsen et al.’s (1999) study outlined above.

If attention is scale-specific, then we would expect there to be consistent findings of a scaling effect whenever shape-specific effects of size are not present (e.g., with foils that are entirely different shapes from match objects) or have otherwise been eliminated (i.e., size invariant shape representations). In other words, there should always be a consistent residual effect of size changes in responses even when size-invariant recognition is employed. However, this is not always the case in the experimental literature. One possibility is that many of the experiments lack the statistical power to detect the relatively subtle effects of the zoom lens. Jolicoeur and Besner (1987) have noted that all the experiments that failed to find any effects of size for different shaped foils, published before their paper, used relatively small size ratios up to and including 1:1.167). Such small ratios may be inadequate to reliably test for size specific effects because the effect of size will be relatively small. In addition, the average N in the studies investigating size effects in recognition tasks was only 12. The use of a small N may be based on the a priori belief that there is little between-subject variability (i.e., individual differences) in size sensitivity on such tasks.

Given findings of individual differences in orientation sensitivity outlined in section 5.11, the notion of similar variability for size effects is worthy of investigation. Regardless of any possible individual differences, Ns should be large enough to adequately reduce interference from experimental noise. Both small Ns and small effect sizes (proportional to size ratios for a given rate of transformation) contribute to low statistical power and this may explain the failure to find significant effects of size in some experiments for some or all trial types. This would also explain why some experiments detected only shape-specific size effects – the rate of scaling of the attentional mechanism is much faster than that of both mechanisms combined and hence the effect size is relatively smaller. Such experiments may have had the power to detect only the larger effect size.

An alternative explanation is that attentional deployment varies with task demands. Farrell and Pelli (1993) have demonstrated that error rates are greater for mixed versus single scale arrays for locating targets. However, no such difference is evident when the task involves identification. This study demonstrates that there may be attentional differences between identifying and locating images - perhaps because greater attentional resources (e.g., a finer detail image analysis) are needed to locate the object. However, when attentional changes are detected in the recognition experiments described previously, they occur very quickly. Such rapid changes may be not be reflected in error rate but only in latency. In other words, the activity of the zoom lens may not have been detected in the identification task. Alternatively, the perceptual scale may be set simply to the largest size of the arrays (which is the same for the large scale and mixed scale arrays) because it may provide sufficient information for recognition within the array in this task. There would be no need to adjust the lens for the different sizes of the array. In contrast, locating requires a higher degree of image detail and may require scale adjustment to make a finer discrimination based on spatial characteristics not needed for simple identification.

## 7.15 Summary

In this chapter, we have shown some of the promising similarities between the processing of size and rotation information in the recognition of objects. Normalization and invariant processing are evident for both types of symmetries and there is some evidence that the conditions influencing which type of processing dominates may be similar for both. Such factors include presentation timing in the experimental paradigm, familiarity with an object in the task, the type of foil images that are used and the complexity of the stimulus. However, more research is needed to establish the strength of these apparent similarities between the influence of the two symmetries. As discussed, there is also evidence for an attentional mechanism that appears to act exclusively for image size. The actions of this mechanism are non-shape specific and it is possible that it could account for some effects in the literature that may be incorrectly ascribed to size normalization for shape recognition.

# CHAPTER 8: The combined effects of size and planar rotation

*Few studies have investigated the orthogonal influence of size and orientation in recognition tasks. The results of these investigations have produced unusual and conflicting results. Practice induces at least partial attenuation of one or both of the influences of the two transformations. However, little is known about how this attenuation is achieved. In addition, none of these studies has replicated the conditions hypothesized to be necessary for recognition invariant to the transformations of orientation and size as outlined in chapters 5 and 7.*

## 8.1 Introduction

Given some of the similarities, both abstract and empirical, between the effects of planar rotation and size in separate recognition experiments, let us now turn our attention to how these two variables may interact in an experimental situation. We have already outlined some of the common factors that influence the choice of either normalization or invariant mechanisms for the two different transformations in isolation. One possibility is that such similarities result from the actions of two recognition mechanisms within the perceptual system – one that requires the normalization of rotation and scale to provide a match and one that is invariant to such transformations. Many theorists have proposed models for one or other of such systems as outlined in chapters 3 and 4 respectively. For example, in Huttenlocher and Ullman's (1987) model the image representation is pictorial and in order to be recognized it must be normalized in both size and orientation with respect to an internal model. In contrast, Cooper, Biederman and Hummel (1992) propose that the stimulus representation is invariant to the properties of size and orientation, so that no transformation is required for identification purposes. In this chapter, we will examine the small number of experiments in the literature that have investigated the combined effects of planar rotation and size on identification behaviours. None of these has convincingly demonstrated recognition invariant to both properties. However, it is suggested that the conditions necessary for invariant recognition in the previous chapters were not present in such experiments.

## 8.2 Studies investigating the simple effects of size and rotation

Several studies have examined the effects of both rotation and size symmetries without crossing these variables. For example, Posner and Mitchell (1967) simultaneously presented nonsense forms (Gibson figures) in a recognition task where these images could differ in either size or angular deviation from the upright, with two levels of each transformation. The results suggested that a normalization process was used for both types of disparity. As mentioned previously, Rösler et al. (1995) used a handedness task to demonstrate normalization effects for rotation and scale differences in separate blocks. Other studies have utilized matching paradigms where only the foil images are transformed along both dimensions simultaneously (Bundesen & Larsen, 1975; Cave & Kosslyn, 1989; Larsen & Bundesen, 1978). The use of rotation in these experiments was limited to a single angular deviation (either $90^0$ or $180^0$) that is used only to differentiate foil and match images. Latencies for foil images, like those for match times, increased with the size ratio difference and overall responses to foils were longer than those for match trials. However, such results may not necessarily be the result of the additive effects of rotation and size scaling but other processes associated with the production of a negative response.

## 8.3 Studies investigating the orthogonal effects of size and rotation

With respect to those studies that have orthogonally manipulated the scale and orientation of match images, two have demonstrated straightforward additive effects of the two variables in response latencies. Both of these studies involved matching tasks. The first of these studies used a successive presentation paradigm utilizing alphanumeric stimuli and distractors that were mirror-reflections of the match objects (Bundesen, Larsen & Farrell, 1981). The additive nature of orientation and size effects paralleled findings from a study by the same authors that investigated the influence of size and rotation changes in the production of apparent motion that has previously been discussed (Bundesen, Farrell & Larsen, 1983).

Despite the additive nature of the effects, several subjects in both the matching and apparent motion experiments reported the phenomenal experience of transformation along both dimensions simultaneously, i.e., a continuous helical motion. Such a helical motion would be the shortest transformational path to match between such images. This should in fact produce a statistical interaction between scale and orientation compensations. However, Bundesen, Larsen and Farrell (1981) argue that the overall transformational path only approximates a continuous, screw-like motion and that the process occurs via "sequentially alternating smaller steps of mental size transformation with smaller steps of mental rotation" (p. 288). Responses to foil trials in this study were similarly affected by the transformations. In addition, the only other study in this chapter that analyzed foil responses also demonstrated comparable effects of transformations on both trial types. This study by Larsen (1985) also used MRFs (mirror reflected foils). Therefore, these findings are consistent with the trend noted for the simple effects of size and transformation in matching tasks in sections 5.6.1 and 7.12.1 – namely, that MRFs require comparable normalization to matching shapes.

The other study demonstrating additive effects of size and rotation is that of Sekuler and Nash (1972). Like Bundesen et al.'s (1981) study, the paradigm was one of successive matching. But, in contrast, the stimuli were rectangles instead of alphanumeric stimuli and the foils were different shapes and not MRFs. Analysis of the response times similarly revealed that the two orthogonal effects were additive. Subsequent analysis of the error data from the study did reveal an interaction between the rotation and scale changes (Besner, 1978). However, Bundesen et al. (1981) have pointed out that one cannot infer from this that the process of size and rotation compensation occur in the same processing stage. According to Sternberg's (1969) *additive stages logic*, individual sequential stages will be reflected in the experimental data by additivity in response times and need not also be evident in error rates. Certainly, the overwhelming majority of studies in this area have demonstrated effects in response times only and the seemingly contradictory effects in error rates from Sekuler and Nash's (1972) study may be attributable to noise or variables not associated with pattern recognition.

The findings from Bundesen et al.'s (1981) and Sekuler and Nash's (1972) studies suggest that normalization for size and planar rotations are separated temporally. Note that this does not necessarily mean that these transformations cannot be carried out by the same mechanism. In addition, it does not mean that the two transformations cannot be influenced by the same factors. What it does suggest is that the two processes do not occur simultaneously.

Like the studies above, Larsen (1985) also found simple additive effects of size and planar orientation in a similar experiment that also used a simultaneous matching paradigm. However, when the experiment was modified to include just two polygonal shapes rather than new shapes on each trial, the effects of rotation and size were lessened and the two factors interacted in an unusual manner (Experiment II, Condition A). Another matching study by Kubovy and Podgorny (1981) failed to find any significant effects of size despite significant effects of rotation. Both these studies are characterized by a very large number of trials – 1296 in both of Larsen's (1985) experiments and 6,400 in Kubovy and Podgorny's (1981) task with only two different root shapes in both. Therefore, as with the individual effects of rotation and size outlined in the chapters 5 and 7, there is some evidence that the effects of these transformations can also be attenuated with familiarity in experiments orthogonally manipulating these variables.

## 8.4 Invariant recognition or multiple representations?

While the influence of rotation and / or size transformations may be reduced in the studies of Kubovy and Podgorny (1981) and Larsen (1985), exactly how these effects are attenuated remains unclear. In chapter 5, it was argued that there were two ways in which the effects of orientation could be reduced with practice on a set of stimulus shapes. The first is a recognition strategy that utilizes a representation of shape that is invariant under changes in orientation, i.e., an orientation-invariant recognition strategy as demonstrated by McKone and Grenfell (1999) and Murray et al. (1993). The second is by the formation of multiple orientation-specific, handedness-specific representations that act as templates for each of the presented orientations, as evidenced by Tarr and Pinker (1989). There is no a priori reason why both strategies could not operate in the context of size transformations and in the context of orthogonal size and rotational transformations. In fact, given the similarities between rotational and size normalization effects, it would seem prudent to test for the presence of multiple representations in this context. In Experiment II of this project, we will test for the presence of multiple size and / or rotation specific representations in a recognition task.

There are three reasons why multiple representations may have been used in Larsen's (1985) and Kubovy and Podgorny's (1981) studies. Firstly, we argued in chapter 5 that multiple representations may be utilized when the images to be discriminated are very complex and confusable (e.g., easily confused as the same objective shape). In particular, multiple representations may be formed when the foil images are simply mirror-reflected variations of the match shapes (i.e., the foils are the same objective shape). For the most part, the effects of rotation are more resilient in studies with MRFs, and when they do attenuate, there is some evidence that this can be the product of multiple representations and not an invariant recognition mechanism (Tarr & Pinker, 1989, Experiment I). Significantly, both Larsen's (1985) and Kubovy and Podgorny's (1981) experiments employed MRFs. Therefore, the reduction in one or both transformation effects in these studies may also be due to the deployment of multiple representations and not invariant recognition strategies.

Secondly, both these studies employed a very large number of trials yet failed to eliminate the effects of both transformations. As mentioned in section 5.7, several studies have extinguished the effects of such symmetries in a significantly smaller number of trials (e.g., Murray et al., 1993) or possibly without any practice at all (e.g., McKone and Grenfell, 1999). In these studies, the strategy used was demonstrated to be orientation invariant recognition and not multiple representations and the paradigms were free from MRFs. In contrast, Tarr and Pinker (1989, Experiment I) did find evidence for multiple representations after repeated presentations. Not only did they use MRFs like Larsen's (1985) and Kubovy and Podgorny's (1981) but they also required a relatively large number of trials (2304 trials on only three root stimuli) to attenuate, but not eliminate, the effects of transformation.

The third reason for hypothesizing the use of multiple representations in Larsen's (1985) and Kubovy and Podgorny's (1981) experiments, is the unusual attenuation effects, especially in the former study. In Larsen's (1985) second experiment, the interaction effect is attributable to a very selective 'invariance' – namely, a lack of size effects only when the angular difference was $60^0$ and $120^0$. A possible explanation for this pattern of results is the formation of representations specific in size and orientation for some but not all of the possible object versions in the experiment. In Kubovy and Podgorny's (1981) experiment, the presence of orientation effects but not size effects, may be explained by the formation of multiple size specific representations. In this case, a separate representation could have been created – one for each object in all possible sizes but always in the upright position – which could have been rotated to match during the experiment. The variety of reasons given above suggest the involvement of multiple representations similar to those proposed by Tarr and Pinker (1989). However, neither study directly investigated whether such representations may have been formed. The second experiment of this project, which involves the orthogonal manipulation of size and rotation, is designed to test more specifically the mechanism behind any attenuation effects and in particular the use of multiple orientation and size specific representations.

## 8.5 Alternative explanations: Experimental design, sample size and encoding strategies

Another possible explanation for the unusual effects of the transformations in these studies is that they are an artefact caused by the experimental design. Both employed simultaneous matching paradigms and in neither study were the effects of the two types of transformations fully crossed with left-right position in the stimulus array. For example, in Kubovy and Podgorny's (1981) task, the image on the left was only ever displayed at a fixed size and in one of only two different orientations ($0^0$ and $180^0$). In contrast, the object on the right could be presented at any combination of 5 sizes and 5 orientations relative to the other, simultaneously presented pattern. Kubovy and Podgorny (1981) raise the possibility that this aspect of the design may have contributed to the failure to find size effects in their study. However, they admit that it is unclear as to exactly how this might have been achieved.

It may be that the simultaneous matching paradigm itself does not involve a holistic comparison between the two images but rather a more piece-meal approach to pattern matching. Based on a study investigating the effects of size ratio on pattern matching performance, Larsen, McIlhagga and Bundesen (1999) argue that, while successive matching is achieved by a single process of encoding, transforming and comparing, simultaneous matching is achieved via multiple processes of encoding, transforming and comparing carried out sequentially. This hypothesis was supported by the failure to demonstrate complexity effects in successive and not simultaneous matching paradigms. In addition, the theory was successfully modeled using a random walk model. The use of multiple processes in the simultaneous matching paradigm increases the likelihood of interference of noise that would be evident at each of these stages and could obscure the effects of transformations in the data. In addition, there is the possibility that the transformation process or processes can occur simultaneously with other stages in the processing. For example, the evidence for transformation would be significantly reduced if this process occurred simultaneously with the shifting of gaze necessary several times in the simultaneous pattern matching procedure. Having encoded an element of one image, before comparing this to an element in the second image the original element must be transformed and the gaze shifted from one image to the other. If the transformation was achieved before the shift was completed then the time to transform the image would not be evident in response times.

It should be noted that both of the aforementioned studies also produced very long response latencies. In Bundesen et al.'s (1981) study, using a successive matching paradigm and both rotational and size transformations, the upper limit on response times was 900 ms. In Bundesen and Larsen's (1975) study, which investigated only size effects in a simultaneous presentation paradigm, response times did not exceed 600 ms. However, in Kubovy and Podgorny's (1981) study, the latencies ranged from 900 to 2,000 ms. Similarly, in Larsen's (1985) study over 7.2% of the RTs (930 scores) were longer than 4 seconds and the removal of these scores did not affect the statistical conclusions. In other words, these were not outliers but responses consistent with the overall pattern of results. These abnormally long response times, together with evidence of multiple processes in similar paradigms, suggest that the processes may have been more complicated than a single transformation to match procedure and also increase the likelihood that such a procedure may be obscured by other influences.

Finally, both studies used relatively small numbers of subjects, which always included the authors – 5 and 9 in Larsen's (1985) experiments and 8 in Kubovy and Podgorny's (1981). All things being equal, the power to detect an effect of a certain size increases with the number of subjects. Examination of the tabulated means and graphical representations did not suggest trends that failed to reach statistical significance. However, while the studies all involve many measures per subject, the validity of these experiments rests on very small between-subject variability. A study by McKone and Grenfell (1999) has questioned this assumption by demonstrating individual differences in sensitivity to orientation changes and the interaction between such changes and practice in a recognition task. It may be that between-subject variability has obscured some effects and that studies with larger Ns may produce different results. Jolicoeur and Landau (1984) have argued that the failure to find statistically significant effects of normalization may often be attributable to inadequate experimental power. The interaction effect found in Larsen's (1985) experiment may also be the result of the same problem. In his examination of the reaction time paradigm, Pachella (1974) has stated that "the data needed in order to demonstrate true additivity requires a precision that few reaction time experiments obtain" (p.52). Similarly, Sternberg (1969) pointed out that the influence of experimental artifacts is more likely to obscure true additivity than true interaction. In other words, confounding factors that influence the reaction time results (e.g., measurement noise) are more likely to produce a misleading statistical pattern of an interaction between variables even when the effects are truly additive.

# CHAPTER 9: Experiment II

*The second experiment investigated the orthogonal effects of size and orientation using the same root shapes and experimental paradigm as Experiment I. Similarly, shape specific orientation effects were limited to the first stages of the experiment. However, scale effects were consistent throughout the task for both match and foil trials. The influence of size was attributed to a non-shape specific mechanism that is attentional or at least pre-recognition in nature. There was no interaction between the effects of the two transformations and therefore no evidence that normalization occurred along the dimensions of size and orientation simultaneously. There was only limited evidence for a correlation between the two effects. There was also support for axis-based representations within the data. This, and other effects in Experiments I and II and previous literature, are incorporated into an algorithmic model of visual processing.*

## 9.1 Introduction

### 9.1.1 Aims

The first experiment demonstrated a 'trend to invariance' with respect to rotated images. Armed with the same stimuli and experimental paradigm, we can now investigate whether there is a similar pattern in the identification of objects under scale *and* planar orientation changes - is there a 'trend to invariance' for rotated and scaled images as well? There are two major aims in this second experiment. The first, as mentioned above, is to examine whether responses become invariant to orientation and size. Such a finding may be consistent with the progressive involvement of a recognition mechanism that is invariant to both transformations (see chapter 4). The second aim is to examine whether and how the two properties may be normalized. Are the effects additive in a way that would suggest that normalization for the two tasks is carried out non-simultaneously, perhaps by separate mechanisms? Or do the effects of size and orientation interact? This latter finding could suggest the action of some kind of 'global transformational mechanism' - a mechanism that deforms a single template or abstract representation along both size and orientation dimensions simultaneously.

### 9.1.2 Experimental design considerations

Experiment I provided more than just stimuli and an experimental procedure for examining further effects of image transformation. It also gave us some idea of the amount of practice necessary for the formation of an invariant strategy as well as an estimate of the magnitude of the normalization effect, at least with respect to orientation. Using this information, several steps were taken to improve statistical power in the second experiment. Firstly, the number of subjects was increased from 18 to 30. Secondly, the absolute angular difference between rotation levels was increased both within blocks (from $60^0$ to $80^0$) and between test and train blocks (from $30^0$ to $40^0$). Therefore, for the same normalization rate, regardless of whether it is associated either with a single image or multiple representations, the effect size will be larger. For example, if the effects of size and orientation are additive and if we anticipate similar variability of response to the first experiment, this increases the power to detect a fast normalization rate for orientation of 1000 deg/s to .93 (at $\alpha$ = .10). Thirdly, there will be no between-group factor like the two different presentation time conditions in the first experiment. While the effect of this factor was negligible statistically, it may have increased between-subject variation that could occlude within-group effects.

One reason for increasing the power of the experiment is the importance of testing for an interaction between orientation and size factors. It is estimated that the experiment will have adequate power (.8) to detect an interaction of 'medium' size at $\alpha = .10$ (i.e., an effect size of .33 where 'medium' effect = .25 and 'large' effect = .40 as defined by Cohen (1988)). As with Experiment I, all comparisons will be considered up to the .10 alpha level. This step was taken because a major research hypothesis predicts *no* differences in means associated with size and orientation transformations. Therefore, every effort was made to allow the research hypothesis to fail. However, care will be taken in the evaluation of effects significant only at the .10 level due to the increased likelihood of spurious or misleading findings.

Previous research has guided the choice of both size and planar transformations used in this study. As mentioned in the review section, Jolicoeur and Besner (1987) have argued that size changes are often too small in similar experiments (no larger than 1:1.67 in experiments prior to their own). As a result, this experiment will use size changes up to 1:2. This is the optimal choice based on the consideration of three factors – the recommendations of Jolicoeur and Besner (1987), the size of the screen used in our experiments and the need to minimize picture resolution differences between different sized versions of the stimuli images. With respect to the orientations used, $0^0$ and $180^0$ versions were again omitted for the reasons outlined in the introduction to the first experiment (see section 6.1.2). As in the previous experiment, all values used in the statistical analysis will be pooled from numerous observations to increase measurement consistency. For example, the test for overall rotation effects in the training blocks will be based on the comparison of just two means – each of which will summarize 144 responses from all of the 30 subjects (i.e., a total of 4320 pairs of observations).

## 9.2 Methodology

All the details of Experiment I were similar to those of Experiment II, except where specified below.

### 9.2.1 Subjects

Thirty subjects took part in the second experiment. Ages ranged from 16 to 52 with an average age of 31.60. Twelve of the subjects were female (40% of the sample).

### 9.2.2 Procedure

The first stimulus was presented for 500 ms for all subjects in the experiment. The shape in the second stimulus of the pair was both rotated and scaled. As with the first experiment, the transformations utilized depended on whether the trial was in a TRAIN or TEST block. For TRAIN trials, the transformation consisted of a combination of one rotation of the set {$50^0$, $310^0$, $130^0$, $130^0$} and one scale change from the set {5/4, 4/5, 7/4, 4/7}, where each ratio represents the relative size of the first image to the second. For TEST trials, the transformation consisted of a combination of one rotation of the set {$90^0$, $270^0$, $170^0$, $190^0$} and one scale change from the set {3/2, 2/3, 2/1, 1/2}. These transformations were always conducted around the centroid of the shape (defined as the first moment of area). Each block consisted of 96 trials and each *match* and *foil* object is presented once at each of the 16 transformations possible in each block. The experiment consisted of three TRAIN blocks followed by a single TEST block such that there were 384 trials in total. As with the first experiment, the 8 practice trials utilized the set of transformations used in the TRAIN blocks.

The information sheet was altered to include references to, and examples of, rotations and scale changes (see appendix E). Similarly, the first set of verbal instructions was altered as follows (alterations in **bold**):

> *In each trial in this experiment you will first see a cross on which to centre your gaze. Then this cross will disappear and you will see an image displayed on the screen. After a time, this image will disappear. Then, after a brief pause, a second image will appear. At this point you must decide as quickly as possible whether the second object is the same as the first by pressing either the red key on the right for "Yes" (i.e., "Yes, this is the same object") or the red key on the left for "No" (i.e., "No, this is a different object"). You should respond with the "Yes" key whenever the second image has the same shape as the first* ***even if its size has changed and / or it***

***has been rotated**. Try to respond as quickly as you can but without sacrificing accuracy, i.e., try to avoid errors caused by trying to respond too fast. Press any key on the keyboard when you are ready to try some practice examples.*

## 9.3 Results

In this *Results* section, the terms "highly", "moderately" and "marginally", when used in relation to the findings of a statistical test, refer to results achieving significance at the .01, .05 and .10 alpha levels respectively. In addition, whenever Mauchly's Test of Sphericity is significant or close to significance at the .05 level, this result will be reported alongside the associated effects. In such cases, the degrees of freedom will be adjusted using the Huynh-Feldt correction.

### 9.3.1 Match trials

As in Experiment I, the data was filtered for excessively long reaction times before analysis. Reaction times longer than two seconds were counted as "guesses" and, regardless of the accuracy of response, were treated as errors. The overall mean response time for correct responses was 330.90 ms (sd = 135.45). For the ANOVAs conducted, there were, at most, four within-subjects factors. The first of these factors was TYPE with two levels. This compared the responses of *matching trials* (where the second shape was the same as the first) to *foil trials* (where the second shape was different from the first).

The second factor, BLOCK, compared response times across different blocks in the experiment (with a maximum of 4 levels – one for each block). In the ANOVAs on practice trials, this was used to test for, among other things, practice and fatigue effects. For the ANOVAs on the test block and the train block which precede it, the BLOCK effect was also used to test for the multiple representation + normalization hypothesis. This was made possible by employing test stimuli depicting forms that were a) all $40^0$ further away from the upright than the corresponding orientations in the practice blocks and b) scaled by a linear size ratio that is larger by .25 than the ratios used in the train trials for both magnifications and demagnifications. Therefore, if the effects of the two transformations were attenuated by the final practice block and this attenuation was the product of multiple, orientation and / or size specific representations, then there should be an overall increase in response latencies in the test block as subjects recognize the test versions by normalizing from representations formed in the practice trials. For example, if the normalization of rotation and size differences was independent, this would result in an increase in latencies equal to the sum of the time to rotate the image through $40^0$ and the time to magnify / reduce an image by a linear size ratio of 1.25. Note that the choice of angular deviations and scale transformations in the train and test blocks still permits detection of conventional normalization from a single image in both types of blocks.

The third factor in the ANOVAs, ROTATION, compared the effects of angular deviations of the stimuli from the upright. As in the first experiment, the different rotational deviations from the upright were combined into two levels within any one block. The reaction times for all training trials (match and foil), using rotations of $50^0$ and $310^0$, were combined to form a Level 1 reaction time, while those for $130^0$ and $230^0$ were combined into a Level 2 reaction time. For testing trials, the rotations of $90^0$ and $270^0$ were combined to form the Level 1 reaction time, while those for $170^0$ and $190^0$ were similarly combined to form the Level 2 reaction time. Note that, in both training and testing, the difference between the two levels was always equivalent to an $80^0$ rotation under the single representation normalization hypothesis. Thus, if subjects were normalizing from the upright to recognize images, this $80^0$ difference would be reflected in similar differences in reaction times in both test and train blocks.

The fourth factor was SIZE. This examined the effect of the size difference between the first and second stimuli in a trial. The different scalings used in the experiment were, like the rotations, pooled into two levels within any one block. The reaction times for all training trials using magnification factors of 1.25 (a ratio of the first stimulus size to that of the second of 4/5) and 0.8 (5/4) were combined to form a Level 1 reaction time while those for 1.75 (4/7) and 0.57 (7/4) were condensed into a Level 2 reaction time. For testing trials, the magnification factors of 1.5 (2/3) and .67 (3/2) were combined to form the Level 1 reaction time, while those for 2 (2/1) and .5 (1/2) were similarly combined to form the Level 2 reaction time. Note that in both training and testing, the *difference* between the two levels was always equivalent under the normalization hypothesis – the difference in linear size ratio between the greatest magnification and the smallest magnification (or, equivalently, the greatest demagnification and the smallest demagnification) was always .5. Thus, if subjects were normalizing for size with respect to the first image and RT is a linear function of size ratio (which is the function that form specific normalization has taken in previous research), the effect of SIZE would be equivalent in test and train trials.

The overall mean response time for correct responses to match trials was 330.90 ms (sd = 135.45). A repeated measures ANOVA was conducted on these response latencies within the three training blocks. This revealed a highly significant difference in mean times across the different blocks (F (2,58) = 17.84, $p < .001$). The function relating RT and practice was a monotonically decreasing one and a trend analysis demonstrated a highly significant linear component (F (1,29) = 23.51, $p < .001$). However, the trend is best described by a quadratic function with a relatively greater improvement in times between the first and second blocks and this is demonstrated by a moderate quadratic component in the trend (F (1,29) = 5.45, $p < .05$). This trend for the first three blocks is visible in figure 9.1.

In addition, there were overall moderate effects of SIZE (F (1,29) = 6.89, $p < .05$) and ROTATION (F (1,29) = 5.85, $p < .05$) in the response time data. The direction of both these main effects was consistent with normalization and none of the interactions involving these two factors was significant (even at the .10 alpha level). As can be seen in figure 9.2, there was only a small, insignificant increase in the influence of rotation at greater scale differences and a large amount of overlap between the two levels. The effect of SIZE did not vary significantly across the train blocks, as can be seen in the top panel of figure 9.3. In this figure, and in many other figures in this chapter, the dotted line represents a hypothetical invariant strategy for comparison with the actual data. Overall, the mean difference between the two size conditions was 15.80 ms. If subjects were in fact normalizing for size and the function relating RT and size ratio is a linear one, then the *hypothetical slope of size normalization* would be 31.61 ms.

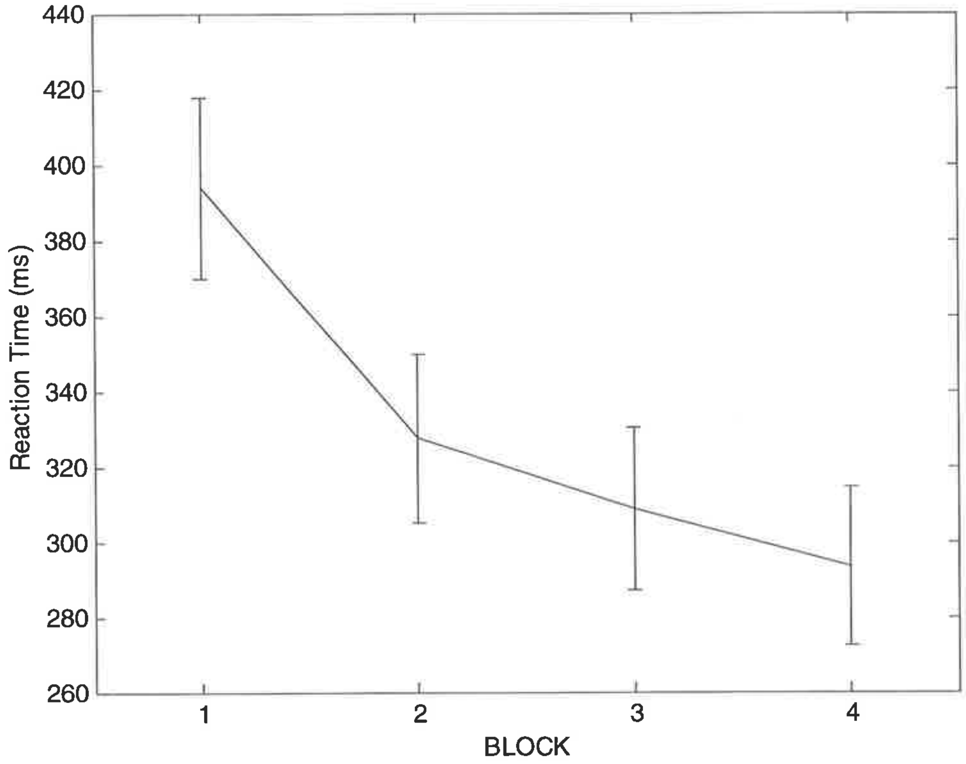


*Figure 9.1. Reaction times for correct matching trials across the four experimental blocks. The error bars represent the mean +/- one standard error.*

In contrast to the consistency of SIZE effects with practice, there was also a BLOCK by ROTATION interaction that was close to moderate significance (F (2,58) = 2.92, p < .10). As can be seen in the top panel of figure 9.4, there was an orientation effect consistent with normalization in response latencies for the first two train blocks. However, by the third block this effect had disappeared. The *hypothetical rates of orientation normalization* were 3703 deg/s and 2262 deg/s for the first and second blocks respectively.

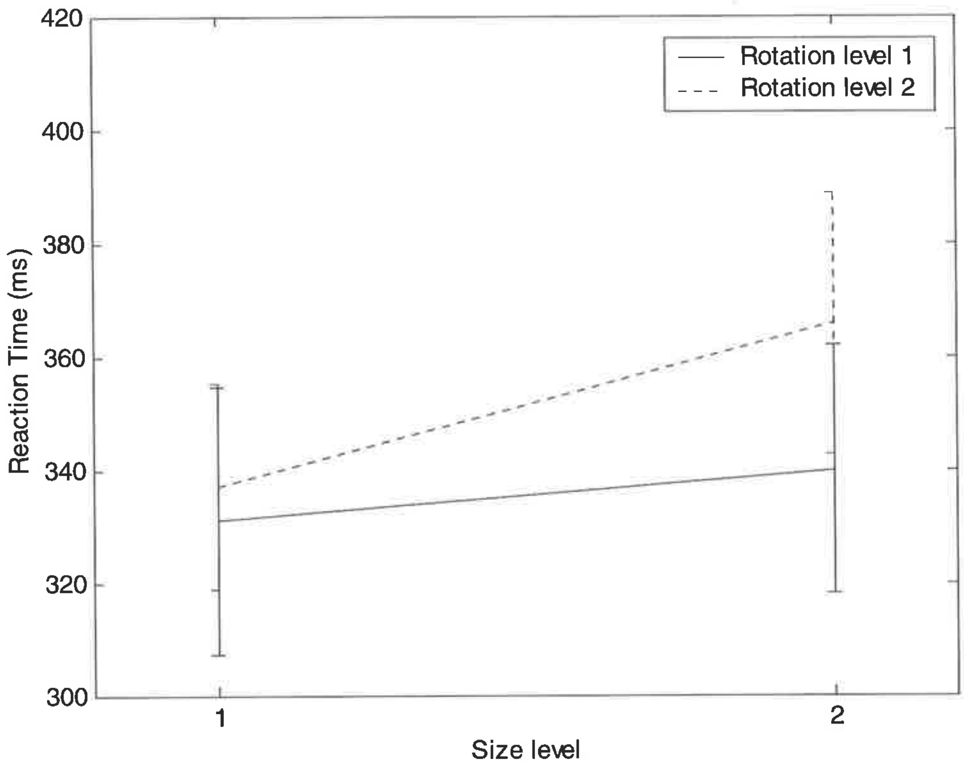


*Figure 9.2. The interaction of ROTATION and SIZE effects on latencies for correct responses to matching trials. The horizontal axis represents magnitude of size changes where 1 represents "smaller" scale ratios of the first image to the second of 5/4, 4/5, 3/2 and 2/3 and 2 represents "larger" scale ratios of 7/4, 4/7, 2/1 and 1/2. The solid line represents "smaller" rotation changes (Level 1) incorporating angular deviations from the upright of $50^0$, $310^0$, $90^0$, and $270^0$ while the dotted line represents "larger" rotation changes (Level 2) utilizing angular deviations from the upright of $130^0$, $230^0$, $170^0$ and $190^0$. The error bars represent the mean +/- one standard error.*

Figure 9.4 also demonstrates an unusual effect of rotation in the test block, i.e., an effect of orientation which is opposite in direction to normalization. An ANOVA involving response latencies to match trials in the last train block and the test block was conducted. This demonstrated the unusual change in orientation effects between the two last blocks by a BLOCK by ROTATION interaction that was close to moderate significance (F (1,29) = 3.92, $p < .10$). Consistent with the trend visible in the top panel of figure 9.3, there was a moderate effect of size in the test block (F

(1,29) = 8.81, $p < .05$) that was consistent with normalization. Similarly, the ANOVA for the last two blocks revealed an overall moderate SIZE effect (F (1,29) = 5.34, $p < .05$) and there was no evidence for any change in this effect between this pair, as demonstrated by the lack of any significant interaction between SIZE and BLOCK factors.

There was no significant change in overall response time (i.e., BLOCK effect) between the test block and the train block directly preceding it. In fact, the raw difference in means of 15 ms was *opposite* in direction to that predicted by a multiple representation + normalization hypothesis (which predicts an increase in response times in the test block). This finding is visible in figure 9.1. The confidence interval for this mean difference was $\{-38.04 < u_d < 7.23\}$. Note that, even if subjects utilized the very fast normalization rate of 3703 deg/s found in the very first block to normalize any multiple representations, the resultant population mean of 10.80 ms would be outside this interval. In other words, we can be 95 % certain that the actual population mean difference is not consistent with a multiple representation + normalization hypothesis.

Overall response accuracy was good at 96% correct (sd = 6.47) and the means and standard errors of the accuracy of responses split by block and transformation level are contained in appendix F. With respect to this measure, there were no significant effects in the repeated measures ANOVA for the training blocks. However, there was an interesting effect of orientation revealed in the ANOVA on the test block and the final train block. This consisted of a BLOCK by ROTATION interaction that was close to moderate significance (F (1,29) = 3.68, $p < .10$) and an overall moderate effect of ROTATION (F (1,29) = 6.22, p < .05). Planned contrasts were conducted on the simple effects of ROTATION and SIZE in the two blocks. These revealed a moderate effect of ROTATION in the test block that was opposite in direction to that predicted by a normalization hypothesis (F (1,29) = 8.91, p < .05]). This amounted to a 3% advantage under the Level 2 condition.

### 9.3.2 Foil trials

An omnibus repeated measures ANOVA was conducted on response latencies to foil trials. This revealed that the difference between latencies for correct responses to match and foil trials was close to moderate significance (F (1,29) = 4.099, p < .10). This amounts to an extra 11 ms of processing time for the distractor trials. The mean response time for foil trials was 341.90 ms with a standard deviation of 124.75. As with the matching responses, a repeated measures ANOVA conducted only on foil trials revealed a significant decrease in response times during the training stage of the experiment. This was demonstrated by a highly significant BLOCK effect (F (2,58) = 20.24, p < .001). While the trend possessed a highly significant linear component (F (1,29) = 21.62, p < .001), it is best described by a quadratic function given the moderately significant quadratic component (F (1,29) = 12.07, p < .05). In other words, the influence of practice in reducing RTs tails off as the experiment progresses.

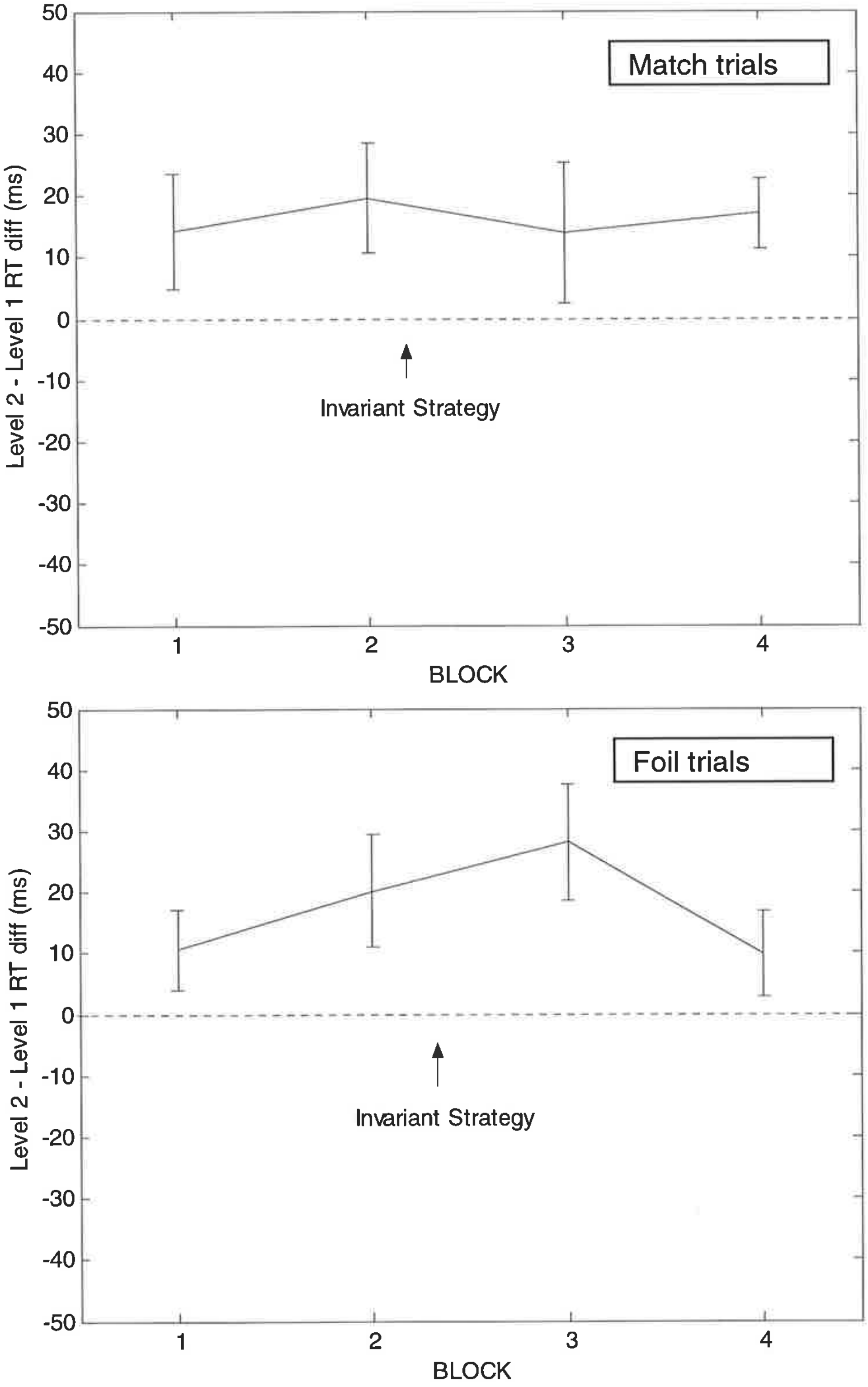


*Figure 9.3. Reaction times differences between the two different scaling levels for correct match (top panel) and foil (bottom panel) trials across the four experimental blocks. Size level 1 represents "smaller" scale ratios of the first image to the second of 5/4, 4/5, 3/2 and 2/3. Size level 2 represents "larger" scale ratios of 7/4, 4/7, 2/1 and 1/2. The error bars represent the mean +/- one standard error.*

There was also an overall moderate effect of SIZE in these three blocks and the direction of this effect was consistent with a normalization procedure (F (1,29) = 10.88, p <. 05). This trend is visible in the bottom panel of figure 9.3. Neither the variation in the influence associated with SIZE across blocks for these foil trials nor the difference in SIZE effects between foil and match trials examined in the omnibus ANOVA was significant even at the .10 level. A similar effect of SIZE, both in terms of magnitude and direction, was also revealed in an ANOVA on the test block and the final train block (F (1,29) = 9.81, p < .01). Overall, the difference between response times to the two size conditions was 19.60 ms. The resulting *hypothetical slope of size normalization* was 39.20 ms.

In contrast, the effect of ROTATION in foil trials was not significant. The mean difference in rotation levels was only 7.80 ms and the direction of this difference was opposite to that predicted by a normalization hypothesis. This pattern is demonstrated in the bottom panel of figure 9.4. The different influence of ROTATION between the two trial types visible in this figure was confirmed by the omnibus ANOVA which demonstrated a moderately significant BLOCK by TYPE by ROTATION (F (3, 87) = 3.282, p < .05) and a close to moderate TYPE by ROTATION effect (F (1,29) = 3.920, p < .10).

In terms of response accuracy in the foil trials, the only significant effect was associated with ROTATION. This was moderately sized, consistent in direction with normalization and occurred in the ANOVA involving the test block and the final train block (F (1,29) = 7.31, p < .05). However, this effect amounted to only a 1%

difference in accuracy between rotation conditions. Overall foil accuracy was good, with an average rate of 97% correct (sd = 5.41).

### 9.3.3 Relationship between size and rotation sensitivity

A measure of size sensitivity, $size_{diff}$, was calculated for each subject by subtracting the mean response time for the smaller size ratio condition (*size level 1*) from the larger size ratio condition (*size level 2*) for correct responses to match trials. There was a large amount of between-subject variability with respect to this measure as demonstrated by the high standard deviation of 28.21 (mean = 16.11 ms). The median score of $size_{diff}$ of 19.03 ms was then used to divide the subjects into two equally sized groups – a low size sensitivity (LSS) group and a high size sensitivity (HSS) group. An omnibus ANOVA on response times with the inclusion of a between-subject factor based on these two groups failed to find any significant main or interaction effects associated with this factor except for the obvious interaction with size effects. However, splitting the sample into two groups significantly reduces the statistical power of tests involving this between-subject factor (i.e., N reduces from 30 to 15). Separate ANOVAs for each size sensitive group yielded different patterns of significant results and the raw differences apparent suggested that the failure to detect between-subject effects may have been attributable to the reduced N rather than the size of the effect. In other words, the magnitude of the effects appears reasonable but the statistical power may not have been sufficient to detect them in the between-subject comparisons.

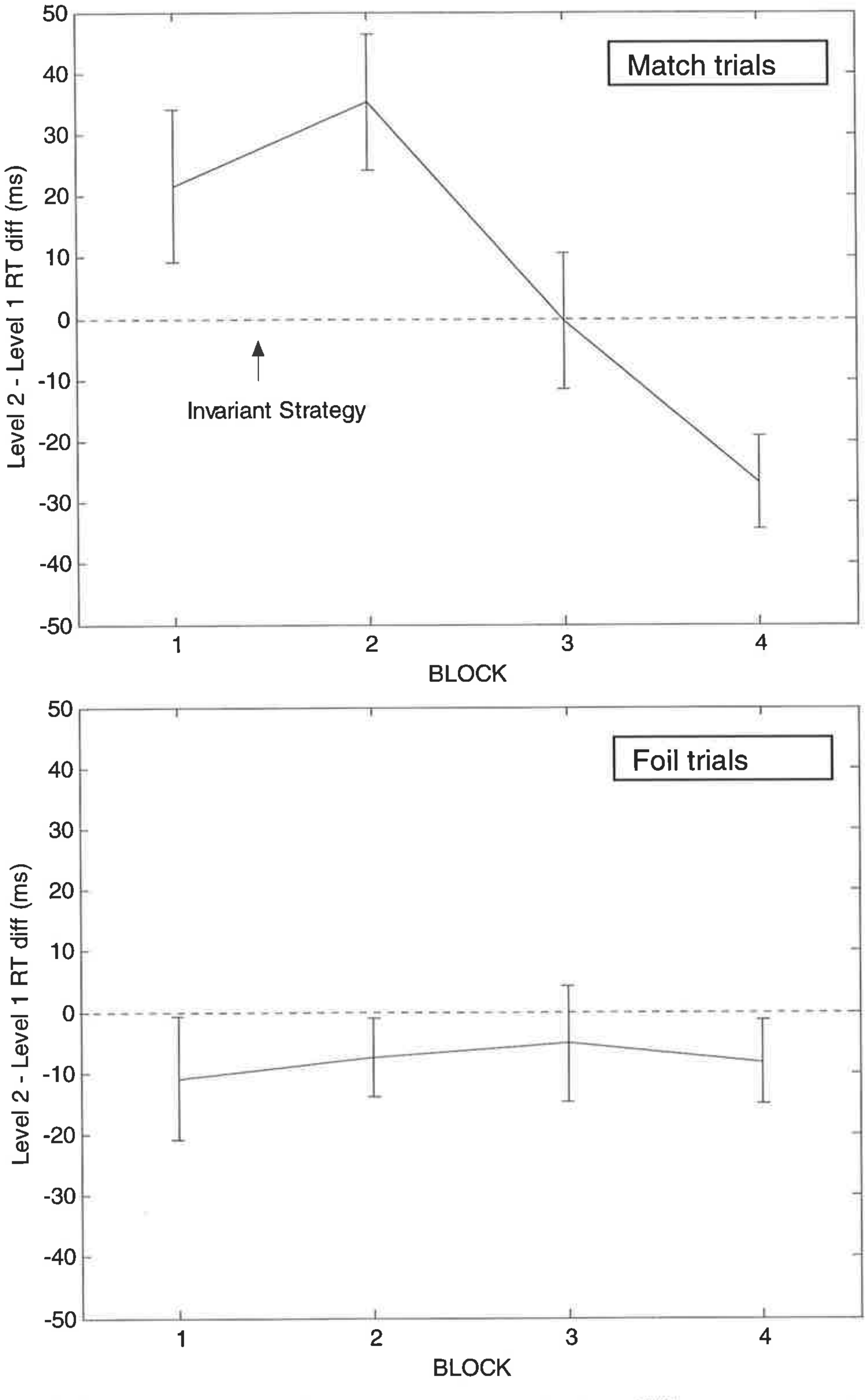


*Figure 9.4. Reaction time differences between the two different rotation levels for correct match trials (top panel) and foil trials (bottom panel) across the four experimental blocks. Rotation level 1 represents "smaller" angular deviations from the upright of $50^0$, $310^0$, $90^0$, and $270^0$. Rotation level 2 represents "larger" angular deviations from the upright of $130^0$, $230^0$, $170^0$ and $190^0$. The error bars represent the mean +/- one standard error.*

Obviously, the effect of size will vary between groups simply due to the manner in which the groups were devised. However, the magnitude of this difference was pronounced. In the LSS group, the effect was not significant even at the .10 alpha level, with a mean difference of only 4.55 ms (which was opposite in direction to that attributable to normalization). In contrast, the effect was highly significant in the HSS group and the mean difference of 36.77 ms was consistent both in size and direction with a normalization strategy (F (1,14) = 70.13, p < .001). These effects of size and their consistency across the blocks in each group are visible in figure 9.5.

The two size sensitivity groups displayed remarkably different patterns of rotation sensitivity. In the LSS group, the effect of ROTATION was not significant, with a mean difference between rotation conditions of less than 1 ms. The BLOCK by ROTATION effect is only significant at the .10 level (F (3,42) = 2.33, p < .10). However, for the HSS group, this interaction is highly significant (F (2.63. 36.80) = 5.14, p < .01; W (5) = .40, p < .05). Similarly, there was moderately significant effect of ROTATION within this group, with a comparatively large mean difference of 36.77 ms, consistent with normalization (F (1,14) = 5.18, p < .05). Analyses for multiple representations were also performed using an ANOVA on the test block and the final train block for each size sensitivity group. The only significant effect of BLOCK between the two groups was in the LSS group (F (1,14) = 4.12, p < .10). However, this was only marginally significant and opposite in direction to that predicted by the multiple representation hypothesis. These variations in the effects of rotation between the two size sensitivity groups can be seen in figure 9.6.

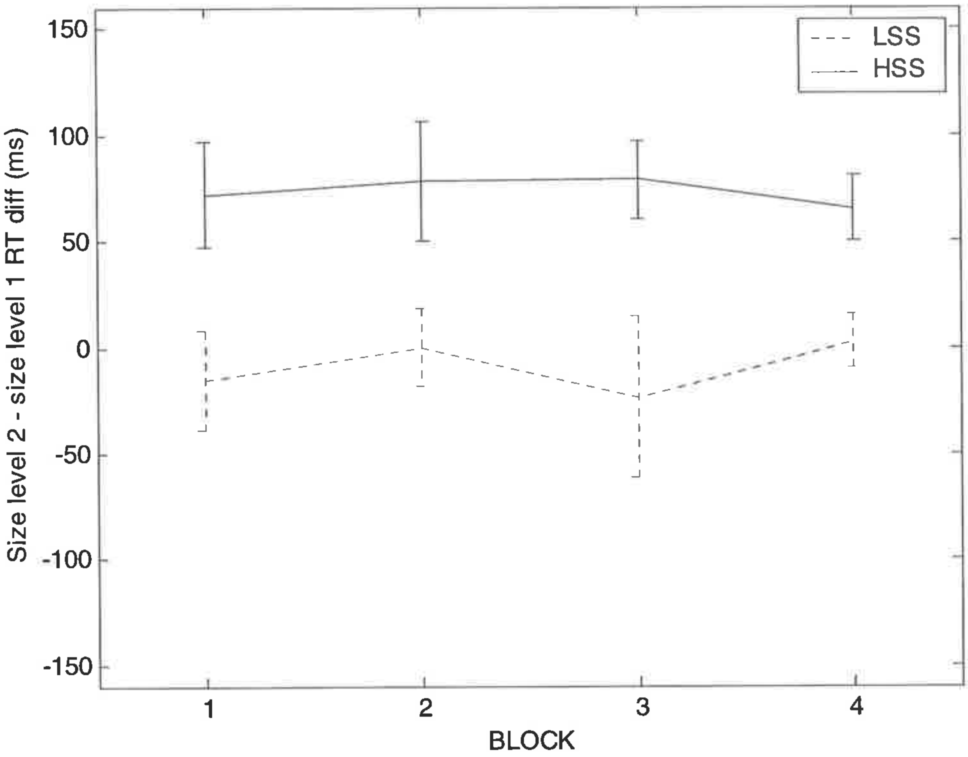


*Figure 9.5. Reaction time differences between the two size levels in the LSS (Low Size Sensitivity) and the HSS (High Size Sensitivity) groups for correct matching trials across the four experimental blocks. Size level 1 represents "smaller" scale ratios of the first image to the second of 5/4, 4/5, 3/2 and 2/3. Size level 2 represents "larger" scale ratios of 7/4, 4/7, 2/1 and 1/2. The error bars represent the mean +/- one standard error.*

To examine the relationship between size and rotation sensitivities further, a similar set of analyses was also conducted for the two groups based on rotation sensitivity. A measure of such sensitivity, $\text{rotation}_{\text{diff}}$, was formed by subtracting the mean response time for the smaller angular difference condition (*rotation level 1*) from the larger angular difference condition (*rotation level 2*) for correct responses to match trials. As with the measure for size differences, the sensitivity to rotation differed greatly between subjects, with a standard deviation of 30.66 (mean = 7.46 ms). The sample was divided into two equally sized groups in accordance with the

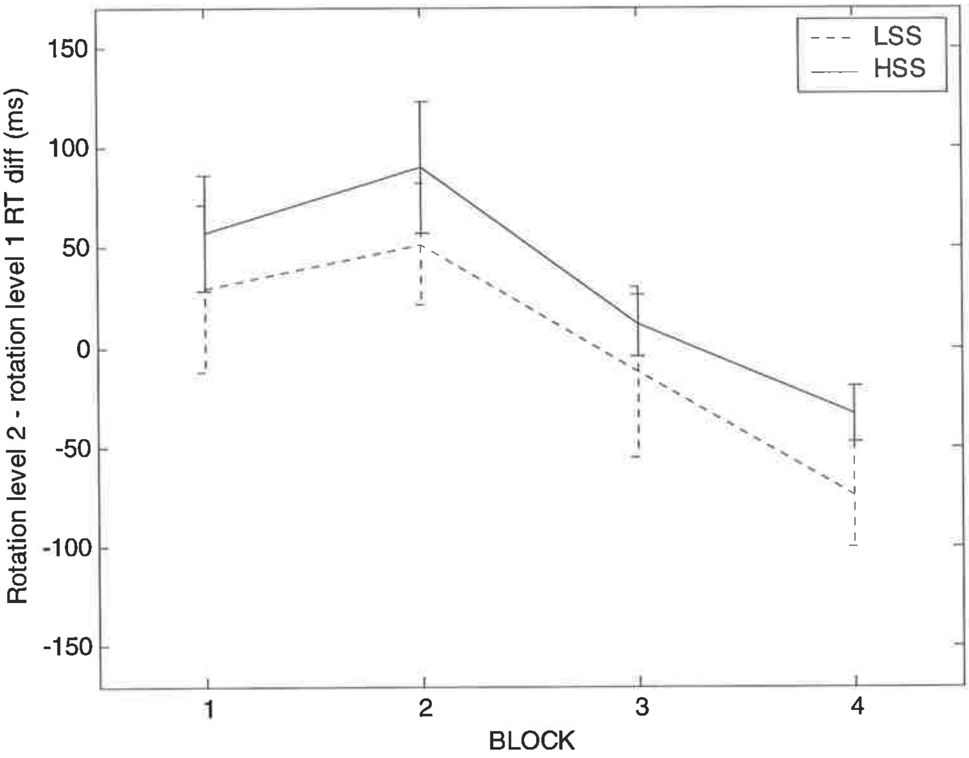


*Figure 9.6. Reaction time differences between the two rotation levels in the LSS (Low Size Sensitivity) and the HSS (High Size Sensitivity) groups for correct matching trials across the four experimental blocks.* *Rotation level 1 represents "smaller" angular deviations from the upright of 50$^0$, 310$^0$, 90$^0$, and 270$^0$. Rotation level 2 represents "larger" angular deviations from the upright of 130$^0$, 230$^0$, 170$^0$ and 190$^0$. The error bars represent the mean +/- one standard error.*

median score of rotation$_{diff}$ (7.26 ms) – a low rotation sensitivity (LRS) group and a high rotation sensitivity (HRS) group. The effect of SIZE was highly significant for the HRS group (F (1,14) = 10.40, p < .01) but failed to reach significance in the LRS group. The effect of ROTATION was highly significant in the HRS group (F (1,14) = 31.50, p < .001), with a mean difference consistent in direction with normalization of 26.90 ms. In addition, the BLOCK by ROTATION interaction was also highly significant (F (3,42) = 5.81, p < .01). However, in the LRS group neither the ROTATION by BLOCK interaction nor the ROTATION main effect was significant and the direction of the mean difference associated with the non-significant effect of orientation was opposite to that predicted by the normalization hypothesis (13.03 ms).

ANOVAs on the test block and the final train block for each rotation sensitivity group were also performed. There was a marginally significant effect in the tests for multiple representations in the LRS group ($F(1,14) = 4.12$, $p < .10$) but the direction of this effect demonstrated that mean latencies reduced between the third and fourth block. Therefore, as with the similar effect in the LSS group, the effect was inconsistent with the multiple representation + normalization hypothesis.

To further examine the association between rotation and size sensitivities, a correlation was calculated between $\text{size}_{\text{diff}}$ and $\text{rotation}_{\text{diff}}$. As mentioned previously, the standard deviation of the $\text{rotation}_{\text{diff}}$ measure was high and the distribution deviated significantly from normality (Shapiro-Wilk - $W(30) = .870$, $p < .05$). A Spearman rank order correlation was conducted between the two sensitivity measures, but failed to reach significance. However, the correlation coefficient was .206 which, according to Cohen's (1988) classification, equates to a small to medium effect size. The failure to find a significant correlation may have been due to inadequate power. A priori, an N of between 120 and 140 would be necessary to achieve appropriate statistical power (.8) for an effect size of such magnitude at the .01 alpha level.

## 9.4 Discussion

### 9.4.1 The Humphreys effect

The overall influence of orientation in the second experiment was remarkably similar to that demonstrated in the first – the effects of orientation were attenuated with practice on the match trials, a phenomenon which we have previously dubbed a 'trend to invariance', and there was no evidence of normalization along this dimension for foil trials. In addition, the 'trend to invariance' was not due to the deployment of multiple orientation-specific representations but was attributable to orientation-invariant recognition. The only major difference in the results for Experiment II with respect to the influence of orientation was the unusual effects of rotation in the test block. We believe that this is not the product of a different processing strategy but that changes in the design of the second experiment allowed us to examine different elements of the same perceptual mechanisms.

The phenomenon in question is the 'reverse' effect of rotation in the test block. This occurred immediately after the final train block, in which no effects of orientation were evident. This trend was not evident in the first experiment and we propose that the explanation for this lies in the variation in angular deviations used between the two experiments. To increase power in the second experiment, the angular deviations were altered. For the testing trials, this involved changing the magnitude of the angular deviation from the upright from $150^0$ to $170^0$. In both experiments, versions at $180^0$ were excluded in an effort to avoid the unusual effect in identification experiments often found at this angle (see section 5.5). Given the design of our experiments, the inclusion of this angle and the corresponding effects in *reducing* reaction times could have obscured the normalization effects – the presence or absence of which was of primary concern to us. But it is likely that the reduction in response latencies at $170^0$ orientations is in fact the same phenomenon as that occurring at $180^0$ in other experiments. In other words, the M-shape effect demonstrated in previous studies is more broadly tuned than anticipated. Note that in previous experiments demonstrating the M-shape effect, angular variations of $170^0$ had not been included. It seems that, while attempting to avoid the interference of what was considered, a priori, to be a limited effect, we may have in fact examined it in more detail. This serendipitous discovery may provide a further clue into the underlying perceptual processing.

In this section, we propose that an effect that we will call the *Humphreys effect* is responsible for the unusual influence of orientation in Experiment II as well as M-shape effects in other experiments in the literature. In a series of experiments, Humphreys (1984) demonstrated a specific advantage in shape recognition under a successive pair matching paradigm that may account for these unusual effects. This advantage was apparent when the principal axes of the two shapes were in the same orientation regardless of the top-bottom direction of the shape (see section 4.2 for a description of the principal axis). Humphreys (1984) compared the effects of matching two identical geometric shapes under three conditions: a) where the orientation of the principal axis of the shape was maintained, but the relative locations of the focal features were altered, b) where the orientation of the principal axis was changed, but the relative positions of the focal features were altered and c) where both orientation of the axis and the relative positions of the focal features are altered. These conditions were all equivalent to either a rotation in the plane, a rotation in depth, or a combination of both of a single planar (two dimensional) shape. In Experiments 1, 2, 4 and 5, there was an advantage in matching performance under condition a). This amounts to a benefit in matching in shapes that have undergone either a $180^0$ rotation in the plane or a $180^0$ rotation in the picture plane and a mirror-reflection in the image's own principal axis (equivalent to a $180^0$ rotation in depth around the principal axis). The advantage conferred on these versions of the shapes cannot be explained by a conventional normalization to match hypothesis because these versions were more quickly matched than other versions where the shortest angular deviations to match were smaller.

To account for this effect, Humphreys (1984) suggested that subjects were generating axis-based shape descriptions such as those postulated in Marr's (Marr, 1982; Marr & Nishihara, 1978) theory of visual processing outlined in section 4.2. Humpreys (1984) suggested that the response time to match was reduced under condition a) because the 'orientation' of the principal axis did not need to be recalculated. It may be that the generation of this pictorial axis was facilitated because the axis from the previous image remains in memory (almost like an after image although at a higher processing level) when the new image was presented. In our experiment, we propose that we witnessed a similar advantage when the images of a pair were separated by a $170^0$ rotation – the principal axes were very closely aligned and this may have facilitated the derivation of this component of the stimulus representation. In other words, there was a computational advantage when the first and second pair of images were separated by a $170^0$ rotation. Figure 9.7 provides an illustration of this axis-based effect.

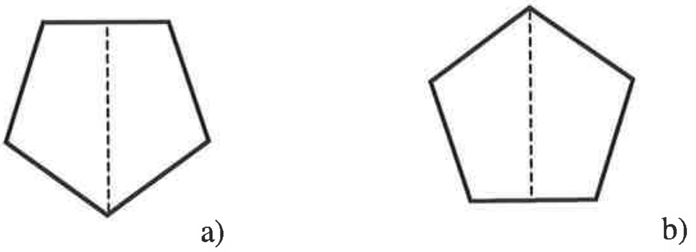


*Figure 9.7. An illustration of the 'Humphreys effect'. The two images, a) and b), represent the same shape rotated by $180^0$ in the picture plane. The bold line depicts the overall shape contour while the dotted line represents the principal axis derived from this contour. Note that in both cases the image of the axis is identical.*

The exact influence of the 'Humphreys effect' in our experiment is difficult to determine. While the computational account is probably most compelling, there is also some evidence for an explanation based on response bias. The near constancy of axis orientation may have created a bias towards a match response. Whenever the shapes were presented at $170^0$ from the upright, this could have produced an increase in perceptual similarity associated with the axis-based structures that biased the subjects towards producing a match response regardless of whether the two shapes matched or not. Note that this explanation involves an overall response bias and not necessarily the involvement of the principal axis at a computational level. This notion is partially corroborated by the error rate data which demonstrated that, for the $170^0$ versions, there was an increase in accuracy for match trials but a decrease in accuracy on foil trials for these versions. In other words, subjects tended to respond as if the two images had the same shape whenever the second image was oriented at $170^0$ regardless of any shape similarity. Perhaps the near 'alignment' of principal axes caused an increase in sameness information that, in turn, produced a response bias.

An explanation in terms of a recognition mechanism does not take into account the error rate and response latency data from both match and foil trials. However, the error rate findings in our experiment were not entirely compelling support for the bias account. Firstly, the increase in errors for the $170^0$ versions in the foil trials was preceded by a similar increase in error rate for versions $130^0$ from the upright anyway. Therefore, the error rate effect may not be specifically related to the $170^0$ versions. Secondly, the magnitudes of all of these error differences were small – between 1% and 3%. Such evidence suggests that the error rate differences may be unrelated to the 'Humphreys effect' but may be better attributed to experimental noise. The bias account was also rejected by Humphreys (1984) because he failed to find any simultaneous error patterns in match and foil trials to match the RT differences in his experiments. The exact cause of, and the conditions necessary for, the influence of principal axis effects in tasks involving planar rotations is worthy of further investigation.

Despite the fact that the 'Humphreys effect' is associated with orientation specificity (i.e., an effect at or near $180^0$), Humphreys (1984) argues that the effect is likely to be the result of an orientation invariant recognition mechanism. We believe our experiments and review have demonstrated some support for this hypothesis. Firstly, there is an association between the 'M-shape' response latency profile and experimental conditions that are conducive to the development of an orientation invariant profile with practice (see section 5.5), i.e., not mental rotation experiments but tasks that are free from MRFs. Secondly, in our experiment, the effect was directly associated with orientation invariant processing - it occurred after the final train block in which no effects of rotational sensitivity were found. However, in section 9.4.5, we will outline a model of processing where the 'Humphreys' effect is available to both orientation-sensitive and insensitive mechanisms.

### 9.4.2 Zoom effects

The influence of scale in Experiment II was not only statistically independent of the effects of orientation but was also qualitatively different. We believe that the effects of size were in fact the result of the zoom lens of attentional deployment. Firstly, the effects of size and rotation did not interact which is necessary for an account based on qualitatively different, independent perceptual mechanisms, i.e., one involving recognition and one involving attention. Secondly, the size-disparity effect was consistent throughout the course of the experiment. There is evidence from previous experiments that, regardless of type of task or foil, size effects which are attributable to a transformation to match procedure, diminish with practice (see section 7.10 and chapter 8). Such a pattern would parallel the findings with regards to orientation in this and other experiments. However, practice did not exert any appreciable influence on the effects of size differences as it did for differences in orientation. If both rotation and size effects were associated with shape-specific normalization, it would be reasonable to assume that similar experimental variables, e.g., practice and foil type, would exert a similar influence on the effects of the two variables even if the two transformations were carried out quite independently. In fact, in chapter 7, we outlined several similarities between the simple effects of the two transformations in the experimental literature on recognition behaviour.

Thirdly, the effects of size were not only consistent throughout the experiment but were similar for both match and foil trials. Unlike rotation effects which were limited to match trials, the influence of size involved a kind of 'blanket' effect that affected all responses. In section 7.12.1, we provided evidence for the notion that normalization for size was evident for foils only when the foil images were the same objective shape as the match image or otherwise highly confusable with the match stimuli (e.g., Jolicoeur & Besner, 1987; Larsen, McIlhagga and Bundesen, 1999 [Simultaneous matching condition only]). In contrast, the foil images used in our experiments were entirely different, independently generated objective shapes. As a result, we would expect results similar to many other recognition studies where the foil images also depict entirely different objective shapes and where these foils are not easily confusable with match images. In such studies, significant scale effects are shape specific, i.e., evident in match but not distractor trials (Besner, 1983 [Experiment II]; Besner & Coltheart, 1975, 1976; Biederman & Cooper, 1992 [Experiment II]; Corcoran & Besner, 1975; Milliken & Jolicoeur, 1992 [Experiment I]; Santee & Egeth, 1980 [Experiment I & III]). More specifically, we anticipated comparable results to Besner's (1983) second experiment, at least at the start of the task, given the many similarities in design and stimuli, i.e., a paired stimuli, same / different matching task using abstract, closed polygons and entirely different shapes for foil images. This would involve shape-specific normalization on the match trials only (i.e., a slow transformation rate) and perceptual scale effects (i.e., a fast transformation rate that is much harder to detect) evident only on the foil trials.

We propose that the consistent size effects over trial type and blocks in our experiment represent only the activity of the zoom lens attentional mechanism. Unlike size normalization, an attention based scale mechanism would explain the equivalent influence of size on both types of trial and the persistence of these size effects despite practice. This mechanism involves preliminary stimulus encoding and is separate from later, higher levels of processing such as recognition. If shape specific normalization was at work then it would be evidenced by a relatively larger size discrepancy effect in the match trials as evidenced in the studies listed above. Thus, we conclude that normalization to match for scale differences was not present or not detectable because shape specific size effects were not present in the experimental data over and above the non-shape specific attentional effects.

The relatively large N used in this experiment may have enabled the detection of the effects of attentional scale which in other tasks may have been undetectable – in this task, the average difference between size conditions was no more than 18 ms. In section 7.14, we suggested that the failure to find evidence for the zoom lens theory may be at times attributable to low statistical power and there is some support for this notion in our experiment. In Experiment II, there was a high degree of between subject variability with respect to size sensitivity. Therefore, the variety of findings with respect to the scale effects in previous empirical studies may be the product of relatively low Ns – most studies are taking relatively small samples from a diverse population of scores. However, it may be that other experimental variables (e.g., stimuli and timing) influence the nature of attentional deployment between studies. The use of scale specific attention in matching and recognition tasks is worthy of further investigation.

### 9.4.3 General transformational vs. general invariant recognition mechanisms

In Experiment II, the independence of the effects of rotation and size were consistent with findings from a number of other matching tasks (e.g., Bundesen et al., 1981; Kubovy & Podgorny, 1981; Sekuler & Nash, 1972). In other words, there was no evidence of a single transformational mechanism that simultaneously alters the stimulus image along both dimensions. Even after repeated presentations there was no interaction between the effects as was evident after practice in Larsen's (1985) study. It may be that the difference lies primarily in the choice of distractors – Larsen (1985) used MRFs while we employed entirely different random polygons. In Larsen's (1985) study, the use of MRFs, the emergence of this interaction effect only after practice and the unusual nature of this interaction effect (see chapter 8) are all consistent with the formation of multiple orientation-, size- and handedness-specific representations. We found no evidence of such representations in our experiment. It may also be that, as discussed in chapter 8, Larsen's (1985) findings are due to an experimental artefact that was absent in our findings. This speculation is worthy of investigating further.

However, as we have already pointed out, we believe that the significant effects of size in our experiment were the result not of normalization but of an attentional mechanism. If this is correct, the effects of this mechanism and that of normalization would be independent and this may be what is evident in our results. Therefore, our ability to evaluate the notion of a single transformational mechanism is hampered by the failure to find shape specific size effects, i.e., normalization for size for the purposes of recognition. As mentioned previously, size normalization effects were not detected in our experiment. If our experiment lacked the statistical power to detect these simple effects of size then it is likely to have also lacked the power to detect any interaction between size and rotation normalization. Therefore, our experiment may not have been a rigorous test of the interaction or independence of size and rotation normalization.

One possible reason for the failure to find size normalization effects is an accelerated function for all shape-specific transformations. The rate of normalization for orientation apparent in Experiment II was incredibly fast, with a mean rate within the first two blocks of 2983 deg/s. This is markedly faster than the fastest rate in Experiment I of 1751 deg/s in the first train block and quicker than any significant normalization rate reported in previous literature. Indeed, based on their implicit or explicit criteria for acceptable rates, it is likely that authors such as Cohen and Kubovy (1993) and Tarr and Pinker (1989) would reject such rates as the effects of normalization for orientation. We believe this represents normalization and one of the reasons for holding this belief is that the attenuation of the effects with practice demonstrated in our study parallels that of other studies where normalization rates are generally slower. However, if the rate of normalization for rotation was comparably faster in this experiment, then it is not unreasonable to expect the corresponding function of size transformation to have been similarly accelerated. This assumption is independent of whether these two transformations are carried out by the same system because other experimental variables, such as stimulus complexity or other task variables, may exert a similar influence on rates along both dimensions. Therefore, shape specific size effects may have been too small to detect accurately.

In addition, not only might the normalization rate have been very fast at the start of the experiment but these effects may have attenuated within the first block making them even more difficult to detect. Note that even with MRFs, which are not strongly associated with 'invariant' matching effects, Larsen (1985, Experiment II) found that practice increased the normalization rates for both rotation and size scaling. Despite several similarities, one of the most important differences between this experiment and that of Besner's (1983) is that we repeatedly presented the same stimuli. Previous research suggested this should attenuate shape specific normalization effects. The problem of detecting such effects may also be exacerbated because such effects would be demonstrated by an increase in overall rates in the match trials over the foil trials. With only two levels of size disparity in any one block, the rates demonstrated will not be as accurate as those derived from functions comparing latencies from many more transformations. Given these qualifications, and the size normalization effects in section 7.2, we have reason to be suspicious about the failure to find shape-specific effects.

This discussion highlights one of the limitations in comparing the relative influences of the two transformations - namely, that it is impossible to design an experiment with 'equivalent' effects for the two variables. Simply because only one of the variables is exerting a statistically significant influence does not guarantee that normalization is occurring along only this particular dimensions. At any combination of size and rotation transformations, it is unlikely that each variable will exert the same influence on RT as measured by effect size. As a result, only one may be significant by virtue of a larger effect size even though the other variable may still be exerting an influence on perceptual processing. Therefore, even if there were underlying qualitative similarities in the processing of the two variables, this would not necessarily be reflected in the pattern of significant results. We will discuss these issues at greater length in chapter 10.

Therefore, while there is no prima facie evidence of a parallel 'trend to invariance' for both size and orientation, it is possible that such shape specific size effects may have been undetectable in our experiment. One way in which this could be evaluated is by replicating the study with a larger N and with an additional condition using MRFs. Certainly, there is no shape specific sensitivity to either size or rotation transformations at the end of the experiment. This is consistent with a recognition strategy that is invariant to both these properties such as those outlined in chapter 4. Rather than assume that our experiments provide the definitive 'result' we have tried to assess our findings in the light of previous literature and other considerations such as statistical power. However, in the next chapter we will provide alternative explanations for findings in such tasks and question some of the fundamental concepts implicit in this interpretation. This involves a critical evaluation of the psychological techniques commonly used to study object recognition.

### 9.4.4 Dependence of effects

While the overall group analysis did not reveal any relationship between size and orientation effects, there was limited evidence for a relationship between the two variables at an individual level. These effects were evident in patterns of significance between separate ANOVAs for groups based on size and rotational sensitivity as well as mean effect sizes for simple effects and correlations. Obviously, one must be cautious about inferring from non-significant results and the raw effect was only small to medium in magnitude. However, the converging evidence from different analyses suggests that the effects may be valid but that the experiment lacked the statistical power to find these effects significant. The lack of power in many comparable matching and naming experiments has already been noted by Jolicoeur and Landau (1984).

At the very least, the experiment provides evidence that individual differences in such tasks merit further investigation. Only McKone and Grenfell (1999) have looked into this issue. In their study, they found individual differences in strategy – while some subjects used an orientation-invariant recognition procedure from the start of the experiment, others developed this only after practice. Although there was variation in orientation sensitivity in our experiment, we found stronger evidence for individual differences in the effects of size and no evidence of practice affecting the use of size-sensitive or -insensitive strategies. Taken together with McKone and Grenfell's (1999) findings, it seems that the notion of individual differences in perceptual and recognition mechanisms, and the relationship between size and rotation effects, requires further investigation. Indeed, this issue seems to have been neglected and may be a key to understanding the variety of experimental findings. However, as the power estimation to detect significant correlation effects demonstrated, the size of N may be prohibitively large.

If the effects of rotation and size possess some common variability at the individual level how might this be caused? We have proposed that the RT effects associated with the two transformations are the product of two qualitatively different mechanisms, i.e., a recognition procedure and an attentional strategy. Perhaps the relationship is the result of a broader influence on perceptual functioning. Performance, in terms of the speed of the different systems that perform such operations, may reflect a general perceptual processing ability that varies between individuals. Both orientation sensitivity in recognition (McKone & Grenfell, 1999) and size specificity of the zoom lens (Pasto & Burack, 1997) vary between individuals. Some of this variability may be accounted for by a kind of general visual processing ability. This may be similar to Spearman's *g* for intelligence – an overall efficiency of processing ability that, for any one person, influences performance on a number of different tasks (Spearman, 1904).

One particularly acute problem in interpreting these results is that we are trying to infer idealized or individual behaviour from group results. As is the norm in such experimental designs, we combined data from several subjects to reduce measurement noise. However, given the individual differences that we know or suspect exist, the combined results do not represent a single uniform perceptual process extracted by summation but a summary of group performance. We are not achieving a picture of 'the' underlying perceptual process because there is evidence that such a thing does not exist. What we have is a summary of individuals that does not reflect the activity of any one individual just as the learning curve does not reflect the progress of one person. Therefore, the individual differences and the statistical averaging process may complicate our interpretation of overall results.

### 9.4.5 An outline of a unified model for the recognition of planar rotated and / or scaled images

#### *9.4.5.1 Model description*

In this section, we will describe a model that can account for results in both experiments of this project. We will also discuss how this general model, which is depicted in figure 9.8, can account for a variety of other findings in the experimental literature. The model is algorithmic in nature and incorporates several novel ideas. It also includes important elements of previous ideas from the literature and, in particular, those of Marr (Marr, 1982; Marr & Nishihara, 1978) and Ullman (Huttenlocher & Ullman, 1987; Ullman, 1989; Shoham & Ullman, 1988). For example, it features general notions of perceptual processing stated by Marr. These include multiple stimulus representations that become increasingly more invariant to features such as orientation and size at higher levels of processing, feedback between processing stages and shape descriptions that are invariant to scale and orientation information. More specifically, it also features Marr's notion of an axis based, 'stick-figure' shape representation scheme. The model also includes a transformation to match mechanism based on Ullman's various alignment schemes. In summary, the model may be considered a synthesis of new and old ideas on object recognition.

The long term memory storage of an object is handled by two separate systems – one which contains size and orientation free representations (*INV store*) and one which contains size and orientation specific representations (*DEP store*). Initial encoding of the stimulus image results in a concrete, pictorial representation of the image contours. This process is influenced by various preliminary mechanisms

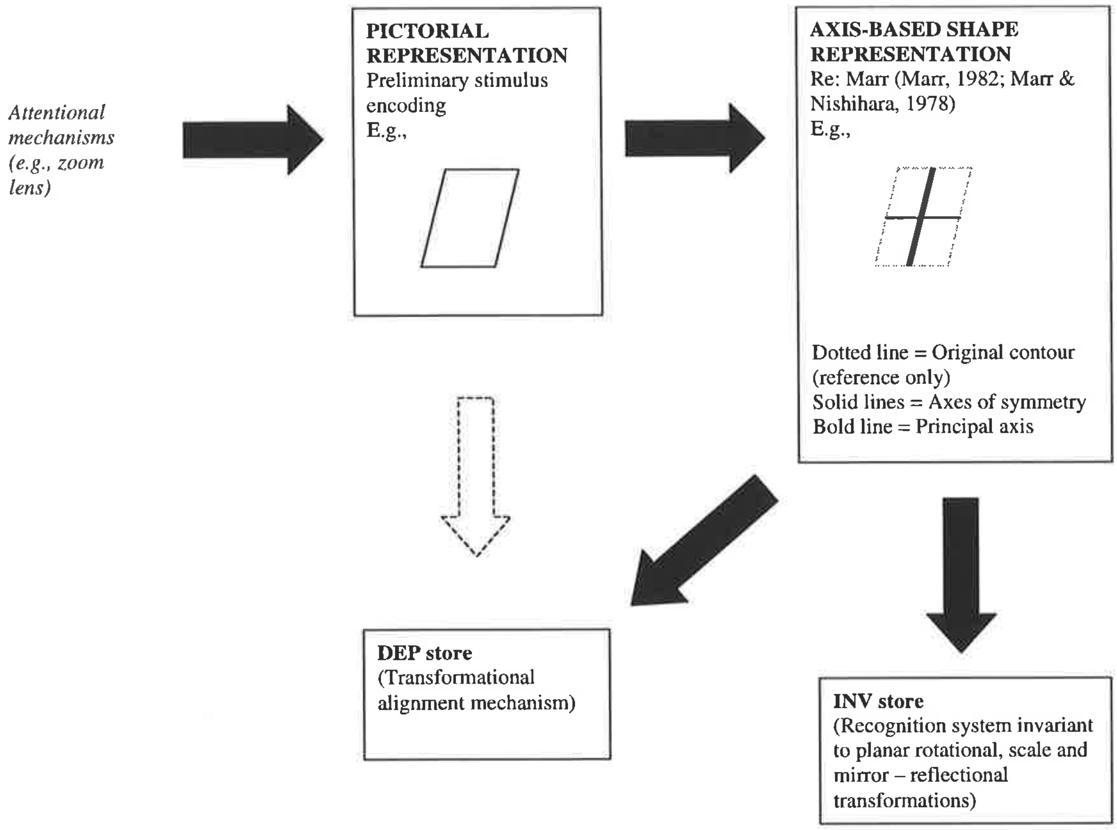


*Figure 9.8. A flow diagram of the proposed unified model of perceptual processing.*

including attentional procedures (e.g., the zoom lens of attentional deployment) and other processes such as image segmentation. The next stage in the flow of processing is the derivation of an axis-based shape representation. This is identical to the representational style described by Marr (1982; Marr & Nishihara, 1978) and outlined in section 4.2. It consists of axes defined by characteristics such as elongation and symmetry that are derived from the image in a low-level fashion. One of these axes is the principal axis (see section 4.2). In other words, the derivation of this representation does not require specific object knowledge (i.e., any preliminary recognition of the shape), but is formed by purely bottom-up, data driven procedures.

Two parallel streams of processing are instigated at this point, both of which are based on the same axis based shape description described above. These streams of processing are largely independent and are capable of initiating a response to the task requirements (e.g., match vs. foil response and triggering a semantic label). Both ‘compete’ in time to be the first module to provide a match. The ‘winner’ instigates the behavioural response. The first stream of processing involves recognition of shape using the stored representation in *DEP store* and involves an alignment procedure such as those outlined by Ullman and colleagues (Huttenlocher & Ullman, 1987; Ullman, 1989; Shoham & Ullman, 1988) previously discussed in chapter 3. Such catalogued representations are orientation, scale and handedness specific. The first step involves the derivation of an alignment key. This key consists of at least three elements of any combination of points or lines as outlined by Shoham and Ullman (1988). The derivation of this key and subsequent search for matching keys in memory would be relatively simple due to the economy of representation – rather than having to derive points and lines from a pictorial representation, an axis based structure reduces the ‘search space’ of elements that could serve as a possible key. If a matching key or keys is found, the transformation to align the two keys is calculated. This transformation is then performed to bring the two depictions into alignment and then a match is tested for.

As suggested by Edelman and Weinshall (1998), the actual transformational device is of limited capacity. This means that that the degree of transformation is reflected in RTs and, if the system is required to transform a large amount of image information, error rate. The transformations of planar rotation and size are not carried out simultaneously - either transforming along each dimension one after the other or more likely by alternating between smaller steps of each type of simple transformation as suggested by Bundesen et al. (1981). Interestingly, although this algorithm is based on the work of Ullman, Marr also hinted at a transformational process based on the alignment of stick-figure representations for 3-D recognition. He suggests that the simplest way of testing for the compatibility of the image and a stored or catalogued model is by "a relaxation process that adjusts the orientation of the [Principal] Axis incrementally, seeking the disposition for which the projections of the angles between the component axes of the catalogue model ... best agree with those of the stick figure" (Marr & Nishihara, 1978, p. 290).

The other stream of processing involves representations invariant to the similarity group of transformations that are also extracted from the axis-based depiction of the stimulus image. This stimulus representation is then compared with orientation and size free representations in a similar format housed in *INV store*. As with the alignment mechanism, processing in this stream is also simplified by using an axis based representation and not a more pictorial one. The type of information contained in this 'invariant' representation could take a variety of forms including a complex log mapping of the axis-based figure (see section 4.4.2). It could also involve a simple analysis and indexing of properties such as length ratios and angular deviations between axes both of which are invariant to overall scale and orientation and easily derivable from a stick figure representation of a planar object.

The ease with which a match is found in *INV store*, i.e., the amount of time it takes, is a function of familiarity of the object in question. This may involve a more efficient representation of the shape in *INV store*. In essence, repeated presentations can prime the representation of a particular shape so that the response of the *INV store* system to subsequent testing of this effect will be accelerated. The ease with which a match is found in *INV store* is also a function of the time delay between shape presentations. The image processing in *INV store* can continue after a response is given by the *DEP store*. This background processing can act to prime the shape representation in *INV store* and, as mentioned above, may form a more efficient shape representation. As a result, even though the initial presentation of an image may be recognized by the *DEP store* first, continuous processing in the *INV store* module after this response can prime the orientation, scale and mirror-reflected image for subsequent testing. One reason why the system may effectively strive for an 'invariant' representation is that this will facilitate later processing and eliminate the need for a taxing and time consuming image transformation procedure in the recognition process.

Finally, we have also included a direct link (depicted by an arrow consisting of dotted lines) between the pictorial representation and the *DEP store*. This link involves an alignment system like that outlined in Ullman (1989), where pictorial representations are transformed to match. This may involve the use of alignment keys, which are based on three or more points in the stimulus image, or a simpler, less constrained approach, such as the simpler orientation alignment scheme outlined in Ullman (1989). This latter method enables the use of clearly defined orientation cues to perform the rotation instead of alignment keys. Overall, this pictorial alignment procedure enables a more complex and detailed stimulus matching procedure than the axis based account, which sacrifices some imagery detail in exchange for economy of representation. However, due to the added complexity of the image, both the search for an alignment key and the rate of transformation are considerably slower. Therefore, it is a more taxing and time consuming operation than alignment using the more abstract axis-based image descriptions.

#### *9.4.5.2 Accounting for the results in Experiments I and II*

This model can explain the findings of our two experiments. At the beginning of Experiment I, stick figure representations are derived from the pictorial representations. Although subsequent processing occurs in both *INV store* and *DEP store*, the alignment mechanism associated with *DEP store* is initially quicker than recognition by *INV store* and RTs will reflect this transformation to match procedure. However, with practice on the matching task, shape representations specific to the shapes repeatedly used throughout the experiment become primed within the *INV store* system and responses to such match stimuli are consequently faster. As a result, RTs reflect the activity of this system, i.e., a flat function of RTs across orientations. For the foil trials, responses may be invariant across all angular deviations for a number of reasons. Firstly, the alignment mechanism fails to find a matching key for the object (and hence no transformation is applied and a "foil" response is given). Secondly, the *INV store* may also fail to find a representation. Thirdly, a time limit may be reached by which point neither mechanism has found a match. Such a limit would always be independent of the orientation of the object but may alter during the course of the experiment (e.g., progressively reduce with practice and confidence in the stimulus set).

In Experiment II, the effects of the zoom lens mechanism on initial stimulus encoding are evident in RTs because of the size differences between stimuli. Because it operates on initial stimulus encoding, it affects RTs to all images, shape and foil, and is independent of subsequent recognition strategies. Following the derivation of an axis-based shape description from the initial stimulus encoding, processing again occurs in parallel across the two recognition systems, *INV store* and *DEP store*. Initially, the latter system provides a faster response than the former. However, with repeated practice the 'invariant' representation of the match shapes become primed and recognition responses accessing these representations become faster so that, on average, the *INV store* system provides the matching response before *DEP store* has completed processing. Note that this explanation rests on the assumption made above that normalization for size effects were small and undetectable in this study. The decrease in overall 'normalization rate' may be caused by one or both of two factors. Firstly, individual differences between subjects in efficiency of either representation style may produce differences in switching between them. For example, a subject who possesses a relatively more efficient *INV store* system may switch to this representational style more readily. Secondly, there may be noise in the system causing subjects to alternate between the two procedures with increased probability of shape specific *INV store* dominance with practice.

Finally, when the two successive images are separate by a $170^0$ rotation, there is an influence of the Humphreys effect at the axis-based representational level. The first image of the pair always leaves a stimulus trace of its various axes at this stage. When the second image is processed, the derivation of this stick figure representation is facilitated and / or a bias towards sameness is produced because the central component of the shape depiction, i.e., the principal axis, is very close to the same orientation in the two figures.

#### *9.4.5.3 Accounting for the results in previous studies*

Any valid model must demonstrate the ability to generalize to a variety of findings. In addition to accounting for the effects in our experiments, the proposed model has the capacity to explain other phenomena in the literature. In what follows, we will briefly describe how the model may account for such findings.

(1) The model accounts for the influence of same shape distractors in recognition tasks under both planar rotational and scale transformations alluded to previously in sections 5.6 and 7.12. The *INV store* system is ineffectual in tasks where the foil images share exactly the same objective shape – i.e., when the match and foil image are differentiated only by transformations of the similarity group. *INV store* cannot discriminate between the two types of trials because, in the eyes of this system, all the images 'match', i.e., the information it is deriving is the same for both match and foil images. Therefore, processing is always the responsibility of the *DEP store* module. A similar explanation can also account for the use of normalization on foil trials when the foil and match images are highly confusable (e.g., Jolicoeur & Besner, 1987; Larsen, McIlhagga & Bundesen, 1999). The two forms are easily confusable as having the same objective shape and the *INV store* system cannot discriminate between the two images in the pair on the basis of the 'invariant' information initially derived. The *DEP store* system dominates processing because it

is capable of finer discriminations between forms, such as when they are highly complex images (see (6) below).

Note that the *DEP store* system can still function to recognize shapes in the absence of MRFs and when the forms used in the experiment are not confusable. With practice however, *INV store* would gradually come to dominate responses. Such an account is relevant to our own experiments (as mentioned above) and to a variety of other tasks showing attenuation of the effects of transformation with practice under such conditions (see sections 5.6.2, 5.7, 7.10 and 7.12.2).

(2) In addition to allowing for the influence of foil images, the model also accounts for the orientation sensitivity evident when tasks involve left-right direction assignment free from MRFs (e.g., Jolicoeur, 1988). In this case, the *INV store* representation is ineffectual because it does not contain information on handedness necessary for the task. As a result, the system must use the representation of the shape in *DEP store* and this requires use of the alignment procedure.

(3) The M-shape effect (section 5.5) can be accounted for in the same manner as the Humpreys effect noted above – namely, by a correspondence between principal axes of the current and former stimulus traces. This is because the advantage or bias occurs before either *INV store* or *DEP store* has processed the image so the Humphreys effect can be evident regardless of the recognition procedure used (i.e., *INV store* in Experiment II or *DEP store* in, for example, Eley (1982)).

(4) The individual differences in orientation sensitivity demonstrated by McKone and Grenfell (1999) may be attributable to differences between subjects in the relative efficiency of the two recognition modules. For example, some subjects may possess an inherently more efficient (= faster) *INV store* module and this system may be more likely to 'win' the race to provide a response even when confronted with relatively unfamiliar images.

(5) Because priming of a shape representation in *INV store* is not only a function of practice but also ISI (or effective ISI when presentations are not consecutive), the model accounts for the attenuation of orientation (section 5.7) and size (section 7.10) effects with practice as well as the influence of longer ISIs and priming experiments in invoking 'invariant' effects (sections 5.14, 7.7 and 7.9).

(6) The model can also account for the influence of stimulus complexity on transformational sensitivity to size (section 7.11) and orientation (section 5.12). In tasks where the images are complex and confusable (e.g., Tarr & Pinker, 1989 [Experiment II]), processing is limited to *DEP store* because the abstract invariant representations in *INV store* cannot discriminate between images. In fact, in cases where the axis-based representation itself is too abstract to distinguish between the objects (e.g., faces), alignment may be employed using only the pictorial stimulus representation. This results in extreme 'sensitivity' to orientation and size in the recognition of such stimuli – the large amount of information contained in the pictorial format taxes the transformational system that, as mentioned above, is a limited capacity device. This produces an increase in error rates associated with the

degree of transformation that has been exhibited in studies using such stimuli (e.g., Kolers et al., 1985; Rock, 1973).

(7) As mentioned above, *DEP store* is used exclusively whenever the match and foil images are separated by only a shape preserving transformation or when the two types of stimuli (or simply the objects to be named) are otherwise highly similar. However, after repeated presentations the *DEP store* system will begin to form multiple representations of the stimulus images that are orientation, scale and handedness specific. This facilitates processing because it reduces the level of time consuming and taxing transformation needed in the alignment process. The model can therefore account for the use of multiple representations in Tarr and Pinker's (1989) tasks, as well as their suspected use in experiments such as Larsen's (1985).

(8) The use of an axis based representation means that the images processed do not have to be pictorially identical to invoke the same recognition response – what is derived is essential information on basic shape structure. This would be advantageous in a real world recognition task because it allows for generalizabilty in recognition, in accordance with schematic shape structure, and not simply template matching. In addition, the use of axis based representations could account for the findings of Leek (1998), where experience in seeing real-world objects in various orientations transfers to orientation invariant recognition for pictorial depictions of similar objects. In this case, training has primed the representation of the objects in *INV store* and this representation is useable in the task involving line drawing pictures of the same objects.

(9) The parallel nature of computation in the model can account for findings attributable to background processing in Jolicoeur's (1988) study, outlined in section 5.10. According to the first account we provided, practice on a handedness task invoked the representations in *DEP store* for the purposes of recognition. Processing of shape also occurred in parallel in *INV store*, although such representations were not useable in the task due to their handedness invariance. However, when the task switched to one of identification, the representation in *INV store* was accessed and performance was immediately invariant across orientations. Alternatively, the model can also account for the other explanation given in section 5.10 – namely, that the facilitation in matching is the product of the formation of multiple stimulus representations, that may be only partially formed in the handedness task, but are adequate to facilitate processing at the same orientations in the identification task.

(10) Because the use of *INV store* does not require experience of a shape in various orientations and sizes, it can explain the results of Jolicoeur and Milliken (1989) where orientation invariance is available for shapes that have been repeatedly presented in only one size and scale. In this case, the representation in *INV store* is primed through repeated presentations. However, this tendency towards using an invariant strategy was only evident within a context of poly-oriented objects in Jolicoeur and Milliken's (1989) study (see section 5.9). This requirement suggests that priming can also be influenced by higher level functioning which decides that there is a likely future requirement of recognition of the shape in various orientations and sizes.

#### *9.4.5.4 Conclusion to model*

This hybrid model, although tentative, can explain a variety of experimental effects in a range of different studies. It makes many assumptions about processing that have not been empirically verified but, in this respect, it also provides a framework for further investigation into the recognition of rotated and scaled objects. However, it may be considered a conventional model. It is couched in the concepts that are commonplace within the experimental literature – i.e., that transformations which are performed are part of the recognition process, that performance unaffected by size and orientation variables represents recognition invariant to such stimulus properties, and that the activities of such mechanisms are examinable in conventional experimental paradigms via latency and accuracy measures. Such assumptions may be considered fundamental tenets of a particular scientific paradigm (Kuhn, 1962). In the next chapter, we will question these basic assumptions. Some of the difficulties in interpreting data in a parsimonious way, not only in the current project but also in other studies, may represent challenges to the conventional experimental psychology paradigm used to investigate object recognition.

# CHAPTER 10: Conclusion

*This chapter discusses the problems with conventional psychological approaches to understanding the mechanisms of object recognition. In particular, insight from response latency and error rate measures may be inadequate to reflect the complexity and parallelism of brain functioning. We propose a range of further experiments that reflect a multidisciplinary perspective on the problem at hand.*

## 10.1 Introduction

Like the other psychology experiments we have reviewed, our own experiments used the traditional dependent measures of reaction time and error rate. However, such measures have particular limitations in what they can tell us. It is unlikely that such measures do justice to the complexity of brain functioning and often provide only limited insight into the underlying mental processing. In this project, we have attempted a multidisciplinary approach to the issues at hand by examining not only traditional psychological tests, but also explicit computational modeling, data from studies in neurophysiology and neuropsychology as well as broader philosophical implications.

A good example of how this multi-faceted approach was employed in this project is the question of the relationship between identification and transformation procedures raised in chapters 2 and 3. If the performance measures of reaction time or error rate are proportional to the degree of stimulus transformation, then the most common explanation within the empirical psychological literature is that the shape is being ‘transformed to match’. However, computationally speaking it is impossible to transform an image into alignment with another without prior knowledge of the image to be matched. Philosophically, it is nonsensical to assume that in order to recognize an object we must already ‘know’ what it is. In addition, neuropsychological evidence suggests that systems associated with the transformation of images are not directly involved in the recognition and long-term storage of object representations (e.g., Rösler, et al., 1995). To begin to understand how we recognize shapes from different viewpoints we need to ask questions of a variety of disciplines. By employing a multi-disciplinary approach we can use various streams of evidence to converge on a more complete understanding of the underlying processing.

In this chapter, we begin by discussing limitations of the traditional psychological paradigms in investigating recognition processes. There are two ways in which these conventions are limiting. Firstly, the dependent measures of accuracy and reaction time may not measure the effects of recognition strategies accurately. Secondly, the dependent measures may not in fact reflect the recognition processes we are hypothesizing, i.e., the measures may be accurate but we are wrong about what this processing being measured actually is. In the last part of this chapter, we will suggest further studies that approach the problem from a variety of disciplines.

## 10.2 Limitations of the experimental paradigm and dependent measures: Inaccuracies of measurement

> *Most of the things we observe in experiments have no relevance to the process of thinking, other than the empty observation that thinking, like most processes, takes a measurable amount of time.*
>
> *(Deese, 1969, p. 518)*

In the overwhelming majority of the psychological experiments reviewed, if the effect of a transformation (e.g., rotation) was deemed significant the authors assumed that a normalization strategy had been adopted. Alternatively, if the effect was insignificant it was assumed that subjects had used an invariant strategy. However, there are many factors besides the underlying visual processing that influence the data gained from such experiments. Different studies used different experimental paradigms and different statistical analyses. For example, the ability to detect a statistically significant effect depends on a number of variables, all of which vary between different studies – e.g., the number of subjects (increasing statistical power, reducing the effect of between subject variation or noise), the magnitude of the transformations used (proportional to effect size) and the number of trials (proportional to measure reliability i.e., minimizing the effect of within-subject variation). Ultimately, studies may differ in their ability to find a statistically significant effect of normalization even when the rate of transformation in both experiments is the same.

In addition to statistically significant effects, authors differ in their judgment as to what an acceptable rate of normalization (measured by effect size) might be in the first place. For example, while Cohen and Kubovy (1993) employed an upper limit of 1000 deg/s on valid rates of normalization for orientation in their experimental data, Tarr and Pinker (1989 [Experiment I]) and Pierret and Peronnet (1994 [Experiment I]) discounted significant effects of orientation as evidence for normalization, even though their rates were lower than 1000 deg/s. In this project, statistically significant effects of rotation consistent with normalization rates in the order of 1500 to 3000 deg/s were judged as valid evidence for normalization, partly because the attenuation of these effects paralleled the trends in other studies demonstrating slower normalization rates (e.g., Jolicoeur et al., 1987). Therefore, there is not only variability between studies as to whether a given rate of normalization will be statistically detectable but also whether such an effect is ultimately judged to be the result of normalization.

The question raised is an important one – at what point (i.e., at what effect size) do we draw the line between ‘normalization’ and ‘invariant’ recognition? Such a cut-off is even more difficult to define due to variation in normalization rates between studies (e.g., identification versus handedness discrimination) and within studies (e.g., the attenuation of the effects of transformation and between subject variation in strategies). The lack of any such criteria hampers the post-hoc comparability of studies and the objective interpretation of results. It also compromises the planning of future research with respect to determining statistical power and adequate N. For example, Jolicoeur and Landau (1984) have argued that some studies, which failed to find statistically significant effects of rotation, may have simply lacked the power to do so. They cite the studies of Simion et al. (1982) and Corballis et al. (1978) as examples where the effects of normalization are evident in mean reaction times but failed to reach statistical significance. However, setting criteria in a non-arbitrary manner seems almost impossible, given the variety of findings in the literature and different theoretical viewpoints.

It may be that normalization rates can be so fast that gaining adequate statistical power to detect these effects is impractical. The overall decrease in orientation effects in Experiments I and II (which parallel findings in other studies, e.g., Murray et al., 1993) may be entirely due to increasing speed of normalization within subjects and not to the gradual deployment of an orientation-invariant strategy. If this is the case, the rate of the transformation to match procedure may become so fast after sufficient practice, that the normalization profile will be virtually undetectable - the effect size associated with transforming the image will be small and there will still be constant interference from measurement and internal processing noise.

Why would the rate of normalization increase with practice? One possibility is that repeated presentations induce a more efficient initial coding of the stimulus pattern that is ‘informationally small’, so that the resulting representation is easier and faster to transform than one which is ‘informationally large’. This account could explain why complex images, such as faces (e.g., Rock, 1973, 1974; Thompson, 1980) and Tarr and Pinker’s (1989) images, invoke slow and resilient normalization procedures despite practice. Similarly, in handedness tasks, additional information is necessary in the preliminary encoding to determine left-right directionality. This, in turn, would account for the slower rates in handedness tasks (see section 5.3) and the need for more practice and motivation to reduce orientation effects than in identification tasks (e.g., Cohen & Kubovy, 1993).

Jolicoeur and Landau (1984) have argued that error rate may be a more sensitive indicator of normalization than reaction time. In their study, and that of Lawson and Jolicoeur (1998), there were persistent effects of orientation in accuracy of response, despite an ample number of presentations for practice effects. However, in order to induce an adequate number of errors (around 20% of responses in Jolicoeur and Landau (1984)) presentations were very brief and pattern masks were used. By halting or disrupting processing in this way, the experiments may be changing, in a qualitative manner, the nature of the perceptual processing. There is some evidence that such conditions may access a different representational code than tasks with more liberal stimulus presentation factors. The failure to find orientation effects in other conventional identification experiments may not be due to a lack of statistical power. Rather it may be the result of accessing a different, orientation-invariant representational code that was unavailable in Jolicoeur and Landau's (1984) and Lawson and Jolicoeur's (1998) experiments. While the notion of multiple representational codes, and the different sensitivities to shape invariant transformations between these codes, was not sustained in Experiment I, there is still post-hoc evidence (see sections 5.14 and 7.9), as well as promising experimental findings within studies (e.g., Ellis et al., 1989; Santee & Egeth, 1980), that support this hypothesis. Perhaps by modifying the experimental conditions to induce effects in alternative measures, we are achieving a kind of quantum effect – by changing the dependent measure we use, we may be altering the very phenomenon we wish to investigate.

## 10.3 Limitations of the experimental paradigm and dependent measures: When the measures do not reflect recognition processes

It may be that the dependent measures traditionally used *are* appropriate measures of the mental processing. Yet we may still be wrong about what these tools are actually measuring, i.e., our assumptions about the underlying processing may be incorrect. In this section, we will discuss alternative explanations for the effects of transformations on the dependent measures. This involves the influence of perceived similarity, alternative types of normalization, and the involvement of perceptual mechanisms that are not used for recognition.

It is normally assumed that the 'trend to invariance' in many experiments is due to a shift from the normalization strategy towards an invariant one. However, it may be that it is a change to a different type of normalization. As mentioned in section 3.3, once the specifications for a transformation to match have been calculated, there is no reason, computationally speaking, why any transformation cannot be completed in a single temporal step. With practice, subjects may be able to move from an analogue normalization strategy to some kind of 'snap' normalization whereby the transformation takes place in a single time frame. For example, if the image is misoriented, the 'snap' normalization transforms the image to its upright orientation without the need for representing the image at the intermediate orientations. Such a transformation may be similar to what is referred to as a *blink transformation*. In

Kosslyn's (Kosslyn, 1987; Kosslyn, Chabris, Marsolek, & Koenig, 1992) theory of mental imagery, this type of transformation is a crude, digital mapping procedure normally reserved for broader categorical distinctions (e.g., left / right, big / small). The other type of procedure in Kosslyn's theory is a *shift transformation* that is a precise analogue transformation, more akin to the normalization procedure often assumed to underlie the processing in identification and discrimination tasks. Perhaps, with practice, the blink transformation process becomes more precise until it is as accurate as the shift transformation.

Alternatively, there are many ways in which sensitivity to orientation and size in object recognition tasks may not be due to a normalization strategy. One possibility is that such results are due to the influence of perceived similarity. It may be that the magnitude of the shape preserving transformation between stored and current stimuli images is a source of perceived similarity and that this similarity information influences the speed or accuracy of processing. In other words, it is the similarity between images, which is a function of the magnitude of the transformation separating them, that is influencing responses and not the actions of a normalization strategy. Consider the case of recognizing or matching rotated stimuli. When the difference is $0^0$ the images are highly similar and, at angles greater than $0^0$, the deviation is proportional to the amount of perceived difference between the two representations (e.g., a $90^0$ separation produces less perceived similarity than a $30^0$ separation). The degree of perceptual similarity may influence the confidence of response. Even though information on objective shape is identical and this information is derived using an orientation-invariant recognition mechanism, subjects may be more inclined to recheck their decision and make a more detailed analysis of the image when there is information on difference.

This account may be especially pertinent in matching tasks such as Experiments I and II. In these experiments, the time for subjects to respond “same” to a pair of similar shaped but rotated stimuli is often proportional to the degree of “difference” in the pair. The ability to eliminate the influence of the transformational difference may reflect increasing confidence in response choice, based on orientation-invariant information. Such an account might explain the persistent advantage in recognizing upright images over disoriented images even after practice (e.g., Maki (1986) and Braine (1965)), because there is no information on difference along the orientation dimension for upright versions. The interference between the elements of objective shape and shape irrelevant transformation may be a kind of Stroop effect (Stroop, 1935). Naming the colour of the ink in which a colour word has been written is easy when the two are consistent (i.e., the word ‘blue’ written in blue ink). However, when the colour of the ink and the descriptor are different, subjects often give the name of the word and not the colour as they are requested to do. Few would dispute that colour and form are separable qualities and that such a distinction has some basis in the perceptual system despite this higher order interference. Similarly, experiments by Glezer and colleagues (see Glezer et al., 1974) have also demonstrated that form is processed in a separate channel to that of size and orientation information. Perhaps a similar interference phenomenon is at work in recognition tasks. Even though identification is irrelevant to size and orientation, such features, at least initially, influence response bias or confidence and cannot easily be ignored.

An alternative explanation for the sensitivity to properties such as size and orientation is the involvement of low-level perceptual mechanisms, i.e., mechanisms that do not directly involve the recognition of the object but occur at earlier stages of visual processing such as stimulus encoding. We have already discussed the influence of the zoom lens theory of attentional deployment and the role of this perceptual scaling is well established empirically (see section 7.14). To what extent are other size effects, and effects of rotation for that matter, attributable to similar encoding mechanisms? Cooper, Biederman and Hummel (1992) have proposed that the effects of size and orientation in some tasks may "represent influences of a visual stage that occurs prior to object recognition" (p.194). Similarly, Carpenter and Just (1978) have demonstrated orientation sensitivity in encoding stages. They found that patterns of eye-movement may reflect orientational differences in a mental rotation task. Tarr and Pinker (1989 [Experiment I]) used this explanation to account for residual rotation effects (which were equivalent to those found in identification tasks) in a highly practised mental rotation task. However, it would be unwise to presume that eye-movement patterns are the same in identification and mental rotation tasks given the qualitative and quantitative differences between performance in the two types of experiments outlined previously (see section 2.6 and chapter 5).

If the locus of temporal sensitivity to transformation effects in the perceptual system is at the encoding level, this would explain some of the results in Jolicoeur and Landau's (1984) and Lawson and Jolicoeur's (1998) experiments. Brief presentation times and masking may cut short the encoding process that is sensitive to the orientation and scale characteristics of the image. Because the encoding process is more likely to be disrupted (and an incomplete stimulus representation formed) with transformations of greater magnitude, the error will reflect the size of the transformation even though no normalization to match process is occurring. Certainly, many authors including Cavanagh (1989) propose that early perceptual analyses are highly sensitive to low-level spatial characteristics and not more abstract properties, such as shape, which may be derived at higher levels of perceptual processing. However, if such mechanisms are responsible for effects traditionally subsumed by normalization, it is not clear how these effects occur for some but not all individuals or how the effects are reduced with practice if the effect occurs at a low level of perceptual processing *prior* to object knowledge.

In addition to stimulus encoding mechanisms, there may also be an influence of higher-level perceptual mechanisms within an object recognition task that do not involve object recognition but which may produce normalization-like effects in the response data. Regardless of the nature of the task, the perceptual system may engage in a variety of analyses and procedures in addition to those of pure shape recognition. A response would not be given until this battery of mechanisms is complete. Therefore, even if the recognition procedure may have already been completed, a behavioural response may not be given until the other procedures, some or all of which may be sensitive to orientation and size, have finished. Leek (1998) raised this issue in considering the effects of stimulus orientation on response time and suggested that these effects may result from mental processes such as “the preparation and planning of motor actions or the calculation of stimulus orientation for the purposes of reaching and grasping even when such actions are not required by the task” (p.650).

Object specific familiarity effects, in which the influence of orientation is attenuated, may therefore be the results of the suppression of the activities of parallel processes that are irrelevant to the task at hand, namely shape recognition. However, the notion that such extraneous procedures involve the planning of motor activities requires experimental validation. It could also be argued that Leek's (1998) findings may not entirely support this hypothesis. The objects for which orientation did not affect response time, i.e., poly-oriented objects, were those that are commonly manipulated by hand (e.g., pencil, pen, razor, cigarette, scissors, banana, knife, comb, toothbrush, carrot, hammer) whereas those that did invoke orientation effects were those less likely or easily manipulated by hand (e.g., computer, truck, washing machine, skyscraper, dresser, crane, church, fridge, car, tree). It seems counterintuitive for the automatic planning of motor activity to be suppressed only for those objects for which such planning would be beneficial in real-life situations.

By proposing separate mechanisms for recognizing and transforming the stimulus image, we can also explain the family resemblances between performance on a variety of visual processing tasks. In chapter 2 we discussed a number of phenomena involving implied transformations between static stimuli – the *motion-recognition continuum.* These mechanisms involve not only an implicit transformation process but often an automatic, phenomenal experience of the transformations themselves. However, such transformations cannot be calculated or performed until shape constancy has been obtained. Therefore, there may be two mechanisms involved – one for recognition and one for transformation. This dual-mechanism account is also appropriate for processing in the experiments we have performed. The experimental paradigm we used is remarkably similar to those under which effects such as apparent motion and representational momentum are evident– namely, successive, identically shaped images differentiated by a transformation from the similarity group. The only differences are in the timing aspects of the task (e.g., stimulus presentation duration, ISI and SOA) and how the subject is required to respond. In all cases, the perceptual system may perform similar analyses on the stimulus image. The most parsimonious solution, i.e., the one that can account for processing in a large number of different tasks, is a two-stage linear model: the image is first recognized in an invariant manner and then the implied motion is calculated and performed.

Such a model can also explain how and why the transformation procedure can be eliminated in recognition tasks on a stimulus specific basis. Once the first stage of processing is complete there is no necessity for the second stage in such tasks – object knowledge has already been accessed and a response can be given without the need for additional processing to interpret implied motion. Because object recognition is primary, there may be a learned association between the representation of a shape and subsequent image transformation such that recognizing a shape can allow or inhibit later processing. In other words, in a recognition task where understanding implied motion is not useful to the task and may even provide an unnecessary burden to the visual processing system, the object representations of the shapes presented repeatedly in the task may inhibit subsequent transformation of that image. This two-stage model avoids the problem with stimulus specific attenuation effects of transformation faced by many of the alternative accounts outlined above. It also explains how we appear to 'know' what the shape is before we transform it – we have in fact already completed recognition before normalization. Finally, it provides a parsimonious account of the results, not only from recognition tasks but from the variety of perceptual phenomena subsumed by the motion-recognition continuum.

This theory could easily be extended to account for some of the influence of timing in the experimental literature. The reduction in effects of shape preserving transformations with increasing ISIs (e.g., Ellis et al. (1989) and Santee and Egeth (1980)) may be the result of the reduced influence of a non-recognition, image transformation procedure. Larger ISIs are less likely to invoke the phenomenal experience of motion For example, in Santee and Egeth's (1980) experiment, strong influences of size transformation were found when the same shape was presented successively at different sizes with no ISI. This is remarkably similar to the experimental paradigm successfully employed by Bundesen et al. (1983) to invoke the apparent motion effects of depth motion – namely, two images differing by only a size transformation presented successively without any inter-stimulus gap. The only fundamental difference between tasks is how the subjects are required to respond – what they actually see is remarkably similar. At larger ISIs (which is in fact proportional to SOA in their paradigm because presentation duration of the first stimulus remains constant) the effect of size diminishes in Santee and Egeth's (1980) experiment. Similarly, apparent motion for a given size of transformation is no longer evident for SOAs that exceed a certain upper limit (Korte's Third Law). However, there is some evidence in our first experiment and in other research that timing does not influence orientation and size sensitivity. For example, Murray (1999) found no corresponding effect of ISI for planar rotations. However, there may be a myriad of variables that influence the role of the motion phenomena besides timing in the experiment. Further research is needed into this area.

The hypothetical two-stage model we are suggesting requires further experimental validation. However, some recent research directly supports this account. A study by De Caro and Reeves (2000) concluded that orientation invariant recognition was primary and that the influence of orientation on response times was due to a post-recognition stage of processing. The authors performed a dual task experiment using line drawings of common objects. After presentation of the image there was a stimulus onset asynchrony (SOA) of either 14, 28 or 41 ms followed by a pattern mask and then one of two tasks. The first task required picture-word verification of the drawings. The second task involved orientation verification in which participants were required to judge whether the arrow currently shown on the display was pointing in the same direction as the object they had just seen. The critical SOA ($SOA_c$), the estimated SOA at which subjects were 75% correct, was estimated across orientations at $45^0$ intervals.

For the orientation task, the $SOA_c$s increased in an approximately linear fashion with angular deviation from the upright. In contrast, the $SOA_c$s for name verification did not demonstrate a similar, significant linear trend across angular deviations and were consistently shorter across all angles than in the orientation task. These results suggest that the identification was orientation invariant but that the determination of orientation requires some kind of mental rotation or normalization procedure. It may be that the post-recognition process of determining orientation is an automatic process and that it is related to the automatic transformational phenomena of the motion-recognition continuum. What the experiment does not confirm is whether the determination of orientation procedure evident in their task is the precise cause of the ‘normalization’ effects in other experiments. It may be that backward masking in their experimental paradigm may be accessing different representational codes to those in other experiments. Nevertheless, De Caro and Reeves’ (2000) study still provides convincing evidence that image transformation may be the result of a post-recognition stage.

There are many incentives for attributing ‘normalization’ effects to these motion phenomena and not to a recognition process. As mentioned above, the influence of these motion phenomena can account for the varying influence of shape preserving transformations in the literature on recognition. In addition, it simultaneously provides a solution to the correspondence problem for apparent motion and all other phenomena in the *motion-recognition continuum*. This explanation has the potential to provide a unified theory of perceptual processing in which invariant recognition precedes the mental transformation and we can explain how we seem to know which image to transform in mechanisms ranging from apparent motion through to ‘normalization’ in recognition tasks.

The notion that the image transformation effects that occur in recognition tasks might be related to the analysis of movement also ties in nicely with the dorsal / ventral debate concerning visual processing (see sections 2.5.6 and 7.6.2). Object recognition would occur in the ventral stream that neuropsychological studies have shown is strongly associated with object recognition and the long-term store of object representations. This would involve an invariant recognition strategy. In contrast, the image transformation procedure would take place in the dorsal stream that is associated with the mental transformation of images and visuo-motor processes. This system is involved, not in recognizing an object, but in understanding motion associated with it and coordinating the manipulation of the object in the real world. Therefore, this hypothesis regarding two processes of visual processing in naming and matching tasks has some correspondence with what we know about visual processing in the brain. At the very least, it provides us with research hypotheses to explore in tasks involving object recognition where brain activity can also be monitored. In the next section, we will discuss how this, and many other experiments, may be conducted to further understand the processes underlying visual object recognition.

In summary, we have tried to look objectively at the traditional psychological approaches to the problem of visual object recognition and to examine approaches from different fields of investigation. The dual system account we have ultimately favoured, whereby one system performs invariant recognition while the other analyses implied motion, requires validation with further studies. However, it does provide a concise way of uniting a number of varied perceptual phenomena under the one umbrella. This theory also takes into account findings from different fields of research. Only by taking such a multidisciplinary perspective can we hope to overcome the limitations associated with each and every research paradigm, whether in psychology, neuropsychology, neurophysiology or philosophy. As researchers, we are often naive enough to assume that our experiment is examining the single perceptual process we are interested in. By drawing comparisons between phenomena which occur in remarkably similar paradigms but which are investigated within ‘different’ fields of vision research, we hope to have shed some light on how these different ideas may be related.

## 10.4 Future research

In discussing the results from this project and others from the experimental literature, we appear to have asked as many questions as we may have answered. In order to gain a more complete understanding of the way we recognize shapes under different transformations, further investigation is sorely needed. In this section, we will offer some suggestions for further research with an emphasis on variety not only in the tasks employed but also in the disciplines in which the questions are asked. The limitations of the conventional reaction time and accuracy paradigms in psychology have been noted above. However, all disciplines operate within well-established scientific paradigms and possess their own particular limitations and biases. Only by following a multi-faceted approach can we hope to converge on a more accurate picture of the underlying psychology.

One general finding of the current research and other recent studies (e.g., McKone & Grenfell, 1999) that should influence future research is that the mechanism(s) for recognition, and other perceptual processes, invoked in tasks involving objects varying in size and orientation, may be influenced by individual differences. As a result, all of the suggested studies listed below require larger Ns then those conventionally employed in the experimental literature. Preferably, studies should be designed to allow for such investigation. Even if the experimenter chooses not to investigate such differences, one needs a large enough sample to minimize their effects. There is reason to believe that the implied assumption of consistency between subjects on recognition tasks should be questioned. In addition, increasing Ns will help to allay suspicions of inadequate statistical power. Given the difficulties mentioned above in determining a priori statistical power, the simplest way to reduce the risk of incorrectly rejecting the scientific hypothesis of normalization is by testing a comparatively larger sample population.

**1) A psychological refractory period (PRP) task investigating the effects of planar rotation and scale changes in recognition**. This experiment could be based on the dual task paradigm employed by Lawson et al. (2000) to demonstrate that planar and in depth rotations are processed in a sequential manner. Despite some similarities in the ways size and planar rotations are compensated for, this may be due to the activities of entirely different systems under a higher level of control and / or the same experimental factors influencing the two systems in a similar manner. A PRP task would examine whether the compensation for, or normalization of, the two symmetries occurred at the same processing stage and, if not, in what order they occurred.

**2) A study investigating neurophysiological changes in mental transformation tasks and identification or matching tasks involving planar rotated and scaled stimuli**. Rösler et al.'s (1995) experiment demonstrated similarities between the processes of intentional, conscious mental rotation and mental size scaling and showed that the transformational process for both is associated with the parietal cortex. Comparing measures of cerebral activation via ERPs or fMRI, in both similar mental transformation tasks and recognition tasks free from MRFs, would shed some light on the similarities and dissimilarities between the two processes. Is the parietal cortex also involved in normalization of objective shape and, if so, is it involved to the same extent for equivalent transformations in the two tasks?

There is some evidence that size normalization at least is associated with the dorsal stream in a task free from MRFs (e.g., Larsen, Bundesen, Kyllingsbæk, Paulson & Law, 2000). Additionally, by employing a sufficient number of trials and a limited number of different objects, one could seek the attenuation of orientation and size effects in psychological measures. Is such an attenuation effect associated with quantitative (e.g., change in slow-wave amplitude in parietal cortex) or quantitative (e.g., change in cortical areas recruited for the task) alterations in cerebral activation?

**3) Identification (single presentation) and matching (successive presentation) tasks investigating the effects of different timing and presentation variables on the recognition of planar rotated, depth rotated and scaled stimuli.** The hypothesis of multiple representational codes, whose sensitivity to shape preserving transformations varies, may be further examined by experiments exploring the effects of presentation duration, ISI, SOA and the presence or absence of backward pattern masking. Are the effects of ISI on size sensitivity in Santee and Egeth's (1980) experiment also generalizable to rotations? Do the effects of presentation time on depth rotation and size sensitivities in Ellis et al.'s (1989) paper generalize to planar rotations? Are the resilient orientation effects in error rates in Jolicoeur and Landau's (1984) and Lawson and Jolicoeur's (1998) experiments due only to the brevity of the presentation time or do they also depend on the presence of backward pattern masking? Will conditions similar to those in the above pair of studies induce comparable effects for size and in depth rotations?

Such experiments could shed more light on the issue of where, in terms of processing stages, the effects of different transformations are compensated for. In

Lawson et al.'s (2000) PRP study, planar rotations were compensated for at a pre-bottleneck stage, whereas depth rotations were dealt with in either a bottleneck or post-bottleneck stage. In addition, Ellis et al.'s (1989) study provided evidence for an early representational code or process that is sensitive to depth rotations but not necessarily scale changes (the VIEW CODE). Therefore, it is possible that size compensation may occur before depth rotations. A PRP study could test the hypothesis that planar rotation and scale aspects are both compensated for before the central bottleneck, while conventional psychological recognition tasks, investigating temporally sensitive codes, could test for codes that are sensitive to planar rotations but not scale changes.

**4) Tasks involving the orthogonal manipulation of planar rotation and size.** Experiments similar to those employed in this study could be administered with a broader range of rotation and scale transformations. As mentioned previously, one difficulty with the orthogonal manipulation of two variables is that, even if the two factors have a similar qualitative effect within the perceptual system, the quantitative effects of the two in the experimental context may not be of a comparable magnitude. If they are not, then we may find significant effects in only one factor despite such underlying similarities because of differences in effect size. Even if subjects are normalizing for both properties, is a $90^0$ rotation equivalent to magnification by a factor of two?

Perhaps the best way to approach this problem is with three experimental conditions – the first two examining the simple effects of the two variables separately and the final condition examining the orthogonal manipulation of the two transformations. This would also allow us to examine whether the simple effects of size and rotation are comparable to those in orthogonal conditions – is the normalization rate for rotated images the same as it is for rotated and scaled images? If the two procedures draw on common resources then this may be reflected in comparatively slower normalization rates in the mixed condition than under the simple effect conditions. Bundesen et al. (1981) have suggested that, despite additivity between the effects of the two transformations, the global transformational path is separated into individual steps of simple rotations and scale changes. Is this alternation between types of transformation additionally taxing to the perceptual system? Our own experiments were inconclusive on this matter. Although the rate of normalization for orientation was faster when orientation was crossed with size than when it was the only transformation, the effects of orientation also took longer to dissipate under the orthogonal condition. However, we used different subjects and orientations in the two experiments, as well as a larger N in the second task, which makes such comparisons somewhat tenuous.

One limitation of the experiments in this and other studies is the range of transformations used. Using a wider variety of transformations would provide a clearer picture of processing. Despite significant differences between a certain set of transformations suggestive of a normalization procedure, a wider range of transformations may demonstrate significant deviations from a normalization strategy (e.g., Lawson & Jolicoeur, 1999). This in turn may provide evidence for concurrent processing sensitive and insensitive to the transformations employed.

**5) Perceptual similarity of forms under different transformations.** Does the theory of objective shape have any psychological reality? Or are mirror-reflected versions of the same form classified as a different 'shape'? One way to begin addressing this question is by means of a similarity task, in which subjects are asked to rate similarities between different images on a scale where the minimum rating (e.g., -5) represents very little similarity and the maximum rating (e.g., 5) represents highly similar forms. These images would vary in terms of transformations of the same root form or forms that vary in terms of the transformations applied. Such transformations could involve those from the similarity group as well as other symmetries, e.g., affine transformations. Are mirror-reflected images of root images judged to be as similar as rotated versions? Are the effects of objective shape preserving symmetries less detrimental to perceived similarity than the affine transformations? Does objective shape play a special role in the judgment of similarity? Are scaled images more perceptually similar than rotated images?

To further investigate the notion of shape one might also include a condition where the stimuli are identical to those in the similarity condition but the task is not one of judged similarity but rather of confidence in shape constancy. Subjects would be asked to rate the relationship between the images where the minimum value (e.g., -5) reflects very low confidence in the two images possessing the same shape, and the maximum value (e.g., 5) represents a very high degree of confidence in shape constancy. Such a study could address the question of whether the notion of pure, objective shape has any psychological reality– is this concept reflected indirectly in a judged similarity task and / or more directly in explicit judgments of the everyday understanding of the term 'shape'?

**6) Neuropsychological studies investigating breakdowns in functioning associated with the ability to consciously transform and recognize images varying along the dimensions of size and planar rotation.** As mentioned previously, Farah and Hammond (1988) have conducted experiments on subjects with brain lesions where, although the ability to mentally rotate was impaired, recognition of objects under the same transformations appeared intact. Do such patients exhibit a similar difficulty in mentally transforming an image along the dimension of scale? Are abilities or disabilities in the recognition of images that are transformed in various ways associated with similar patterns of brain lesions? Are our abilities to estimate an object's orientation and size functionally related? Layman and Greene's (1988) study demonstrated a selective breakdown in the ability to recognize images that have been depth rotated versus those that have been planar rotated. Do such patients demonstrate similar problems with the same types of conscious mental transformation? And do those with difficulties in such recognition tasks also possess difficulties recognizing images under different transformations, e.g., mirror reflections and scale changes?

Neuropsychological studies of selective impairments after brain trauma may provide insight into whether or not normalization procedures for both conscious transformations and implied transformations in recognition tasks are located in similar cortical areas. For example, Rösler et al.'s (1995) study provides evidence that transformations for both size and rotation can occur within the parietal cortex. Do lesions in these areas also manifest themselves in corresponding deficiencies in recognition under these transformations? Such studies would shed further light on the important issue of how conscious mental transformations of rotation and size might be related to normalization mechanisms under the two symmetries for the purposes of recognition.

**7) Computational modeling of candidate recognition procedures under scale and rotation changes.** Explicit modeling not only provides mechanisms that can account for existing psychological data but it can also be used in a more experimental manner to provide hypotheses to test psychologically. The latter can be even more thought provoking when, like Schwartz's (1980a) account, they are derived from neurophysiological data. We believe that both the idea of computational anatomy (i.e., that the topological mapping of stimuli images can provide an important computational, analytical role) and the concept of multiple representational levels in the perceptual system are worthy of further investigation. In addition, such approaches capture important characteristics of brain structure and anatomy. Such modeling is important because it allows us to entertain novel hypotheses for perceptual functioning and also exclude hypotheses that may be unrealistic from a computational point of view. We can then design experiments to test potential hypotheses in the psychological domain. In addition, Churchland (1986) has pointed out that Connectionist modeling can also guide further research into neurophysiology by providing testable hypotheses of neural mechanisms. Therefore, Connectionism provides an ideal framework for the modeling of such processes, given its relationship with brain hardware, its ability to implement theories at a computational level and its natural ability to mimic the parallelism and complexity of brain functioning.

One particular advantage of the Connectionist paradigm is that it allows the modeling of non-linear, dynamic systems – a feature that we believe is fundamental to visual processing. As Marr (1982) has pointed out, there is likely to be a good deal of feedback between processing stages with the result that the final recognition response may be viewed as the product of a relaxation process between these stages. Low-level information from the stimulus image triggers responses at a higher-level which can in turn guide further informational extraction at the earlier stage. It is unlikely that any system can recognize objects in complex, real world visual scenes in a single feedforward pass, as entertained in our Connectionist model. However, such models may capture broader principles of perceptual mechanisms. At the present time, it is arguable whether enough is known about human performance on such tasks to warrant more precise modeling.

**8) Studies investigating relationships between performance on various perceptual tasks including recognition tasks involving rotated and scaled shapes.** These would involve large Ns and investigate correlations between performance on handedness, identification, matching tasks with and without same shape distractors, as well as tests of attentional scale (e.g., Pasto & Burack, 1997), and visual intelligence (e.g., Ravens Advanced Progressive Matrices). Correlational data and comparisons between different rates of 'transformation' may yield commonalities in processing speed, relationships between strategy choice and measures of intellectual functioning, or both. Are similar representational structures utilized for the different tasks? Is there a meaningful, general measure of visual processing speed? Such investigation appears warranted in light of the preliminary evidence for individual differences in task performance and strategy.

**9) Attentional versus shape specific size effects.** When are effects associated with differences in the scale of images attributable to attention and not recognition? What factors influence the scale sensitivity of early visual encoding? Is there a reliable way of discriminating between the two size–specific mechanisms experimentally? Such questions may be addressed by experiments using similar stimuli and scales in experiments involving recognition and attention. Is the time necessary to expand or contract the zoom lens comparable to that for shape specific size normalization?

One route of investigation could be to investigate serial performance on a matching task and a target identification in multiple arrays task. This matching task would be similar to the paradigm we have employed in our experiments with the exception that only size transformations are used. The target identification task would consist of an array of letters (evenly spaced along the horizontal and vertical dimensions) all presented in an upright position and in the same scale presented on an evenly spaced grid. In each array, all letters would be identical except for one – the target letter. Subjects would be required to provide a match or foil response to a single trial of the matching task. As soon as this response is given, a single trial of the target identification task would be presented and subjects would be required to name the target stimulus. If the zoom lens is adjusted for the size of the second pair of the match trial then performance on the subsequent target identification task should be facilitated if the target is within the area subsumed by the second image of the matching trial. If the target is outside this area then the lens would need to be adjusted and this would have a negative effect on response capabilities. This experiment would investigate the connection between space and shape based attributes of the zoom lens.

# APPENDIX A

# VOLUNTEERS WANTED!

**Dear student,**

**I am a PhD student in Psychology at the University of Adelaide and I would like to invite you to participate in my study. The experiments are designed to investigate the effects of rotation and size changes in the recognition of images and, in so doing, contribute to our understanding of the mental representation of objects. This type of information can have important real-world implications – for example, in improving identification in situations where fast response is vital (e.g., in automobile safety) and in helping understand how identification behaviour can break down (e.g., in diagnosing visual deficiencies following brain lesions). The simple tasks involve abstract images presented one after the other on a computer screen and require you to respond only "yes" or "no" – "yes" if the image is of the same object as that in the previous image and "no" if the depicted object is a different one. The tasks are timed but none of the tasks involve any intelligence tests and all results are strictly confidential. The tests should take no longer than about 30 minutes. If you would like to participate, please contact me by phone, e-mail or pigeonhole (details below) letting me know when you would be available and how I can get in contact with you.**

**Thanks,**

**Marcus Butavicius**
**E-mail: marcus.butavicius@psychology.adelaide.edu.au**
**Phone: 8303 4674**
**Pigeonhole (just inside the psychology office, to the left of the counter)**

# APPENDIX B

## *INFORMATION SHEET*

Dear Subject,

My name is Marcus Butavicius and I am a PhD student in Psychology. I would like to invite you to participate in a study on visual perception. This research is investigating how changes in orientation can alter the way we perceive and identify images.

Currently, there is considerable debate as to whether, and how, such factors influence our recognition of objects in the environment around us. This type of research is important in helping to understand how the objects we view are represented in the human perceptual system. Through understanding the very nature of the internal representation of objects, we can not only improve identification behaviour (e.g., in conditions where fast response times are important, like road safety) but also learn how it can break down (e.g., in diagnosing deficiencies in perceptual processing following brain lesions).

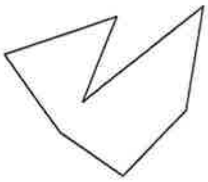
On the following page are some examples of the type of image that will be included in the experiment and the ways in which an image can be presented (i.e., rotated). If you would like to take part in this experiment, please take the time to look over these examples before you begin.

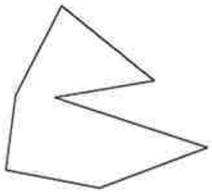
Thanks,

Marcus Butavicius

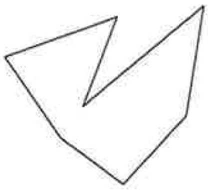
# EXAMPLES OF ROTATIONS

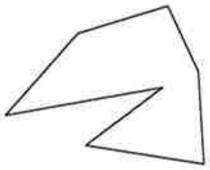
**1. Rotating** an image is just like turning an object (like turning a door knob).





ORIGINAL ➜ ROTATED





ORIGINAL ➜ ROTATED

# APPENDIX C

*Table Appendix C.*

***Experiment I – Means and SEs (bracketed) for Test vs. Train comparison 1***

Match trials – Reaction time

| | Train 1 | | Test 1 | |
|---|---|---|---|---|
| | *Rotation level 1* | *Rotation level 2* | *Rotation level 1* | *Rotation level 2* |
| *Condition 1 (100ms)* | 370.44 (43.56) | 414.75 (44.26) | 360.93 (40.68) | 349.64 (41.39) |
| *Condition 2 (500ms)* | 353.72 (43.56) | 378.18 (44.26) | 324.57 (40.68) | 350.75 (41.39) |

Match trials - Accuracy

| | Train 1 | | Test 1 | |
|---|---|---|---|---|
| | *Rotation level 1* | *Rotation level 2* | *Rotation level 1* | *Rotation level 2* |
| *Condition 1 (100ms)* | 95.37 (2.02) | 91.67 (2.22) | 98.15 (1.67) | 97.22 (1.73) |
| *Condition 2 (500ms)* | 96.30 (2.02) | 96.30 (2.22) | 95.37 (1.67) | 96.30 (1.73) |

Foil trials – Reaction time

| | Train 1 | | Test 1 | |
|---|---|---|---|---|
| | *Rotation level 1* | *Rotation level 2* | *Rotation level 1* | *Rotation level 2* |
| *Condition 1 (100ms)* | 377.22 (36.05) | 400.62 (39.44) | 378.55 (31.03) | 326.45 (27.98) |
| *Condition 2 (500ms)* | 310.36 (36.05) | 395.11 (39.44) | 362.91 (31.03) | 330.57 (27.98) |

Foil trials – Accuracy

| | Train 1 | | Test 1 | |
|---|---|---|---|---|
| | *Rotation level 1* | *Rotation level 2* | *Rotation level 1* | *Rotation level 2* |
| *Condition 1 (100ms)* | 97.04 (1.05) | 95.77 (1.43) | 96.30 (1.73) | 100.00 (.66) |
| *Condition 2 (500ms)* | 100.00 (1.05) | 99.21 (1.43) | 97.22 (1.73) | 99.10 (.66) |

# APPENDIX D

*Table Appendix D.*

***Experiment I – Means and SEs (bracketed) for Test vs. Train comparison 2***

Match trials – Reaction time

| | Train 2 | | Test 2 | |
|---|---|---|---|---|
| | *Rotation level 1* | *Rotation level 2* | *Rotation level 1* | *Rotation level 2* |
| *Condition 1 (100ms)* | 308.42 (36.87) | 341.36 (36.40) | 359.02 (42.10) | 316.51 (31.21) |
| *Condition 2 (500ms)* | 304.51 (36.87) | 307.73 (36.40) | 309.39 (42.10) | 312.45 (31.21) |

Match trials - Accuracy

| | Train 2 | | Test 2 | |
|---|---|---|---|---|
| | *Rotation level 1* | *Rotation level 2* | *Rotation level 1* | *Rotation level 2* |
| *Condition 1 (100ms)* | 95.37 (1.67) | 99.07 (.93) | 97.22 (1.18) | 97.78 (2.15) |
| *Condition 2 (500ms)* | 98.15 (1.67) | 99.07 (.93) | 99.07 (1.18) | 95.19 (2.15) |

Foil trials – Reaction time

| | Train 2 | | Test 2 | |
|---|---|---|---|---|
| | *Rotation level 1* | *Rotation level 2* | *Rotation level 1* | *Rotation level 2* |
| *Condition 1 (100ms)* | 362.29 (32.41) | 340.19 (29.15) | 365.35 (31.01) | 371.43 (31.90) |
| *Condition 2 (500ms)* | 332.98 (32.41) | 320.78 (29.15) | 324.71 (31.01) | 329.64 (31.90) |

Foil trials – Accuracy

| | Train 2 | | Test 2 | |
|---|---|---|---|---|
| | *Rotation level 1* | *Rotation level 2* | *Rotation level 1* | *Rotation level 2* |
| *Condition 1 (100ms)* | 95.37 (1.67) | 99.07 (.93) | 97.22 (1.18) | 97.78 (2.15) |
| *Condition 2 (500ms)* | 98.15 (1.67) | 99.07 (.93) | 99.07 (1.18) | 95.19 (2.15) |

# APPENDIX E

## *INFORMATION SHEET*

Dear Subject,

My name is Marcus Butavicius and I am a PhD student in Psychology. I would like to invite you to participate in a study on visual perception. This research is investigating how changes in orientation and size can alter the way we perceive and identify images.

Currently, there is considerable debate as to whether, and how, such factors influence our recognition of objects in the environment around us. This type of research is important in helping to understand how the objects we view are represented in the human perceptual system. Through understanding the very nature of the internal representation of objects, we can not only improve identification behaviour (e.g., in conditions where fast response times are important, like road safety) but also learn how it can break down (e.g., in diagnosing deficiencies in perceptual processing following brain lesions).

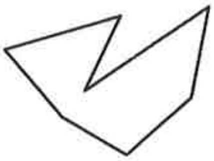
On the following page are some examples of the type of image that will be included in the experiment and the ways in which an image can be presented (i.e., rotated and scaled). If you would like to take part in this experiment, please take the time to look over these examples before you begin.

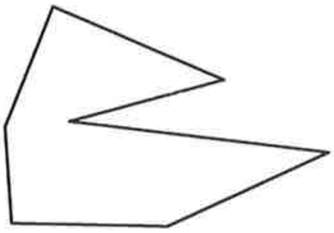
Thanks,

Marcus Butavicius

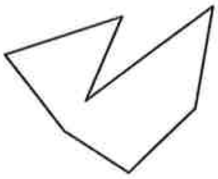
# EXAMPLES OF ROTATED AND SCALED IMAGES

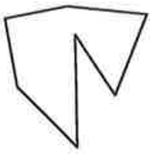
**Rotating** an image is just like turning an object (like turning a door knob).
**Scaling** an image simply means changing the size of the image (i.e., making it larger or smaller).
In this experiment the images you see can be both rotated and scaled



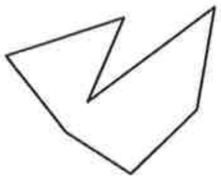


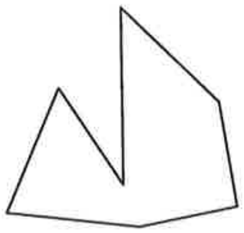
ORIGINAL → ROTATED and SCALED



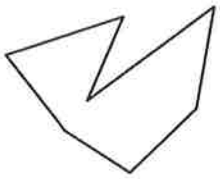


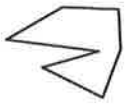
ORIGINAL → ROTATED and SCALED





ORIGINAL ➜ ROTATED and SCALED





ORIGINAL ➜ ROTATED and SCALED

# APPENDIX F

*Table Appendix F*

***Experiment II – Means and SEs for accuracy (expressed as a percentage)***

**Match trials – accuracy**

| *Block #* | *Size level* | *Rotation level* | *Mean* | *SE* |
|---|---|---|---|---|
| 1 | 1 | 1 | 95.56 | 1.25 |
| | | 2 | 95.28 | 1.11 |
| | 2 | 1 | 95.56 | 1.11 |
| | | 2 | 95.56 | 1.18 |
| 2 | 1 | 1 | 94.44 | 1.28 |
| | | 2 | 95.56 | 1.18 |
| | 2 | 1 | 95.00 | 1.53 |
| | | 2 | 94.44 | 1.15 |
| 3 | 1 | 1 | 96.67 | 1.03 |
| | | 2 | 97.78 | 0.79 |
| | 2 | 1 | 95.56 | 1.43 |
| | | 2 | 95.83 | 1.04 |
| 4 | 1 | 1 | 94.73 | 1.02 |
| | | 2 | 97.78 | 0.88 |
| | 2 | 1 | 92.78 | 1.68 |
| | | 2 | 96.11 | 0.87 |

**Foil trials – accuracy**

| *Block #* | *Size level* | *Rotation level* | *Mean* | *SE* |
|---|---|---|---|---|
| 1 | 1 | 1 | 97.23 | 1.01 |
| | | 2 | 96.39 | 1.11 |
| | 2 | 1 | 95.56 | 1.18 |
| | | 2 | 96.67 | 0.94 |
| 2 | 1 | 1 | 96.94 | 0.93 |
| | | 2 | 96.94 | 1.09 |
| | 2 | 1 | 95.83 | 1.04 |
| | | 2 | 96.67 | 0.86 |
| 3 | 1 | 1 | 98.89 | 0.53 |
| | | 2 | 96.11 | 1.11 |
| | 2 | 1 | 96.94 | 1.29 |
| | | 2 | 96.39 | 1.18 |
| 4 | 1 | 1 | 98.89 | 0.53 |
| | | 2 | 97.23 | 0.93 |
| | 2 | 1 | 97.50 | 0.82 |
| | | 2 | 96.67 | 0.94 |

# REFERENCES

Ahlfors, L. (1966). *Complex Analysis.* New York: McGraw Hill.

Anderson, J.A. (1968). A memory storage model utilizing spatial correlation functions. *Kybernetic, 5(3),* 113-119.

Anderson, J.A. (1970). Two models for memory organization using interactive traces. *Mathematical Biosciences, 8,* 137-160.

Anderson, J.A. (1972). A simple neural network generating an interactive memory. *Mathematical Biosciences, 14,* 197-200.

Anderson, J.A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review, 80(6),* 417-438.

Anderson, J.A. (1995). *An Introduction to Neural Networks.* Cambridge, Massachusetts: MIT Press.

Anstis, S., Giaschi, D., & Cogan, A.I. (1985). Adaptation to apparent motion. *Vision Research, 25 (8),* 1051-1062.

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

Ashbridge, E. Perrett, D.I. (1998). Generalizing across object orientation and size. In V. Walsh, & J. Kulikowski (Eds.) *Perceptual constancy: Why things look the way they do* (pp. 192-209). Cambridge, UK: Cambridge University Press.

Attneave, F., & Arnoult, M.D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin, 53,* 452-471.

Barriopedro, M.I., & Botella, J. (1998). New evidence for the zoom lens model using the RSVP technique. *Perception and Psychophysics, 60(8),* 1406-1414.

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

Besner, D., & Coltheart, M. (1975). Same-different judgments with words and nonwords: The differential effects of relative size. *Memory & Cognition, 3(6),* 673-677.

Besner, D. (1978). Pattern recognition: Are size and orientation additive factors? *Perception and Psychophysics, 23(1)*, 93.

Besner, D. (1983). Visual pattern recognition: Size preprocessing re-examined. *Quarterly Journal of Experimental Psychology, 35A,* 209-216.

Besner, D., & Coltheart, M. (1976). Mental size scaling examined. *Memory & Cognition, 4(5)*, 525-531.

Bethell-Fox, C.E., & Shepard, R.N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14(1)*, 12-23.

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

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

Biederman, I., & Cooper, E.E. (1991b). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23(3),* 393-419.

Biederman, I., & Cooper, E.E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance, 18(1),* 121-133.

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

Binford, T. O. (1971). *Visual perception by computer.* Paper presented at the IEEE Conference on Systems and Control, December 1971, Miami.

Braccini, C., Gambardella, G., & Tagliasco, V. (1982). A Model of the Early Stages of the Human Visual System: Functional and Topological Transformations Performed in the Peripheral Visual Field. *Biological Cybernetics, 44,* 47-58.

Braddick, O.J. (1974). A short-range process in apparent motion. *Vision Research, 14,* 519-527.

Braddick, O.J. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London B, 290*, 137-151.

Braine, L.G. (1965). Disorientation of forms: An examination of Rock's theory. *Psychonomic Science, 3*, 541-542.

Bregman, A.S. (1977). Perception and behaviour as compositions of ideals. *Cognitive Psychology, 9,* 250-292.

Brennen, T., Baguley, T., Bright, J., & Bruce, V. (1990). Resolving semantically induced tip-of-the-tongue states for proper nouns. *Memory and Cognition, 18(4)*, 339-347.

Brosgole, L., PlaHovinsak, T.J., Roig, M. & Notaro, J.P. (1985). *The role of perceived distance in determining apparent visual size.* Bulletin of the Psychonomic Society, 23(6), 489-492.

Brousil, J.K., & Smith, D.R. (1967). A threshold logic network for shape invariance. *IEEE Transactions on Electronic Computers EC-16,* 818-828.

Bundesen, C., Farrell, J.E., & Larsen, A. (1983). Visual apparent movement: Transformations of size and orientation. *Perception, 12(5),* 549-558.

Bundesen, C., & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance, 1(3),* 214-220.

Bundesen, C., Larsen, A., & Farrell, J.E. (1981). Mental transformations of size and orientation. In A. Baddeley & J. Long (Eds.), *Attention and performance IX* (pp. 279-294). Hillsdale, NJ: Erlbaum.

Butavicius, M.A. (2001). The recognition of planar rotated and scaled images: Shape and its internal representation. *Australian Journal of Psychology – Supplement, 53,* 142.

Butavicius, M.A. (2002). The recognition of rotated and scaled objects: The normalization versus invariant features debate. *Australian Journal of Psychology, 54,* 49.

Caelli, T.M., Squire, D.M. & Wild, T.P. (1993). Model-based neural networks. *Neural Networks, 6(5),* 613-625.

Carpenter, P.A., & Just, M.A. (1978). Eye fixations during mental rotation. In J.W. Senders, D.F. Fisher, and R.A. Monty (Eds.), *Eye movements and the higher psychological functions* (pp. 115-133). Hillsdale, NJ: Erlbaum.

Casasent, D., & Psaltis, D. (1976). Position, rotation and scale invariant optical correlation. *Applied Optics, 15,* 1795-1799.

Cavanagh, P. (1978). Size and position invariance in the visual system. *Perception, 7,* 167-177.

Cavanagh, P. (1989). Multiple analyses of orientation in the visual system. In D. M.-K. Lam and C. Gilbert (Eds.), *Neural Mechanisms of Visual Perception* (pp. 52-61) . Houston: Gulf Publishing Co.

Cave, K.R., & Kosslyn, S.M. (1989). Varieties of size-specific visual selection. *Journal of Experimental Psychology: General, 118(2),* 148-164.

Cherry, C. (1961). Two ears - one world. In W.A. Rosenblith (Ed.), *Sensory communications: Contributions to the symposium on principles of sensory communication* (pp. 99-119). Cambridge, Massachusetts: MIT Press.

Churchland, P.S. (1986). *Neurophilosophy.* Cambridge, London: Bradford books, MIT Press.

Cohen, D., & Kubovy, M. (1993). Mental rotation, mental representation, and flat slopes. *Cognitive Psychology, 25*, 351-382.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd edition).* Lawrence Erlbaum Associates: Hillsdale, N.J.

Cohen, M.S., Kosslyn, S.M., Breiter, H.C., DiGirolamo, G.J., Thompson, W.L., Anderson, A.K., Bookheimer, S.Y., Rosen, B.R., & Belliveau, J.W. (1996). Changes in cortical activity during mental rotation: A mapping study using functional MRI. *Brain, 119,* 89-100.

Cooper, E.E., Biederman, I., & Hummel, J.E. (1992). Metric invariance in object recognition: A review and further evidence. *Canadian Journal of Psychology, 46(2)*, 191-214.

Cooper, L.A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology, 7*, 20-43.

Cooper, L.A. (1976). Demonstration of a mental analog of an external rotation. *Perception and Psychophysics, 19*, 296-302.

Cooper, L.A., & Shepard, R.N. (1973). Chronometric studies of the rotation of mental images. In W.G. Chase (Ed.) *Visual information processing* (135-142). New York: Academic Press.

Corballis, M.C. (1988). Recognition of disoriented shapes. *Psychological Review, 95(1)*, 115-123.

Corballis, M.C., & Beale, I.L. (1976). *The psychology of left and right.* New York: Wiley.

Corballis, M.C., & Blackman, A.R. (1990). The effect of apparent movement on mental rotation. *Memory and Cognition, 18 (5)*, 551-555.

Corballis, M.C., & McLaren, R. (1982). Interaction between perceived and imagined rotation. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 215-224.

Corballis, M.C. & McLaren, R. (1984). Winding one's ps and qs: Mental rotation and mirror-image discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 318-327.

Corballis, M.C., Zbrodoff, N.J., & Roldan, C.E. (1976). What's up in mental rotation? *Perception and Psychophysics, 19*, 525-530.

Corballis, M.C., Zbrodoff, N.J., Shetzer, L.I., & Butler, P.B. (1978). Decisions about identity and orientation of rotated letters and digits. *Memory and Cognition, 6(2),* 98-107.

Corcoran, D.W.J., & Besner, D. (1975). Application of the Posner technique to the study of size and brightness irrelevancies in letter pairs. In P.M.A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 613-629). London: Academic Press.

De Caro, S.A., & Reeves, A. (2000). Rotating objects to determine orientation, not identity: Evidence from a backward-masking/dual-task procedure. *Perception and Psychophysics, 62(7),* 1356-1366.

De Haan, E.H.F. (2001). Face perception and recognition. In Rapp, B. (Ed.) *The handbook of cognitive neuropsychology: What deficits reveal about the human mind* (pp. 75-99). Philadephia, PA: Psychology Press / Taylor & Francis.

Deese, J. (1969). Behaviour and fact. *American Psychologist, 4,* 515-522.

Dennett, D.C. (1978). *Brainstorms.* Massocks: Harvester Press.

Dodwell, P.C. (1971). On perceptual clarity. *Psychological Review, 78,* 275-289.

Dow, B.M., Snyder, A.Z., Vautin, R.G., & Bauer, R. (1981). Magnification factor and receptive field size in foveal striate cortex of Monkey. *Experimental Brain Research, 44,* 213-228.

Edelman, S., & Weinshall, D. (1998). Computational approaches to shape constancy. In V. Walsh, & J. Kulikowski (Eds.), *Perceptual constancy: Why things look the way they do* (pp. 124-143). Cambridge, UK: Cambridge University Press.

Eley, M.G. (1982). Identifying rotated letter-like symbols. *Memory and Cognition 10(1),* 25-32.

Ellis, R., Allport, D.A., Humphreys, G.W., & Collis, J. (1989). Varieties of object constancy. *The Quarterly Journal of Experimental Psychology, 41A(4),* 775-796.

Engel, G.R. (1969). The autocorrelation function and binocular brightness mixing. *Vision Research, 9,* 1111-1130.

Engel, G.R., Dougherty, W.G., & Jones, G.B. (1973). Correlation and letter recognition. *Canadian Journal of Psychology, 27,* 317-326.

Eriksen, C.W., & St. James, J.D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics, 40(4),* 225-240.

Eriksen, C.W., & Yeh, Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 11(5)*, 583-597.

Farah, M.J., & Hammond, K.M. (1988). Mental rotation and orientation-invariant object recognition: Dissociable processes. *Cognition, 29*, 29-46.

Farah, M.J., & Peronnet, F. (1989). Event-related potentials in the study of mental imagery. *Journal of Psychophysiology, 3*, 99-109

Farah, M.J., Hammond, K.M., Levine, D.N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology, 20*, 439-462.

Farell, B., & Pelli, D.G. (1993). Can we attend to large and small at the same time? *Vision research, 33(18)*, 2757-2772.

Farrell, J.E., & Shepard, R.N. (1981). Shape orientation and apparent rotational motion. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 477-486.

Finke, R.A., & Freyd, J.J. (1985). Transformations of visual memory induced by implied motions of pattern elements. *Journal of Experimental Psychology: Language, Memory and Cognition, 11*, 780-794.

Fiser, J., & Biederman, I. (1995). Size invariance in visual object priming of grey scale images. *Perception, 24(7)*, 741-78.

Frank, H. (1925). Untersuchung über Sehgrößenkonstanz bei Kindern / Investigations of the constancy of visual size in children. *Psychologische Forschung / Psychological Research, 7*, 137-145.

Freyd, J.J., & Finke, R.A. (1984). Representational Momentum. *Journal of Experimental Psychology: Language, Memory and Cognition, 10(1)*, 126-132.

Freyd, J.J., & Johnson, J.Q. (1987). Probing the time course of representational momentum. *Journal of Experimental Psychology: Language, Memory and Cognition, 13*, 259-268.

Freyd, J.J. (1987). Dynamic Mental Representations. *Psychological Review, 94(4)*, 427-438.

Garner, W.R. & Clement, D.E. (1963). Goodness of pattern and pattern uncertainty. *Journal of Verbal Learning and Verbal Behavior, 2(5-6)*, 446-452.

Glezer, V.D., Leushina, L.I., Neuskaya, A.A., & Prazdukova, N.V. (1974). Studies on visual pattern recognition in man and animals. *Vision research, 14*, 555-583.

Gogel, W.C. (1965). Equidistance tendency and its consequences. *Psychological Bulletin, 64*, 153-163.

Goodale, M.A., & Milner, A.D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15(1),* 20-25.

Goodale, M.A., Milner, A.D., Jacobsen, L.S., & Carey, D.P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature, 349(6305)*, 154-156.

Green, M. (1983). Inhibition and facilitation of apparent motion by real motion. *Vision Research, 23(9)*, 861-865.

Halligan, P.W., & Marshall, J.C. (1993). Homing in on neglect: A case study of visual search. *Cortex, 29*(1), 167-174.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea.* Cambridge, Massachusetts: MIT Press.

Hayes, A., Sacher, G., Thornton, I.M., Sereno, M.E., & Freyd, J.J. (1996). Representational momentum in depth using stereopsis [ARVO Abstract No. 2120]. *Investigative Opthamology and Vision Science, 37 (supp. 3)*, 3467.

Hayward, W.G., & Tarr, M.J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1511-1521.

Holmes, E.J., & Gross, C.G. (1984). Stimulus equivalence after inferior temporal lesions in monkeys. *Behavioural Neuroscience, 98,* 898-901.

Holway, A.H., & Boring, E.G. (1941). Determinants of apparent visual size with distance variant. *American Journal of Psychology, 54,* 21-37.

Holyoak, K.J. (1977). The form of analog size information in memory. *Cognitive Psychology, 9(1),* 31-51.

Horgan, T., & Tienson, J.L. (1989). Representations without rules. *Philosophical Topics, 17,* 147-174.

Howard, I.P. (1982). *Human visual orientation.* New York: Wiley.

Howard, J.H. Jnr, & Kerst, S.M. (1978). Directional effect of size change on the comparison of visual shapes. *American Journal of Psychology, 91(3),* 491-499.

Hubbard, T.L. (1995). Displacement in depth: Representational momentum and boundary extension. *Psychological Research, 59*, 33-47.

Hubel, D.H., & Wiesel, T.N. (1974). Stereoscopic vision in macaque monkey. *Nature, 225,* 41-44.

Hubel, D.H., & Wiesel, T.N. (1979). Brain mechanisms of vision. *Scientific American, 241(3),* 150-162.

Humphrey, G.K., & Lupker, S.J. (1993). Codes and operations in picture matching. *Psychological Research, 55,* 237-247.

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

Humphreys, G.W. (1984). Shape constancy: The effects of changing shape orientation and the effects of changing the position of focal features. *Perception and Psychophysics, 36(1),* 50-64.

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

Huttenlocher, D.P., & Rucklidge, W.J. (1992). *A multi-resolution technique for comparing images using the Hausdorff distance.* (Tech. Rep. No. 93-1321). Ithaca, NY: Cornell University, Department of Computer Science.

Huttenlocher, D.P. & Ullman, S. (1987). Object recognition using alignment. *Proceedings of the First International Conference on Computer Vision, IEEE* Computer Society Press, 102-111.

Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology, 73*, 218-226.

Johansson, G., von Hofsten, C., & Jansson, G. (1980). Event perception. *Annual Review of Psychology, 31,* 27-63.

Jolicoeur, P., & Cavanagh, P. (1992). Mental rotation, physical rotation, and surface media. *Journal of Experimental Psychology: Human Perception and Psychophysics, 18(2),* 371-384.

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

Jolicoeur, P., & Landau, M.J. (1984). Effects of orientation on the identification of simple visual patterns. *Canadian Journal of Psychlogy, 38(1),* 80-93.

Jolicoeur, P., & Milliken, B. (1989). Identification of disoriented objects: Effects of context of prior presentation. *Journal of Experimental Psychology: Learning, Memory and Cognition. 15(2),* 200-210.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory and Cognition, 13(4),* 289-303.

Jolicoeur, P. (1988). Mental rotation and the identification of disoriented parts. *Canadian Journal of Psychology, 42(4),* 461-478.

Jolicoeur, P. (1990). Identification of disoriented objects: A dual system theory. *Mind and Language, 5(4),* 387-410.

Jolicoeur, P. (1992). Identification of disoriented objects: A dual system theory. In G.W. Humphreys (Ed.), *Understanding vision: An interdisciplinary perspective* (pp. 180-198). Oxford, UK: Blackwell.

Jolicoeur, P., Corballis, M.C., & Lawson, R. (1998). The influence of perceived rotary motion on the recognition of rotated objects. *Psychonomic Bulletin and Review, 5(1),* 140-146.

Jolicoeur, P., Snow, D., & Murray, J. (1987). The time to identify disoriented letters: Effects of practice and font. *Canadian Journal Of Psychology, 41(3),* 303-316.

Jolicouer, P., & Besner, D. (1987). Additivity and interaction between size ratio and response category in the comparison of size-discrepant shapes. *Journal of Experimental Psychology: Human Perception and Performance, 13(3),* 478-487.

Jolicouer, P. (1987). A size-congruency effect in memory for visual shape. *Memory & Cognition, 15 (6),* 531-543.

Julesz, B. (1971). *Foundations of Cyclopean Perception.* Chicago: University of Chicago Press.

Just, M.A., & Carpenter, P.A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92(2),* 137-172.

Karwoski, T. (1933). The memory value of size. *Journal of Experimental Psychology, 14,* 539-554.

Kelly, M.H., & Freyd, J.J. (1987). Explorations of representational momentum. *Cognitive Psychology, 19,* 369-401.

Kolers, P.A., Duchnicky, R.L., & Sandstroem, G. (1985). Size in the visual processing of faces and words. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 726-751.

Koriat, A., & Norman, J. (1984). What is rotated in mental rotation? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 421-434.

Koriat, A., & Norman, J. (1988). Frames and images: Sequential effects in mental rotation. *Journal of Experimental Psychology: Learning, Memory and Cognition 14(1),* 93-111.

Koriat, A., Norman, J., & Kimchi, R. (1991). Recognition of rotated letters: Extracting invariance across successive and simultaneous stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 17(2),* 444-457.

Korte, A. (1915). Kinematoskopische Untersuchungen. *Zeitschrift für Psychologie, 72,* 193-206.

Kosslyn, S.M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

Kosslyn, S.M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review, 94,* 148-175.

Kosslyn, S.M., Chabris, C.F., Marsolek, C.J., & Koenig, O. (1992). Categorical versus coordinate spatial relations: Computational analysis and computer simulations. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 562-577.

Kubovy, M., & Podgorny, K. (1981). Does pattern matching require the normalization of size and orientation? *Perception and Psychophysics, 30(1),* 24-28.

Kuhn, T.S. (1962). *The Structure of Scientific Revolutions.* Chicago: The University of Chicago Press.

Langdon, P.M., Mayhew, J.E.W., & Frisby, J.P. (1991). In search of "characteristic view" 3D object representation in human vision using ratings of perceived differences between views. In J.E.W. Mayhew and J.P. Frisby (Eds.) *3D model recognition from stereoscopic cues* (pp. 244-248). Cambridge, MA: MIT Press.

Larsen, A. (1985). Pattern matching: Effects of size ratio, angular difference in orientation, and familiarity. *Perception and Psychophysics, 38,* 63-68.

Larsen, A., & Bundesen, C. (1978). Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance, 4(1),* 1-20.

Larsen, A., Bundesen, C., Kyllingsbæk, S., Paulson, O.B., & Law, I. (2000). Brain activation during mental transformation of size. *Journal of Cognitive Neuroscience, 12(5),* 763-774.

Larsen, A., McIlhagga, W., & Bundesen, C. (1999). Visual pattern matching: Effects of size ratio, complexity and similarity in simultaneous and successive matching. *Psychological Research, 62*, 280-288.

Lawson, R., & Humphreys, G.W. (1996). View-specificity in object processing: Evidence from picture matching. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 395-416.

Lawson, R., & Humphreys, G.W. (1998a). The neuropsychology of visual object constancy. In V. Walsh, & J. Kulikowski (Eds.), *Perceptual constancy: Why things look the way they do* (pp. 210-228). Cambridge, UK: Cambridge University Press.

Lawson, R., & Humphreys, G.W. (1998b). View-specific effects of depth rotation and foreshortening on the initial recognition and priming of familiar objects. *Perception and Psychophysics, 60,* 1052-1066.

Lawson, R., & Jolicoeur, P. (1998). The effects of plane rotation on the recognition of brief masked pictures of familiar objects. *Memory and Cognition, 26(4),* 791-803.

Lawson, R., & Jolicoeur, P. (1999). The effect of prior experience on recognition thresholds for plane-disoriented pictures of familiar objects. *Memory and Cognition, 27(4),* 751-758.

Lawson, R., & Humphreys, G.W. (1996). View specificity in object processing: Evidence from picture matching. *Journal of Experimental Psychology: Human Perception and Performance, 22(2),* 395-416.

Lawson, R., Humphreys, G.W., & Watson, D.G. (1994). Object recognition under sequential viewing conditions: Evidence for viewpoint specific recognition procedures. *Perception, 23(5),* 595-614.

Lawson, R., Humphreys, G.W., & Jolicoeur (2000). The combined effects of plane disorientation and foreshortening on picture naming: One manipulation or two? *Journal of Experimental Psychology: Human Perception and Performance, 26(2),* 568-581

Layman, S., & Greene, E. (1988). The effect of stroke on object recognition. *Brain and Cognition, 7,* 87-114.

Leek, E.C. (1998). Effects of stimulus orientation on the identification of common polyoriented objects. *Psychonomic Bulletin & Review, 5(4),* 650-658.

Leushina, L.I. (1967). Evidence for independence of channels for recognition of form and estimation of the size of an image. In (Anonymous author) *The Mechanisms of Visual Recognition* (pp. 128-134). Leningrad: Nauka (in Russian).

Licklider, J.C.R. (1959). Three auditory theories. In S. Koch (Ed.), *Psychology: A study of science. Study 1. Conceptual and systematic. Volume 1. Sensory, perceptual, and physiological formulations* (pp. 41-144). New York : McGraw-Hill.

Lowe, D.G. (1985). *Perceptual organization and visual recognition*. Boston: Kluwer.

Lowe, D.G. (1986). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence, 31(3),* 355-395.

Lu, Z.L., & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research, 35(19)*, 2697-2722.

Lueschow, A., Miller, E.K., & Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. *Cerebral Cortex, 5(4)*, 523-531.

Maki, R.H. (1986). Naming and locating the tops of rotated pictures. *Canadian Journal of Psychology, 40(4)*, 368-387.

Marg, E., & Adams, J.E. (1970). Evidence for a neurological zoom in vision from angular changes in some receptive fields of single neurons with changes in fixation distance in the human visual cortex. *Experientia, 26,* 270-271.

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

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

Marr, D. (1977). Analysis of occluding contour. *Proceedings of the Royal Society (London), B, 197*, 441-475.

Martin, M. (1979). Local and global processing: The role of sparsity. *Memory and Cognition, 7*, 476-484.

McClelland, J.L., & Rumelhart, D.E. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models.* Cambridge, Massachusetts: MIT Press.

McCulloch, W.S., & Pitts, W.H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5,* 115-133.

McKone, E., & Grenfell, T. (1999). Orientation invariance in naming rotated objects: Individual differences and repetition priming. *Perception and Psychophysics, 61(8),* 1590-1603.

McMullen, P.A., & Jolicouer, P. (1989). *Processing the identity and orientation of object line drawings.* Paper presented at the 30th Annual Meeting of the Psychonomic Society, 1989, Atlanta, Georgia.

McMullen, P.A., & Jolicoeur, P. (1990). The spatial frame of reference in object naming and discrimination of left-right reflections. *Memory and Cognition, 18(1),* 99-115.

McMullen, P.A., & Jolicoeur, P. (1992). Reference frame and effects of orientation on finding tops of rotated objects. *Journal of Experimental Psychology: Human Perception and Performance, 18(3)*, 807-820.

Metzler, J., & Shepard, R.N. (1974). Transformational studies of the internal representation of three-dimensional objects. In R. Solso (Ed.) (1974). *Theories in cognitive psychology: The Loyola Symposium* (pp. 147-202). Potomac, Md: Lawrence Erlbaum Associates.

Metzler, J., & Shepard, R.N. (1982). Transformational studies of the internal representation of three-dimensional objects. In R.N. Shepard, & L.A., Cooper (Eds.), *Mental images and their transformations* (pp. 25-71). Cambridge, Massachusetts: MIT Press.

Milliken, B., & Jolicoeur, P. (1992). Size effects in visual recognition memory are determined by perceived size. *Memory and Cognition, 20(1),* 83-95.

Milner, A.D., Perrett, D.I., Johnston, R.S. Benson, P.J., Jordan, T.F. Heeley, T.R., Bettucci, D.W., Mortara, D., Mutani, R., Terazzi, E., et al. (1991). Perception and action in visual form perception. *Brain, 114*, 405-428.

Mishkin, M., Ungerleider, L.G., & Macko, K.A. (1983) Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6*, 414-417.

Moyer, R.S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception and Psychophysics, 13(2),* 180-184.

Munger, M.P., Solberg, J.L., Horrocks, K.K. (1999). The relationship between mental rotation and representational momentum. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25(6),* 1557-1568.

Murray, J.E. (1995). Imagining and naming rotated natural objects. *Psychonomic Bulletin and Review. 2(2)*, 239-243.

Murray, J.E. (1999). Orientation-specific effects in picture matching and naming. *Memory and Cognition, 27(5),* 878-889.

Murray, J.E., Jolicouer, P., McMullen, P.A., & Ingleton, M. (1993). Orientation-invariant transfer of training in the identification of rotated natural objects. *Memory and Cognition, 21(5),* 604-610.

Newell, F.N., & Findlay, J.M. (1997). The effect of depth rotation on object identification. *Perception, 26(11)*, 1231-1257.

Niall, K. (1997). “Mental rotation", pictured rotation and tandem rotation in depth. *Acta Psychologica, 95(1)*, 31-83.

Pachella, R.G. (1974). The interpretation of reaction time in information-processing research. In B.H. Kantowitz (Ed.), *Human information processing: Tutorials* *in performance and cognition* (pp. 41-82). Hillsdale, New Jersey: Lawrence Erlbaum.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory and Cognition, 3(6)*, 635-647.

Palmer, S.E. (1982). Symmetry transformation and the structure of perceptual systems. In J. Beck (Ed.), *Organization and representation in perception* (pp. 95-144). Hillsdale, NJ: Lawrence Erlbaum.

Palmer, S.E. (1989). Reference frames in the perception of shape and orientation. In B.E. Shepp & S. Ballestros (Eds.), *Object perception: Structure and process* (pp.121-163). Hillsdale, NJ: Lawrence Erlbaum.

Palmer, S.E. (1999). *Vision science: Photons to Phenomenology.* Cambridge, Massachusetts: MIT Press.

Parker, D.E., Poston, R.L., & Gulledge, W.L. (1983). Spatial orientation: Visual-vestibular-somatic interaction. *Perception & Psychophysics, 33,* 139-146.

Parsons, L.M. (1987a). Imagined spatial transformations of one's body. *Journal of Experimental Psychology: General, 116(2)*, 172-191.

Parsons, L.M. (1987b). Imagined spatial transformations of one's hands and feet. *Cognitive Psychology, 19*, 178-241.

Pasto, L., & Burack, J.A. (1997). A developmental study of visual attention: Issues of filtering efficiency and focus. *Cognitive development, 12(4),* 427-439.

Peitgen, H-O., Jürgen, H., & Saupe, D. (1992). *Fractals for the classroom - Part One: Introduction to fractals and chaos.* New York: Springer-Verlag.

Peronnet, F., & Farah, M.J. (1989). Mental rotation: An event-related potential study with a validated mental rotation task. *Brain and Cognition, 9*, 279-288.

Pierret, A., & Peronnet, F. (1994). Mental rotation and mirror-image discrimination. *Perceptual and Motor Skills, 78*, 515-524.

Pomerantz, J.R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General, 112*, 516-540.

Posner, M.I., & Mitchell, R.F. (1967). Chronometric analysis of classification. *Psychological Review, 74(5)*, 392-409.

Preparata, F.P., & Shamos, M.I. (1985). *Computational geometry.* New York: Springer-Verlag.

Pylyshyn, Z.W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 87*, 16-45

Reitboeck, H.J., & Altmann, J. (1984). A model for size- and rotation-invariant pattern processing in the visual system. *Biological Cybernetics, 51,* 113-121.

Richter, W., Somarjai, R., Summer, S.R., Jamarsz, M., Menon, R.S., Gati, J.S., Georgopoulos, A.P., Tegeler, C., Vgurbill, K., & Kim, S.G. (2000). Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *Journal of Cognitive Neuroscience, 12(2),* 310-320.

Rock, I., & DiVita, J. (1987). A case of viewer-centred perception. *Cognitive Psychology, 21,* 280-293.

Rock, I. (1956). The orientation of forms on the retina and in the environment. *American Journal of Psychology, 69,* 513-528.

Rock, I. (1973). *Orientation and form.* New York: Academic Press.

Rock, I. (1974). The perception of disoriented figures. *Scientific American, 230,* 78-85.

Rock, I. (1975). *An introduction to perception.* New York: Collier-Macmillan.

Rock, I., & Heimer, W. (1957). The effect of retinal and phenomenal orientation on the perception of form. *American Journal of Psychology, 70,* 493-511.

Rösler, F., Heil, M., Bajric, J., Pauls, A.C., & Hennighausen, E. (1995). Patterns of cerebral activation while mental images are rotated and changed in size. *Psychophysiology, 32,* 135-149.

Rösler, F., Schumacher, G., & Sojka, B. (1990). What the brain reveals when it thinks: Event-related potentials during mental rotation and mental arithmetic. *German Journal of Psychology, 14,* 185-203.

Ross, H.E., & Plug, C. (1998). The history of size constancy and size illusions. In V. Walsh, & J. Kulikowski (Eds.) *Perceptual constancy: Why things look the way they do* (pp. 499-528). Cambridge, UK: Cambridge University Press.

Rumelhart, D.E., & McClelland, J.L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations.* Cambridge, Massachusetts: MIT Press.

Santee, J.L., & Egeth, H.E. (1980). Selective attention in the speeded classification and comparison of multidimensional stimuli. *Perception and Psychophysics, 28(3),* 191-204.

Schacter, D.L., Dulaney, S.M., & Merikle, E.P. (1990). Priming of nonverbal information and the nature of implicit memory. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp.83-123). New York: Academic Press.

Schwartz, E.L. (1976). Retinotopic mapping in striate cortex. *Physiologist, 19,* 358.

Schwartz, E.L. (1977a). Spatial mapping in primate sensory projection: Analytic structure and relevance to perception. *Biological Cybernetics, 25,* 181-194.

Schwartz, E.L. (1977b). The development of specific visual connections in the goldfish and the monkey: Outline of a geometric theory of receptopic mapping. *Journal of Theoretical Biology, 69,* 658-683.

Schwartz, E.L. (1977c). Afferent geometry in the primate visual cortex and the generation of neuronal trigger features. *Biological Cybernetics, 29,* 1-24.

Schwartz, E.L. (1980a). Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding. *Vision Research, 20,* p. 645-669.

Schwartz, E.L. (1980b). A quantitative model of the functional architecture of human striate cortex: Cortical anatomy and visual illusion. *Biological Cybernetics, 37,* p. 63-76.

Schwartz, E.L. (1981a). Cortical anatomy, size invariance, and spatial frequency analysis. *Perception, 10,* 455-468.

Schwartz, E.L. (1981b). Cortical mapping and computational anatomy: A projective invariant recursive flow model of visual coding. In E. Grastyán and P. Molár (Eds.), *Advances in neurophysiological sciences* (pp. 431-436). New York: Elsevier.

Schwartz, E.L. (1982). Columnar architecture and computational anatomy in primate visual cortex: Segmentation and feature extraction via spatial frequency coded difference mapping. *Biological Cybernetics, 42,* 157-168.

Schwartz, E.L. (1983). Cortical mapping and perceptual invariance: A reply to Cavanagh. *Vision Research, 23(8),* 831-835.

Sekuler, R., & Nash, D. (1972). Speed of size scaling in human vision. *Psychonomic Science, 27(2),* 93-94.

Sergent, J., & Signoret, J-L. (1993). Implicit access to knowledge derived from unrecognised faces in prosopagnosia. *Cerebral Cortex, 2(5),* 389-400.

Shepard, R.N., & Judd, S.A. (1976). Perceptual illusion of rotation of three-dimensional objects. *Science, 191,* 952-954.

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

Shepard, R.N. (1981). Psychophysical complementarity. In M. Kubovy, & J.R. Pomerantz (Eds.), *Perceptual Organization* (pp. 279-341). Hillsdale, NJ: Lawrence Erlbaum.

Shepard, R.N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review, 91(4),* 417-447.

Shepard, R.N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin and Review, 1(1),* 2-28.

Shepard, S. & Metzler, D. (1988). Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance, 14(1),* 3-11.

Shinar, D., & Owen, D.H. (1973). Effects of form rotation on the speed of classification: The development of shape constancy. *Perception and Psychophysics, 14(1),* 149-154.

Shoham, D., & Ullman, S. (1988). Aligning a model to an image using minimal information. *Proceedings of the Second International Conference on Computer Vision, IEEE* Computer Society Press, 259-263.

Shulman, G.L., & Wilson, J. (1987). Spatial frequency and selective attention to spatial location. *Perception, 16(1),* 103-111.

Simion, F., Bagnara, S., Roncata, S., & Umilta, C. (1982). Transformation processes upon the visual code. *Perception and Psychophysics, 31(1),* 13-25.

Smith, J.D., & Marg, E. (1975). Zoom neurons in visual cortex: Receptive field enlargements with near fixation in monkeys. *Experientia, 31(3),* 323-326.

Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174-215.

Somjen, G. (1972). *Sensory Coding in the Mammalian Nervous System.* New York: Appleton-Century-Croft.

Spearman, C.E. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology, 15,* 72-101.

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

Stefanova, N. (1970). On the role of size in the process of recognition of visual objects. In (Anonymous author) *Investigation of Principles of Information Processing in Visual Systems* (pp. 161-168). Leningrad, Nauka (in Russian).

Sternberg, S. (1966). High speed scanning in human memory. *Science, 153,* 652-654.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica, 30,* 276-315.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643-662.

Takano, Y. (1989). Perception of rotated forms: A theory of information types. *Cognitive Psychology, 21,* 1-59.

Talbot, S.A., & Marshal, W.H. (1941). Physiological studies on neural mechanisms of visual localisation and discrimination. *American Journal of Ophthalmology, 24,* 1255-1263.

Tanaka, K., Saito, H.-A., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology, 66,* 170-189.

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

Tarr, M.J., & Pinker, S. (1990) When does human object recognition use a viewer-centred reference frame? *Psychological Science, 1(4),* 253-256.

Tarr, M.J., & Bülthoff, H.H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1494-1505.

Templeton, W.B. (1973). The role of gravitational cues in the judgment of visual orientation. *Perception and Psychophysics, 14*, 451-457.

Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception, 9,* 483-4.

Tyler, C. W. (1996). *Human symmetry and its computational analysis*. Utrecht, The Netherlands: VSP.

Tyler, C.W., & Miller, R.H. (1994). Pattern identification by trajectory analysis in autocorrelation hyperspace. *Proceedings of the World Congress on Neural Networks, III,* 312-316.

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

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

Uttal, W.R. (1975). *An Autocorrelation Theory of Form Detection.* Hillsdale, New Jersey: Lawrence Erlbaum.

Vickers, D. (2002). *A generative transformational model of human visual perception.* Client Report, Defence Science and Technology Organisation.

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

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

Weiskrantz, L., & Saunders, R.C. (1984). Impairments of visual object transforms in monkeys. *Brain, 107*, 1033-1072

Welford, A.T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years progress. *Ergonomics, 3,* 189-230.

White, M.J. (1980). Naming and categorization of tilted alphanumeric characters do not require mental rotation. *Bulletin of the Psychonomic Society, 15*, 153-156.

Woodford, R.S., & Schlosberg, H. (1954). *Experimental psychology*. New York: Holt, Rinehart & Winston.

Yonelinas, A.P., & Jacoby, L.L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language, 34*, 622-643.

Yoshino, A., Inoue, M., & Suzuki, A. (2000). A topographic electrophysiological study of mental rotation. *Cognitive Brain Research, 9(2)*, 121-124.

Zelinsky, G.J., Rao, R.P.N., Hayhoe, M.M., & Ballard, D.H. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science, 8(6)*, 448-453.