Shape from Gradients

A psychophysical and computational study of the role complex illumination gradients, such as shading and mutual illumination, play in three-dimensional shape perception

Glen HARDING

Submitted for the degree of Doctor of Philosophy

Bradford School of Optometry and Vision Science

Supervisor: Dr Marina Bloj

2013
Abstract

Shape from Gradients

Glen Harding

Keywords: Shape Perception, Colour Vision, Psychophysics, Gradients, Shading, Bayes, Cue Combination.

The human visual system gathers information about three-dimensional object shape from a wide range of sources. How effectively we can use these sources, and how they are combined to form a consistent and accurate percept of the 3D world is the focus of much research. In complex scenes inter-reflections of light between surfaces (mutual illumination) can occur, creating chromatic illumination gradients. These gradients provide a source of information about 3D object shape, but little research has been conducted into the capabilities of the visual system to use such information.

The experiments described here were conducted with the aim of understanding the influence of chromatic gradients from mutual illumination on 3D shape perception. Psychophysical experiments are described that were designed to investigate: If the human visual system takes account of mutual illumination when estimating 3D object shape, and how this might occur; How colour shading cues are integrated with other shape cues; The relative influence on 3D shape perception of achromatic (luminance) shading and chromatic shading from mutual illumination. In addition, one chapter explores a selection of mathematical models of cue integration and their applicability in this case.
The results of the experiments suggest that the human visual system is able to quickly assess and take account of colour mutual illuminations when estimating 3D object shape, and use chromatic gradients as an independent and effective cue. Finally, mathematical modelling reveals that the chromatic gradient cue is likely integrated with other shape cues in a way that is close to statistically optimal.

Acknowledgements

I would like to thank my principal supervisor Marina Bloj for her invaluable help and friendship throughout the process of producing this work.

Special thanks also go to Julie Harris and George Lovell for their many helpful discussions, and to my wife Emma for her support.

Publication of Content

Some of the content of this thesis has been published or submitted for publication in the following articles:


Harding G., Lovell P.G., Harris J.M., Bloj M. Chromatic Gradients can be an Effective Cue to Shape. Submitted to Journal of Vision, April 2014.
Table of Contents

ABSTRACT ............................................................................................................................................. I
ACKNOWLEDGEMENTS ......................................................................................................................II
PUBLICATION OF CONTENT .............................................................................................................II
TABLE OF CONTENTS .......................................................................................................................... III

CHAPTER 1 - GENERAL INTRODUCTION .......................................................................................1

CHAPTER 2 - DEPTH AND 3D SHAPE PERCEPTION ........................................................................2
  2.1 SHAPE CUES FROM ILLUMINATION PHENOMENA ..............................................................5
    2.1.1 Shadows ................................................................................................................................7
    2.1.2 Highlights ............................................................................................................................9
    2.1.3 Three Dimensional Shape-from-Shading ........................................................................... 12
    2.1.4 Perception of 3D Shape from Shading .............................................................................. 15
      2.1.4.1 Separating Shading from Reflectance Changes .......................................................... 15
      2.1.4.1.1 Mutual Illumination and Surface Reflectance ......................................................... 17
      2.1.4.1.2 Colour as a Means of Disentangling Shape and Reflectance ................................. 19
      2.1.4.2 Mechanisms by which Humans Derive Shape from Illumination Cues ...................... 22
        2.1.4.2.1 Illumination Direction and Shape-from-Shading ................................................. 25
        2.1.4.2.2 Mutual Illumination and Object Shape/Position ................................................. 28
    2.1.5 Shading and Shape Constancy ......................................................................................... 30
  2.2 SHAPE FROM CONTOUR / PERSPECTIVE ............................................................................. 31

CHAPTER 3 – 3D SHAPE CUE INTEGRATION ..................................................................................35
  3.1 MODELS OF DEPTH / 3D SHAPE CUE INTEGRATION ......................................................... 35
    3.1.1 Weighted Linear Combination ......................................................................................... 36
    3.1.2 Bayesian Cue Integration .................................................................................................. 39
      3.1.2.1 Weighting of Prior Constraints ............................................................................. 40
    3.1.3 What type of cue combination model is best? ................................................................. 41
  3.2 CUE CONFLICTS .......................................................................................................................... 45

CHAPTER 4 - VISUAL PSYCHOPHYSICS ......................................................................................49
  4.1 MEASUREMENT OF DETECTION AND DISCRIMINATION THRESHOLDS ....................... 49
    4.1.1 Method of Adjustment ....................................................................................................... 51
    4.1.2 Ascending and Descending Limits ................................................................................. 51
    4.1.3 Staircase Methods .............................................................................................................. 52
    4.1.4 Method of Constant Stimuli ............................................................................................. 52
    4.1.5 Forced Choice Procedures ............................................................................................... 53
  4.2 SUPRA-THRESHOLD AND OTHER EXPERIMENTAL PROCEDURES ................................ 54
    4.2.1 Magnitude Estimation ....................................................................................................... 55
    4.2.2 Ranking and Scaling ......................................................................................................... 55
    4.2.3 Reaction Times .................................................................................................................. 56
  4.3 NOISE AND SIGNAL DETECTION THEORY ......................................................................... 57
    4.3.1 Receiver Operating Characteristic (ROC) Curves .......................................................... 60
  4.4 OTHER CONSIDERATIONS ........................................................................................................ 61

CHAPTER 5 – GENERAL METHODS ............................................................................................63
  5.1 COMPUTER RENDERING ........................................................................................................... 63
    5.1.1 Ray Tracing and Monte Carlo methods .......................................................................... 64
    5.1.2 RADIANCE Synthetic Imaging System .......................................................................... 65
    5.1.3 ‘Hyperspectral’ rendering .................................................................................................. 66
  5.2 STIMULI DISPLAY ...................................................................................................................... 69
    5.2.1 Colour Spaces ................................................................................................................... 69
    5.2.2 Cambridge Research Systems ViSaGe ........................................................................... 75
    5.2.3 Display Devices ................................................................................................................ 76
      5.2.3.1 Display Calibration ...................................................................................................... 78
CUES

CHAPTER 6 - EXPERIMENT 1: COLOUR ILLUMINATION GRADIENTS AS A CUE TO 3D SHAPE ................................................................. 92

6.1 METHOD ............................................................................. 94
6.1.1 Observers ................................................................... 97
6.2 RESULTS AND ANALYSIS .................................................. 98
6.2.1 Discriminability of the stimuli ........................................ 98
6.2.2 Results of the Monocular Conditions ............................. 101
6.2.2.1 Comparison of perceived shape across monocular conditions ........................................ 108
6.2.3 Results of the Binocular Conditions .............................. 112
6.2.3.1 ‘Stereo-outline’ Condition Results ............................ 112
6.2.3.2 ‘Stereo Gradient+Outline’ Condition Results ............ 114
6.2.3.3 Effects of the Addition of the Gradient Cue to Stereo Stimuli ........................................... 117
6.3 DISCUSSION ..................................................................... 119
6.3.1 Do observer settings need correcting because of the task? 119
6.3.2 How effectively are the individual cues used?.......... 121
6.3.3 What are the Sources of Bias in Observer Settings? .... 123
6.3.3.1 Assessing Biases Inherent in the Task .................... 128
6.3.4 Cue Combination ........................................................ 133
6.4 EXPERIMENT 1 SUMMARY ............................................... 138

CHAPTER 7: MODELLING THE INTEGRATION OF THE GRADIENT AND OUTLINE CUES ......................................................... 140

7.1 RELIABILITY WEIGHTED LINEAR COMBINATION MODEL ........................................................................ 142
7.1.1 Predictions of the Weak Fusion Model ....................... 143
7.1.2 Cue Reliability ........................................................... 146
7.2 BAYESIAN CUE-COMBINATION MODELS .................. 148
7.2.1 Bayesian Modelling of the Single Cue Data ................ 150
7.2.1.1 Posterior distribution and decision rule .................. 151
7.2.1.2 Fitting the Models to the Data .............................. 153
7.2.2 The Models .............................................................. 154
7.2.2.1 Model 1: Gaussian Prior and Likelihoods ............. 154
7.2.2.2 Model 2: Non-Gaussian Prior ................................. 158
7.2.2.3 Model 3: Non-Gaussian Likelihoods ........................ 161
7.2.2.4 Model Likelihood .................................................. 165
7.2.2.5 Model 4: ‘Best Guess’ Simplified Model ................. 167
7.2.3 Predicting the Combined-Cue Data ......................... 172
7.2.3.1 Model Predictions ................................................. 172
7.2.4 Discussion of Bayesian Modelling .............................. 177
7.2.4.1 Sources of Bias ................................................... 178
7.2.4.2 Model Performance .............................................. 179
7.2.4.3 Model Validity .................................................... 182
7.2.4.3.1 Why Use a Bayesian model? ......................... 182
7.2.4.3.2 How Much Can Bayesian Model Really Tell Us? 183
7.3 MODELLING SUMMARY .................................................. 185

CHAPTER 8 - EXPERIMENT 2: LEARNING TO USE THE GRADIENT CUE ........................................................... 187

8.1 METHOD AND STIMULI .................................................. 190
8.1.1 Training Videos ........................................................ 191
8.1.2 Observers .................................................................. 192
8.2 RESULTS AND ANALYSIS ............................................... 193
8.3 DISCUSSION ................................................................. 199
8.3.1 Can Priors Disambiguate the Gradient Cue? ............. 199
8.3.2 What do Observers Need to Learn in Order to Use the Gradient Cue? 201
CHAPTER 9 - EXPERIMENT 3: SHAPE FROM CHROMATIC GRADIENTS

9.1 METHODS

9.1.1 Apparatus

9.1.2 Stimuli

9.1.3 Observers

9.1.4 Single-Cue Conditions Procedure

9.1.5 Cue-Conflict Condition Procedure

9.2 RESULTS AND ANALYSIS: SINGLE-CUE CONDITIONS

9.3 RESULTS AND ANALYSIS: CUE-CONFLICT CONDITION

9.4 DISCUSSION

9.4.1 Luminance and Chromatic Gradients as Separate Cues to Shape

9.4.2 Luminance and Chromatic Gradient Cue Combination

9.5 EXPERIMENT 3 SUMMARY

CHAPTER 10 - GENERAL DISCUSSION

10.1 SUMMARY OF EXPERIMENTAL FINDINGS

10.1.1 Shape Perception from Realistic Complex Colour Gradients

10.1.2 Disambiguating Complex Colour Gradient Cues - Are 'Priors' Enough?

10.1.3 Chromatic Gradients as an Independent Shape Cue

10.2 FINDINGS FROM THE MODELLING

10.3 CONCLUDING REMARKS AND SUGGESTIONS FOR FUTURE RESEARCH

REFERENCES
Chapter 1 - General Introduction

This thesis is primarily concerned with the perception of three-dimensional object shape. The work described here aims to investigate the role of complex illumination gradients on shape perception, particularly the contribution of chromatic gradients from mutual illuminations. Colour is not often thought of when considering the visual factors that contribute towards depth or 3D shape perception, yet evidence exists that in some cases a significant amount of information about three-dimensional scene structure is available through colour, and that the human visual system may be capable of making some use of this information.

In this thesis I first describe the background literature associated with three-dimensional object shape perception, concentrating on illumination cues. In the second part of this thesis I describe general methods used for the study of the topics in questions. Thirdly I detail a number of experiments, and theoretical models, designed to test and understand the contribution chromatic gradients from mutual illuminations may have on human perception of 3D object shape, together with a presentation and discussion of the results. Finally, I provide a general discussion of the experimental results and modelling, in the context of what is currently known and described in the literature.
Chapter 2 - Depth and 3D Shape Perception

The ability to properly perceive all the spatial dimensions through which we move and live our lives is crucial to human function. It would be very difficult to perform even simple tasks such as walking from place to place, or picking up objects without knowledge of our three-dimensional world. The visual system plays a significant role in how we gain information about the shape and position of ourselves and other objects in three dimensions, but it must solve a complex problem in order to reconstruct all three spatial dimensions from the two-dimensional images presented at the retina of the eyes. Several mechanisms have been shown to be used by the human visual system to reconstruct the three-dimensional world and provide perception of depth and 3D object shape (an overview is given in Bruce, Georgeson, & Green (2003)). Perhaps the most obvious is the process of stereopsis, from binocular disparity. We have two eyes that provide, by virtue of their different positions, two slightly different images of the world. The disparity between these two images is a source of depth information, since the disparity is dependent of the distance of objects, greater for objects closer to the observer (see, for example, Cumming & DeAngelis (2001); Marr & Poggio (1979)). An early demonstration of depth perception from binocular disparity was given by Charles Wheatstone, with the invention of the stereoscope (Wheatstone, 1838). This device used mirrors to present drawings with slightly different viewpoints (a stereo-pair) to each eye. However, disparity is of little use when objects are a significant distance from the observer as the difference between the two retinal images is very small. For people with normal visual acuity, the theoretical limit for obtaining information from
disparity is of the order of a few hundred meters. While stereopsis is known to play an important role in the perception of 3D object shape and position at close distances, it has been shown to have a diminished role for objects more than ten metres from the viewpoint (Cutting & Vishton, 1995; Nagata, 1993).

Stereopsis from binocular disparities does not provide the only means of perceiving depth. This is clear from the fact that we do not perceive a two dimensional world if we close one eye. Other, monocular, mechanisms also contribute to depth perception. The most important of these are briefly outlined below:

**Pictorial size cues**

Cues to depth are contained within the size information in the observed image. For example, familiar objects may be a known size, so it is reasonable to assume them to be far away if they are small and vice versa. A related cue is known as linear perspective: parallel lines appear to converge as they get further away from the observer. Texture gradients can also provide similar size related cues to depth, as the individual elements of a surface texture (for example pebbles on a beach) get smaller the further away they are (see for example Andersen, Braunstein, & Saidpour (1998) for texture and Olson (1974) for perspective).
Motion effects

When there is relative motion between objects and an observer, objects that are closer to the observer will move through a larger visual angle, during any particular time period, than those objects further away. This is known as motion parallax (Gibson, Gibson, Smith, & Flock, 1959; Helmholtz, 1925; Rogers & Graham, 1979). A similar relative motion cue to three-dimensional shape is known as the Kinetic Depth Effect. Here, 3D shape can be determined from two-dimensional image projections when the object is in motion. For example, the cast shadow of a rotating object elicits a much greater impression of solid structure than a stationary shadow (Ullman, 1979; Wallach & O'Connell, 1953).

Shading and shadows and occlusions

Surface shading and shadows, along with the occlusion of objects by others, are phenomena that are created by the three-dimensional geometry (the positions and shapes of objects and light sources) of a scene. As such they contain a wealth of information about depth and 3D shape that the visual system is able to make use of. More details are given below in section 2.1.

Ocular-motor feedback

The relative angular positioning of the eyes (vergence – observer’s eyes must turn inwards if the field of view of both is to be centred on a close object) and optical accommodation (the focal distance of the crystalline lens of the eye) that are required to focus on an object provide feedback
information on the depth of attended objects. See for example (Richards & Milller, 1969).

An important consequence of the fact that multiple cues to depth are available is that the visual system must have some means of combining the information from these different sources, if it is to make best use of the available knowledge, and provide a single percept of depth or 3D shape. This process, known as cue combination or cue integration is central to the study of 3D shape perception, and the work described in this thesis. Details of some of the many previous studies of cue integration are given in Chapter 3. As noted in the general introduction (Chapter 1), this thesis is concerned with the use of illumination phenomena as cues to 3D object shape. Specifically I will investigate chromatic illumination gradients that arise from mutual illuminations. Therefore, in the next section, I describe in detail the shape cues provided by surface lighting phenomena and how colour gradients arise and may provide cues to 3D shape.

2.1 Shape Cues from Illumination Phenomena

Variations in colour and luminance within an image, due to the shape and arrangement of objects and light sources in the scene, provide an important source of 3D shape information for the visual system. For example, see Figure 2.1. In this image, some areas such as the base of the cone, or the bright spot on the top of the sphere have very different luminance (or perceptually, lightness) to other areas of the white objects. This is despite the material of the objects being the same in all these areas, and is a
consequence of the angle of the object surface relative to the observer and the direction of the incident light. In the sections below, I describe how the visual system can often correctly interpret variations in luminance as areas of shadow or highlight due to the geometry of the scene and lighting, and not as areas of different surface reflectance. Similarly colour variations due to 3D shape occur because of mutual illumination (light reflected from one object or surface to another). In Figure 2.1 mutual illuminations are present on the undersides of the objects, so that they take on the colour appearance of the surface they are placed upon. What is known about the capabilities of the visual system to interpret mutual illumination as a product of the geometry of the scene, and not a change in the surface reflectance of the objects, is also outlined in the following sections.

Figure 2.1 An example scene showing colour and luminance gradients arising from both the angle of the object surface relative to the light direction, and light inter-reflections between surfaces/objects (mutual illuminations).

Colour and luminance variations caused by the interaction of light with three-dimensional objects are often collectively referred to as shading or gradients. Ruppertsberg, Bloj, & Hurlbert (2008) and Ruppertsberg, Hurlbert, & Bloj
(2007) define an extrinsic gradient (i.e. one that is not an actual variation in surface reflectance) as a “change in luminance and/or chromaticity that occurs across the surface of an object because of the complex interaction between the positions of light sources, the object on which the gradient appears, and other surrounding objects”.

As outlined in the example above, extrinsic gradients can arise from several different sources: shadows, shading and mutual illumination. Perceptual studies have shown that gradients arising from each of these sources are important for human vision. The following sections describe some of the literature that has been published relating to the information about three-dimensional structure contained in such lighting phenomena, and how this information is used by the human visual system to create both an accurate and consistent reconstruction of three-dimensional objects and scenes.

2.1.1 Shadows

Shadows can be broadly split into two groups: cast shadows (areas that are occluded from a direct light source by another object) and attached shadows (areas of a curved surface that are occluded by other parts of the surface). See Figure 2.2, from Mamassian, Knill, & Kersten (1998).
Cast shadows contain information about object shape, object arrangement and illumination direction in a scene. However, in their review of the perception of cast shadows, Mamassian et al. (1998) assert that cast shadows are not effectively used by the visual system to determine surface shape, but are used as cues to the spatial layout of objects in a scene. Additionally the same authors previously studied the effects of shadows on the perception of depth for moving objects, finding that “cast shadows can produce a very vivid impression of an object moving in depth” (Kersten, Mamassian, & Knill, 1997). The fact that the visual system does not appear to use shadows as a cue to object shape is an interesting result, since information about shape is contained in these shadows. However, because the size and shape of shadows is a product of object shape and light direction, the extraction of reliable object shape information may be a difficult task, due to the wide variations in possible lighting.
While shadows may not be a reliable cue to object shape, it has been shown that cast shadows can be used by the visual system to determine (and disambiguate) light source position (Koenderink, van Doorn, & Pont, 2004). Later work by Mamassian (2004) suggests a possible reason that cast shadows are not used for assessing object shape is that they are processed on a coarse spatial scale, such that fine scale shape features are ignored, for the benefit of a fast interpretation of a scene. Other work such as that by Jacobson & Werner (2004), Lovell, Gilchrist, Tolhurst, & Troscianko (2009) and Porter, Tales, & Leonards (2010) also points towards coarse sampling or insensitivity for cast shadow details.

2.1.2 Highlights

When light reflects from the majority of real surfaces, part of the incident light is scattered by the internal structure of the surface; the reflection of this light is therefore diffuse and the spectrum is modified by the material colour; although this does not hold for all material types, for example many metals. Some of the incident light is however directly reflected, mirror like, from the interface and for many materials is not significantly spectrally modified. This directly reflected light forms specular highlights. An example of a highlight can be seen on the top of the sphere in Figure 2.1 (page 6). Because highlights can have very similar spectral power distributions to the illumination, they could be used by the visual system to gain information about the illuminant. Work by Lee (1986) showed that, computationally, this is possible and suggests that highlights may therefore be important for colour
constancy; a phenomenon where perceptual surface colour appearance can remain very stable despite a significant change in illumination and subsequent change in the chromaticity of light reflected from surfaces. Yang & Maloney (2001) performed a perceptual study where they presented observers with a scene containing highlights and asked observers to adjust a patch within the scene so it appeared achromatic (grey). Since the appearance of achromaticity depends on the perception of the illuminant (a surface with a flat spectral reflectance curve will still have varying chromaticity under different illuminations), they could test if the highlights provided a suitable cue to the illuminant colour. The results showed that the visual system is capable of making use of illuminant information in highlights, with chromaticity changes in specular highlights producing changes in the observer’s achromatic patch settings.

The work described above suggests that the visual system does make some use of highlights as a source of information about the illumination in a scene, at least for estimating illumination chromaticity. However, in theory highlights can also provide the visual system with information about object shape, because the position and shape of highlights in an image are dependent on 3D object shape. For example, estimation of highlight position is used in computer vision research to retrieve light source position when calculating bi-directional surface reflectance distributions (BDRFs) - e.g. Debevec et al. (2004). This is done by measuring the position of highlights on objects of know geometry. Similarly, video eye-tracking software also often relies on the relative positions of the pupil and corneal highlights, known as ‘Purkinje
glints’, to calculate eye position on the assumption of known geometry (Crane & Steele, 1985). Clearly however, the reverse should also be possible: if the light source position is known, or can be accurately estimated, then an estimate of the three-dimensional shape of the object could be made from the position of any highlights. For example, a ‘trick’ long used by painters is to add highlights to the eyes of people in paintings to enhance the perception of a spherical eyeball, rather than a flat circle. Since highlight positions are determined by both the light source and the viewing positions, they also vary slightly in position and shape for each of the observer’s eyes, providing an additional binocular cue.

There is some debate as to the influence of highlights on the perception of three-dimensional object shape. Some studies such as those by Doorschot, Kappers, & Koenderink (1999), Liu & Todd (2004) and Norman, Todd, & Orban (2004) have used psychophysical experiments, asking observers to estimate surface orientation and shape for objects with and without highlights, finding that highlights produce changes and improvements in three-dimensional shape perception. Blake & Bülthoff (1990) also found that observers judged surface gloss to be most realistic if the disparity of the highlights was close to veridical, and that highlights could affect perception of surface curvature. Fleming, Torralba, & Adelson (2004) argue that highlights can be used by the visual system as a cue to 3D shape because systematic patterns of distortion are present across the image of a specular surface, and showed that observers could accurately determine surface orientation for an object defined by specular (mirror like) reflection only. Importantly, this
mechanism does not require any additional cues to shape or knowledge of illumination direction. However, other authors have not been able to find evidence of the use of highlights for the perception of three-dimensional shape. For example Nefs, Koenderink, & Kappers (2006) found no difference between observers' estimated surface shape for matt and glossy objects and Mingolla & Todd (1986) also found little influence of highlights on surface shape perception.

2.1.3 Three Dimensional Shape-from-Shading

As noted at the beginning of this chapter, the interactions of light and object surfaces in three dimensions result in variations in luminance dependent on the relative angle of surfaces, known as shading (Figures 2.1 and 2.2). Therefore, the shading visible in an image of a 3D object provides a source of information about the object shape. This is the basis for 'shape-from-shading' theory, where three-dimensional structure is recovered from surface shading in two-dimensional images.

Much work has been done on solving the shape-from-shading problem computationally. See for example Horn (1975), Horn & Brooks (1986), Ikeuchi & Horn (1981) and Pentland (1984, 1988). This work has concentrated on the application of image processing routines, with the aim of solving specific computer vision problems or creating better algorithms. However, in a more general sense the same problem must be solved by the visual system if it is to make use of luminance and chromatic gradients to recover 3D shape information. Unfortunately shape-from-shading is a difficult
computational problem, since shading information is inherently ambiguous (an example being the classic bump/dimple stimuli, shown in Figure 2.3), and generally a continuous set of solutions exist for any given shading example, since there are multiple unknowns (Prados & Faugeras, 2005). To be able to find a solution traditional algorithms must reduce these unknowns, normally by assuming that surface reflectance at any particular point is a function of only the directly incident light (direct from the light source) and the direction of the surface normal. Thus constant albedo must be assumed (i.e. no colour or other changes in surface reflectance) along with lambertian (perfectly diffuse) surface reflection (Zhang, Tsai, Cryer, & Shah, 1999). Additionally, known lighting conditions are required, normally including direction (Pentland, 1982), although more recent work has provided algorithms that work with diffuse light sources (Langer & Zucker, 1994; Langer & Bulthoff, 1999).
Figure 2.3 An example of shading as a cue to three-dimensional shape. For most observers, the column of circles in the 1st and 3rd columns (from the left) appear concave, while those in the 2nd and 4th appear convex. The shading applied to all the circles is a simple linear luminance gradient and is identical other than orientation for all the circles. The change in perceived shape with shading orientation is linked to assumptions about light source position (see section 2.1.4.2.1 for details).

Some shape from shading algorithms have been developed that use more sophisticated illumination models that incorporating non-direct (i.e. mutual) illumination to recover shape (this is discussed further in section 2.1.4). Nayar, Ikeuchi, & Kanade (1991) showed how traditional shape-from-shading algorithms create erroneous surface shape when mutual illuminations are present, and that by an iterative process these erroneous shapes can be used to find the true shape (since the mutual illumination is related to the true object shape).
2.1.4 Perception of 3D Shape from Shading

2.1.4.1 Separating Shading from Reflectance Changes

Although shading provides 3D shape information to an observer because of the effects of surface orientation on luminance, this information is inherently ambiguous. Spatial variations in illumination and surface reflectance properties can also create differences in luminance. Understanding which variations in luminance are due to surface/object shape, and which are due to variations in reflectance and/or illumination is an important task for the visual system.

Ernst Mach gave what is thought to be the first known demonstration of the interactions of surface shading and 3D shape perception in the late nineteenth century. The ‘Mach Card’ he described consists of a paper card containing a single vertical central fold so that it is convex to an observer (see Figure 2.4). The card is illuminated from the side so that one side has a higher luminance than the other. Observers will usually demonstrate lightness constancy, judging both sides of the card to be made of the same material, despite the differing luminance. If an observer’s perception of the 3D shape is reversed, either by optically reversing the binocular disparity, or even simply by the observer attempting to convince themselves that the card is concave, the two sides are no longer seen as the same surface reflectance (with different illumination), but as surfaces with different lightness (lightness inconstancy).
More recent studies have also investigated the visual system’s capability to account for 3D shape when estimating surface reflectance properties. For example, Boyaci, Maloney, & Hersh (2003) used a complex scene with several objects and surfaces to provide potential information about light source position to the observer. They showed that observers could accurately estimate the orientation of surfaces and thus make consistent judgements of surface lightness by accounting for shading effects, despite significant luminance gradients on the surfaces. A similar result was found by Kitazaki, Kobiki, & Maloney (2008), who found that lightness was modulated by the 3D shape indicated by a number of cues. In another related study, Ripamonti et al. (2004) measured perceived colour of a real grey card as it was rotated with respect to the observer and light source. However, in this study the authors found that while observers could estimate the angle of the
card accurately, their estimates of the light source position contained errors. Bloj et al. (2004) suggested that these light source position errors were due to some of the change in brightness of the card being wrongly attributed to a change in surface reflectance.

2.1.4.1.1 Mutual Illumination and Surface Reflectance

A number of studies have shown that mutual illuminations provide a potentially useful source of information about the relationship between 3D shape and surface reflectance. In their 1984 paper, Gilchrist & Jacobsen describe two real three-dimensional rooms, one painted uniform white and one uniform black. The rooms were identical other than their surface albedo and illumination intensity. They found that even when the illumination in the black room was much higher than the white room, such that reflected luminance values were higher, the white room was correctly identified as white, and the black room estimated as middle grey (Gilchrist & Jacobsen, 1984). The authors suggested that the rooms were distinguishable via the indirect (mutual) illumination within the rooms and provided luminance profiles to support this. Later, Funt, Drew, & Ho (1991) showed that it is computationally possible, by examining the mutual illumination of two nearby surfaces, to recover the surface reflectances, the illumination spectrum and also some information about the geometry of the surfaces (expressed by a factor summarising the amount of mutual illumination). They developed an algorithm and demonstrated its ability to recover the scene information. A subsequent study by Ruppertsberg & Bloj (2007) took these ideas further, indicating that the visual system may have some similar capabilities to those
demonstrated by the algorithm of Funt et al. (1991), as previously suggested by Gilchrist & Jacobsen. Ruppertsberg and Bloj created two physically accurate computer renderings of a room, where the illumination spectra of one room matched the surface reflectance of the other and vice versa. Without mutual illumination, both scenes would be identical. However, the presence of mutual illumination results in differences between the two rooms: when there is only a single light ‘bounce’ (direct reflection) the multiplication of illumination and reflectance spectra are identical for both situations. However, when the light is subject to multiple reflections before reaching the observer’s eye, the resulting spectrum depends increasingly on the surface reflectance. These differences in the scenes due only to mutual illuminations were shown to be perceptually distinguishable with many combinations of surface reflectance and illumination functions.

The studies mentioned above showed that mutual illumination can make perceptually distinguishable differences to scenes, but a study by Bloj, Kersten, & Hurlbert (1999) took this further, providing evidence that the visual system is capable of accounting for mutual illuminations when estimating object shape and reflectance properties. They asked observers to match the colour of the white side of a ‘chromatic Mach card’ (a colour version of the Mach card – see Figure 2.4 – with one side white and the other coloured) to a range of colour chips under the same illumination as the card. The card was viewed either with or without a pseudoscope that inverted the binocular disparity and switched perception of the card from concave to convex. They found that the matched colour was affected by the shape perception of the
card. Observers correctly accounted for the change in colour of the white side of the card due to light reflected from the coloured side when the card was viewed with the correct disparity information (concave), but matched a significantly more coloured chip when the disparity information was reversed (and the card appeared convex). This result indicates that the visual system is capable of accounting for mutual illumination in three-dimensional scenes and this knowledge is incorporated into colour perception at an early stage. This result provides an important indication that the visual system may be able to use chromatic mutual illuminations as a cue to 3D shape, since shape and colour perception are shown to be linked.

### 2.1.4.1.2 Colour as a Means of Disentangling Shape and Reflectance

In many real world situations, chromatic and luminance shading gradients are correlated (such as gradients that arise from mutual illumination, or changes in surface chromaticity with changing viewing angle that coincide with shading). However, a number of studies, using both computer renderings of three-dimensional scenes, as well as more simple stimuli, have shown that the human visual system can discriminate between both luminance and chromatic gradients and also between gradients of different intensities or spatial frequency content (Garcia-Suarez, Ruppertsberg, Hurlbert, & Bloj, 2007; Garcia-Suarez, Ruppertsberg, & Bloj, 2008; Ruppertsberg et al., 2008; Ruppertsberg et al., 2007). It is possible that this is because there are important situations where chromatic and luminance gradients are not correlated: Variations in surface reflectance (either chromatic or achromatic) such as those that occur on patterned surfaces,
such as flowers or many man-made objects, give rise to un-correlated chromatic and luminance gradients. Three-dimensional scene geometries leading to attached and cast shadows also give rise to this effect. If chromatic and luminance gradients are processed separately, in these situations the visual system could use chromatic and luminance changes independently to determine where gradients are due to surface shape and scene geometry and where they are due to changes in surface reflectance. Additionally, Hansen & Gegenfurtner (2009) showed that in natural scenes colour and luminance edges are statistically independent and suggest that separate processing of chromatic information may be useful to segment objects in scenes.

Kingdom (2003) showed a simple effect of the independent operation of chromatic and luminance gradient detection by overlaying a sinusoidal luminance grating and a sinusoidal chromatic grating such that they are out of phase (not spatially aligned, see figure 2.5). In this ‘colour-shading effect’, the chromatic pattern is typically perceived as a striped surface reflectance and the luminance variation as due to changing depth across the surface (see section 2.1.3 - Three Dimensional Shape-from-Shading). This perception of three-dimensional structure does not occur when either the chromatic or luminance gratings are in isolation, or when they are aligned. Kingdom surmises that discrepancy between changing luminance and colour information provides a means for the identification of which image properties are due to surface reflectance and which are due to 3D shape. i.e. that the visual system assumes that luminance changes that do not correlate with
chromatic changes must be due to shading of shadows produced by 3D shape variations.

Figure 2.5 Kingdom plaid stimuli showing increase perceived depth when luminance and chromatic gradients are combined (Kingdom, 2003). (a) Luminance sinusoid only. (b) Chromatic sinusoid only. (c) Out of phase luminance and chromatic sinusoids.

Later, Kingdom, Beauce, & Hunter (2004) used a psychophysical study to similarly show that consistent colour information assisted observers in the identification of shadows, but also extended this to show that random colour contrasts at shadow boundaries made identification more difficult. Kingdom, Rangwala, & Hammamji (2005) looked at how sensitive the ‘colour-shading effect’ was to both colour contrast and the colours involved. They found the effect was strongly dependent of colour contrast, but not on chromaticity or which colour opponent system, L-M or S-(L+M) was stimulated (for more details on colour opponent theory see Dacey & B. (1994), A. M. Derrington & Lennie (1984), DeValois & DeValois (1993), Hendry & Yoshioka (1994) and Hering (1964)).
Summary

The results of studies described in this section suggest that illumination gradients, from both shading and mutual illumination, are used by the visual system when judging 3D shape and surface reflectance properties. They also indicate that the visual system is also, to some extent, capable of separating shape and reflectance properties, so as to account for each when judging the other. However, our knowledge of how these processes interact, and under what circumstances the visual system is capable of correctly assigning variations in luminance and colour, particularly from mutual illuminations, to shape or reflectance properties is limited. Most of the studies that have investigated these capabilities have been either theoretical/computational or unable to describe the precise way in which the complex gradients that arise from realistic colour scenes affect 3D shape perception. Bloj & Hurlbert (2000) showed that the addition of luminance and colour gradients make a contour-defined shape appear to have more depth, but a clear understanding of how and under what circumstances complex, realistic, illumination gradients that include colour effects due to mutual illuminations influence shape perception is still to be determined. It is this question that this work attempts to, at least in part, answer.

2.1.4.2 Mechanisms by which Humans Derive Shape from Illumination Cues

Despite the limitations of shape-from-shading algorithms noted in section 2.3.1, it is possible that they can provide us with clues to how human perception of three-dimensional shape is achieved. Perhaps the visual
system has mechanisms that work, or can be modelled, in a similar way to shape-from-shading algorithms?

Previous studies that have investigated the perception of three-dimensional solid shape have found that the efficacy of illumination related shape cues is seemingly dependent on the specific stimuli and task. For example, in their 1983 study Todd and Mingolla found that shininess enhanced perception of surface curvature from shading (Todd & Mingolla, 1983). A similar study by Liu & Todd (2004) also found that observer performance was much better when highlights and cast shadows were included. However in another perceptual study, in which observers instead judged the orientation of surfaces in shaded images, rather than curvature, Mingolla & Todd (1986) found that highlights (and cast shadows) had no effect of the performance of observers in judging correct slant or tilt. Nefs et al. (2006) later confirmed this result, finding no differences in observers’ perceptions of surface orientation between glossy and matte versions of computer rendered three-dimensional objects.

Mingolla & Todd propose that the results of their 1986 study suggest that the key assumptions made by shape-from-shading algorithms are not valid when considering the human visual system. This is an important finding, as the ideas behind these algorithms have been proposed as the basis for models of human perception of solid shape (e.g. Pentland (1988)). Specifically, Mingolla & Todd point out that the initial assumption of Lambertian reflectance, the requirement for a known light source direction and the
hypothesis that surface orientation detection occurs locally are not supported by their data. Further to this, some authors, such as Erens, Kappers, & Koenderink (1993) suggest that global shading (i.e. that related to illuminant direction) has little influence on estimation of local surface shape. This again indicates that knowledge of illuminant direction is not critical for human perception of shape. Mingolla & Todd suggest that alternative models of human perception of solid shape such as those offered by Koenderink and van Doorn may be more likely to accurately describe the human visual system, since they do not rely on the seemingly invalid assumptions outlined above (e.g. Koenderink & van Doorn (1980), Koenderink & van Doorn (1982)). These models derive information about global surface shape from local surface information. The authors show that theoretically local maxima, minima and saddle points in the image intensity can provide information about the local object shape. For example, intensity maxima occur at positions in which the surface normal is aligned with the illuminant direction, and minima only occur on the boundary curves between hyperbolic and elliptic regions. This idea is consistent with the work of Fleming et al. (2004) who showed that local properties of specular highlights can provide 3D shape information.

Further confirmation of the human visual system’s apparent ability to perceive depth and surface shape from illumination cues without knowledge of a specific illumination direction has been provided by Langer & Bulthoff (1999) who performed a psychophysical experiment assessing observers’ ability to judge surface shape under both point source and diffuse illumination
conditions. They found that observers’ performance in the diffuse condition was better than that predicted by models that require knowledge of a specific illumination direction, or what would be expected from just the fall off of intensity with distance (‘dark means deep’). In the point source condition however, observer’s performance was only good with the light direction from above, but poor when it was from below, indicating some assumptions about illumination direction (specifically of elevated light source position - see section 2.1.4.2.1 below for more detail on light source position). Results from both conditions also correlated with the results of a ‘dark means deep’ model. The authors conclude that the human visual system is not limited to only one mechanism for perceiving three-dimensional shape-from-shading.

There is one further mechanism related to shading through which the visual system has been shown to obtain information about 3D shape. Disparities between the patterns of shading at each eye provide an additional cue to shape. This is because relative light source, object and viewing positions are different for each eye, so small differences in object shading are present between the left and right eye images. Studies have shown that the addition of disparity to shading can increase perceived depth (Bülthoff & Mallot, 1988) and surface curvature (Mingolla & Todd, 1986) of shaded 3D stimuli, while others have shown that shape estimates from shading are more reliable when the shading includes binocular disparities (Vuong, Domini, & Caudek, 2006).

2.1.4.2.1 Illumination Direction and Shape-from-Shading
Knowledge of the direction of a light source is key to the recovery of three-dimensional geometry by the majority of computer algorithms. As noted in the previous section, a number of studies have suggested that the same is not true for human perception of surface shape, and that knowledge of the light source direction is not required. However, studies such as Nefs, Koenderink, & Kappers (2005) and Nefs et al. (2006) that have measured observers’ estimates of the surface shape and orientation of shaded stimuli, have shown that systematic changes occur in observers’ estimations when the light source position is altered. These studies do seem to suggest that light source position has some influence on shape perception, although not necessarily though means of a direct estimation of illumination direction. When Boyaci et al. (2003) studied observers’ estimates of the angle of a tilted surface and its albedo, they found that although observer estimates of surface angle were very good, they still made consistent (and idiosyncratic) errors when estimating the surface albedo. The authors concluded that this was due to observer errors in estimating and discounting the effects of light source position. A similar study by Bloj et al. (2004) found analogous results and the authors developed a model describing the predicted lightness constancy of an observer based on their estimates of parameters describing the geometry of the 3D scene, including illumination direction, by fitting model parameters to observer data.

There is evidence that the influence of light source position on perception of depth is largely due to prior knowledge or preference for light source positions. It has been known for some time that human observers exhibit a
bias towards assuming light is coming from a particular direction (Brewster, 1826), typically from above, and this is reflected in historical preferences for lighting in artwork. In an experiment designed to test this ‘light from above assumption’, Sun & Perona (1998) used bump/dimple (or bubble) stimuli (see Figure 2.3) that have concave-convex ambiguity depending on the light source direction, and recorded how long it took observers to find a differently orientated bump/dimple among many distracters. They found that the lighting direction that gave the fastest response times was from above, but also around 15 degrees to the left of the observer on average. Additionally they found a maximum performance for right-handed observers was achieved with lighting further to the left than for left-handed observers. The authors suggest that this ‘prior’ (an assumption based on prior knowledge) for light direction, which helps the visual system extract shape information from ambiguous stimuli, is modified by observers' preferred light source orientation. Using a shape discrimination task for similar ambiguous shaded stimuli, Mamassian & Goutcher (2001) found comparable results, but showed a larger bias of 26 degrees to the left and did not find a difference between left and right handed observers. Adams, Graf, & Ernst (2004) also performed a similar experiment, again confirming a bias towards light sources above and to the left of the observer. However, as well as measuring the 'light from above prior', the authors also introduced haptic feedback and showed that it was possible to modify the light direction bias through training. More recently the ability of observers to modify their use of the light from above prior differently in multiple illumination contexts has also been demonstrated (Kerrigan & Adams, 2013). These results suggest that the visual system is
adaptable in its construction and use of prior knowledge, modifying these biases in response to environmental conditions. Other biases, or priors, are thought to influence depth and shape perception from shading in adult humans, such as a bias towards perceiving convex shape (Liu & Todd, 2004). A more detailed discussion of priors is given in section 3.1.2.

The work described in this section indicates that although knowledge of the light source direction may not be required by the human visual system to derive object shape, this does not necessarily mean that it has no influence. In fact, it seems that knowledge of light source position has an important role to play in the disambiguation of depth information from shading, but this knowledge is apparently largely collected in the form of a distribution of likely directions (i.e. a prior), rather than direct estimates for every scene.

2.1.4.2.2 Mutual Illumination and Object Shape/Position

As well as providing information about surface reflectance, the theoretical work by Funt et al. (1991) suggests that mutual illumination can also give cues to object shape. Forsyth & Zisserman (1990) showed computationally that, with the presence of mutual illumination, discontinuities in surface radiance only occur at actual surface discontinuities, at the edge of cast shadows and at abrupt changes in surface reflectance. Because of this, these radiance discontinuities are reliable cues to object shape and three-dimensional structure. Similarly, Nayar et al. (1991) developed a shape-from-shading algorithm (see section 2.1.3) that incorporates mutual illumination, showing that because the mutual illumination is related to three-dimensional
shape, it can be used to recover both surface reflectance (as per Funt et al., 1991) and object shape. This work lays the theoretical groundwork for mutual illuminations as a cue to 3D shape, but little work has directly investigated how well the human visual system can use this source of information. While some of the large number of perceptual studies into shape-from-shading (for example Khang, Koenderink, & Kappers (2007), Kleffner & Ramachandran (1992), Liu & Todd (2004), Nefs (2008), Nefs et al. (2006), van Doorn, Koenderink, & Wagemans (2011) and Wagemans, van Doorn, & Koenderink (2010)) have include the effects of inter-reflections, nearly all such work has used achromatic shading, ignoring the chromatic aspect of mutual illuminations.

The importance of mutual illumination in the perception of object and scene properties other than shape has been demonstrated by a few recent studies: Madison, Thompson, Kersten, Shirley, & Smits (2001) investigated the effects of mutual illuminations on the perception of three-dimensional scene layout (i.e. the relative position of objects within a scene). They created computer renderings of a three-dimensional scene, rendering images with “1) no shadow plus no interreflection, 2) shadow only, 3) interreflection only, and 4) shadow plus interreflection”. After assessing observer’s perception of the scene structure within the different renderings, results showed that both shadows (supporting the work of Mamassian et al. (1998) - see section 2.1.1) and inter-reflection provided cues to object layout, with the combination of both producing the most accurate perception for observers. In addition, the results of the experiments of Bloj et al. (1999), which showed that perception
of shape influences the perception of object colour (see section 2.1.4.1.1 for details) indicate that the visual system is capable of accounting for mutual illumination in three-dimensional scenes. This result provides an important indication that the visual system may be able to use chromatic mutual illuminations as a cue to 3D shape, since shape and colour perception appear to be linked.

2.1.5 Shading and Shape Constancy

As described in section 2.1.4.1 - Separating Shading from Reflectance Changes, the human visual system is able to determine to some extent which changes in image properties such as luminance and chromaticity are due to surface reflectance changes and which are due to the geometry of both object and light source positions (e.g. Kraft, Maloney, & Brainard (2002), Ripamonti et al. (2004)). This ability enables the visual system to keep, to some extent, a constant perception of 3D shape under widely varying lighting and surface reflectance conditions, object orientations and distances. This is known as shape constancy.

A number of studies have been carried out with the aim of both measuring and modelling human shape constancy, showing consistent ability in humans even from a very young age (see Walsh & Kulikowski (1998) for an in depth review of this research). Although human shape constancy is generally quite good, Khang et al. (2007) have shown shape constancy can be relatively poor in certain situations. They found that features that arise from surface material type or lighting parameters (e.g. highlights) gave rise to inaccurate
shape estimations when observers were asked to estimate surface orientations and object cross sections. This finding holds a similar importance to the results of Ripamonti et al. (2004) (see section 2.1.4.1), in that it suggests that the interactions of shape illumination and reflectance may not be fully accounted for by the visual system. However, Khang’s study used achromatic conditions that do not fully represent the real world, and studies such as those by Kingdom (2003) and Hansen & Gegenfurtner (2009) suggest that luminance and chromatic variations are treated differently by the visual system when it comes to determining if such gradients are due to surface reflectance or shape (see section 2.1.4.1).

Given this possible different treatment of luminance and chromatic gradients when determining object shape, it seems likely that if the perception of 3D shape is to be thoroughly investigated, the specific effects of colour must be considered and an achromatic treatment will not be sufficient. For this reason, the experiments described in the later chapters of this thesis investigate the shape cues provided by realistic colour shading.

2.2 Shape from Contour / Perspective

In a number of the experiments described in this thesis, colour shading cues to 3D shape are studied in combination with a perspective / object outline cue. Therefore, I will provide more detail in this section regarding 3D shape perception from this type of cue.

Information about 3D object shape is provided in the contours present on and at the edges of surfaces. Contours arise from the outline of objects at
occluding boundaries, but also from changes in surface reflectance and illumination (shading and shadows) related to object shape. A powerful demonstration of the importance of contours for perception of three-dimensional object shape is provided by line drawings. Such drawings using only lines to indicate contours (both surface and outline) are well known to provide robust perception of object shape and scene geometry (Kennedy, 1974; Mamassian & Landy, 1998). For example, see Figure 2.6. A strong impression of a three-dimensional cube is perceived, despite the very limited amount of information available.

![Figure 2.6](image)

**Figure 2.6** Line drawing of a 3-D cube, demonstrating shape cues from perspective and outline contour.

The effect of contours on shape perception can also be seen in the Orbison Illusion (Orbison, 1939). Figure 2.7 shows this illusion. The radiating blue lines provide erroneous contours that distort the perception of both the red square and the bounding rectangle of the image, making both appear rotated from the plane of the page.
Figure 2.7 Orbison Illusion (Orbison, 1939), demonstrating depth from contour. For most observers, both the red square and the rectangular background appear rotated from the plane of the page.

While contours and perspective can provide a robust impression of 3D shape, it has been shown that perception is typically biased. In an experiment designed to investigate biases in the perceived shape of objects defined by simple line drawings, Mamassian & Landy (1998) found that observers were biased towards perceiving convex surfaces and surface orientation consistent with an object viewed from above. A number of studies have also suggested that observers are biased towards seeing lower than veridical surface slant for objects defined by outline (e.g. Andersen et al. (19980, Freeman (1966), Saunders & Backus (2006), Smith (1967), Todd, Thaler, & Dijkstra (2005) and van Ee, Adams, & Mamassian (2003)).
Saunders & Backus (2007) studied the depth perception of human observers in response to line drawings of slanted rectangles. Their aim was to determine, in more detail, which cues in the outline provided the percept of three-dimensionality. Saunders and Backus suggested that two assumptions could be used to provide the depth percept: “(1) converging lines in an image are parallel in the world, and (2) skewed angles in an image are orthogonal in the world”. By using stimuli that were scaled and then rotated, Saunders and Backus were able to provide conflicting cues from the two assumptions outlined above. The results showed that, for non-conflicting cues, observers mainly relied upon the parallelism assumption (converging lines are parallel) when assessing slant. However, when the parallelism and orthogonality (skewed angles are orthogonal) assumptions conflicted, performance was consistent with the use of the orthogonality assumption. It seems that, as with shape-from-shading, multiple methods may be employed by the visual system to generate three-dimensional perception from contours. If multiple mechanisms are used to determine the structure of the three-dimensional world, both from the different classes of cues (shading, contour, disparity etc.) and also different aspects of each class (see Saunders & Backus (2007) for contour and Langer & Bulthoff (1999) for shading), then some method of combining information from different cues must exist. A discussion of how this combination of depth cues might occur is given in the next chapter.
Chapter 3 – 3D Shape Cue Integration

Under the majority of circumstances only a single percept of object shape is apparent to an observer at any one time. It therefore seems clear that if the visual system gathers shape information from multiple cues, mechanisms must exist to integrate the information from these sources. In addition, theoretically, if shape information from multiple cues is combined optimally, it should also lead to more precise estimates than when using any cue alone (see Cochran (1937) and Ernst & Banks (2002)). This has been show in behavioural experiments to occur in humans for some combinations of cues (e.g. Ernst & Banks (2002), Hillis, Watt, Landy, & Banks (2004), Lovell, Bloj, & Harris (2012) and Oruc, Maloney, & Landy (2003)). There is also evidence from functional magnetic resonance imaging studies that specific areas of the brain are active in combining depth information from multiple sources (Ban, Preston, Meeson, & Welchman, 2012). The literature surrounding perceptual cue integration is extensive (for example, Alais & Burr (2004), Jacobs (2002), Johnston, Cumming, & Landy (1994), Johnston, Cumming, & Parker (1993), Landy, Maloney, Johnston, & Young (1995) and Young, Landy, & Maloney (1993)), and much of this literature attempts to model the processes by which cue integration might occur. The following section details some of the types of models that are often used to explain experimental results.

3.1 Models of depth / 3D shape cue integration

One of the most frequently used categories of cue combination / integration model is known as the “Weak Fusion” type (Clark & Yuille, 1990; Landy et
al., 1995). In this kind of model, depth/shape information from each cue is processed independently and estimates from each cue then combined linearly. Clark & Yuille (1990) also describe an alternative class of models they call “Strong Fusion”, where cue processing is not necessarily modular, and cue information may interact resulting in non-linear combination. Strong fusion models are not as frequently used as modular, weak fusion, types. Because the interactions of cue information in strong fusion models can take many forms, meaningful predictions can be difficult to make, as they may not be falsifiable. However, some authors have used this general approach to model cue combination (e.g. Nakayama & Shimojo (1992)), and others have found evidence for strong fusion. For example, Rosas, Wichmann, & Wagemans (2007) studied the integration of depth cues from contour (via texture) and object motion, using a slant estimation task. They investigated if the weighting of the cues could be altered by changing the texture of their stimuli, but found no reliable pattern of linear cue combination across their observers. They suggest that this may be due to a strong coupling of the cues, either intrinsic to the nature of the information, or within the visual system.

### 3.1.1 Weighted Linear Combination

One of the most popular weak fusion models is the reliability-weighted linear combination model, which has been used by a number of authors to successfully predict the combination of depth or 3D shape information from a range of cues (e.g. Ernst & Banks (2002), Hillis et al. (2004), Lovell et al. (2012) and Oruc et al. (2003)).
In this model, the estimates that are made from each cue are combined linearly in proportion to the reliability of relevant cue. Reliability is defined as the inverse of the variance of observer estimates made using the cue. Thus, in the case of integrating two cues, if Cue 1 has half the variance of Cue 2, then it is given twice the weight and the combined cue estimate is given by \((2/3 \times \text{Cue 1 estimate}) + (1/3 \times \text{Cue2 estimate})\).

The advantage of this particular weighting strategy is that, assuming normally distributed estimates and uncorrelated variance, it provides the minimum variance estimate (Cochran, 1937). This is the statistically optimal combination that results in the most reliable combined percept and is a maximum likelihood estimator (MLE) – the estimate that maximises the probability of the current data (Ernst & Banks, 2002). This reliability-weighted linear combination model is therefore often referred to as the MLE model.

Strict weak fusion models, that are completely modular, have a number of drawbacks. They cannot explain how the visual system ‘averages’ information that is of very different types from different cues (most cue information is relative and does not provide absolute estimates of depth), or how estimates of cue reliability are calculated. Additionally, they cannot provide means of robust cue combination behaviour, where cues that suggest outlying values are down weighted or ignored (for an example of this behaviour, see Bülthoff & Mallot (1988)). These issues led Landy et al. (1995) to suggest a “Modified Weak Fusion” (MWF) model. In MWF cues are combined linearly, weighted by their reliability estimates, but prior to this, each cue is “promoted” to a common representation (i.e. an absolute depth
value) by the filling of ‘missing’ parameters from the information present in other cues. Cue combination weightings are determined by cue reliability (as in the MLE model), but this reliability information is explicitly provided by localised ancillary cues. For example, vergence provides information on absolute viewing distance that can be used to constrain the reliability of depth estimates from binocular disparity (since disparity information is more reliable for closer viewing distances). In this model, cues therefore have a limited form of interaction, allowing improved estimates to be made in some cases (for example outlying cues), and a more complete description of how cue combination may be achieved. Vuong et al. (2006) propose a similar method of constrained cue interaction. In their experiment the authors found that shading information appeared to constrain disparity estimates in areas where disparity could not be directly measured. The authors however point out that their theory differs from MWF in that absolute depth values are not calculated for each cue, but simply that the similarity of likely shape maps from individual cues can be used to improve performance.

Landy et al. (1995) also describe a ‘perturbation analysis’ method (see also Young et al. (1993)) to measure the weights given to different depth cues by the visual system. Because the parameters of the system may be affected by the strength (i.e. reliability) of the inputs (depth/shape cues), which are likely to be different for different 3D shapes/depths, small changes (perturbations) in individual cues in isolation must be analysed (with the assumption of local linearity) to determine the underlying mechanisms. In their summary Landy et al. (1995) also point out that care must be taken when performing
psychophysical experiments (in the general case but also specifically for depth cue combination studies) to ensure that all aspects of stimuli are as veridical as possible, since processing may be affected by inconsistent or redundant information by means of information fusion mechanisms. They suggest that in cases where stimuli contain significant conflicts, the assumption of linear fusion of information may not hold and the visual system may be performing in a range where it is not optimised.

3.1.2 Bayesian Cue Integration

Another (although related) commonly used type of cue integration model is the Bayesian variety. Here, cue integration is based on Bayesian statistics, and incorporates the effects of prior knowledge. Such models are based on the optimal combination of statistical shape/depth information provided by image cues and the known probability of possible shapes or depths present in the world (prior probability distributions, or simply ‘priors’). It has been suggested that Bayesian models are ideally suited to explaining how the visual system generates reliable object perception, despite typically highly complex and ambiguous images (Kersten, Mamassian, & Yuille, 2004), and this type of model has been used successfully to explain the results of a number of experiments that contain ambiguous stimuli (Adams & Mamassian, 2004; Kersten et al., 2004; Knill, 2007b; Mamassian, Landy, & Maloney, 2002; Yuille & Bülthoff, 1996). In addition, some authors have suggested that in situations of large cue conflicts, a simple linear fusion of cues may not be suitably robust and Bayesian models may provide a more accurate representation of observed robust integration behaviour. For
example, if two cues are similarly reliable and provide conflicting information (and other cues are minimally weighted), how can we best decide what is the most likely shape or depth? To solve this problem, Knill (2007b) suggests a Bayesian cue integration model and tested its predictions against the performance of human observers when judging the slant of dot filled ellipses. In this situation, a strong cue conflict can be present between binocular disparity and contour, since the visual system appears to be strongly biased to perceive ellipses as slanted circles. Knill found that the weight given to the contour cue decreased smoothly with increasing cue conflict, but was never ignored entirely. Knill suggests that the visual system may integrate the slant information in a non-linear way, by incorporating knowledge of a prior distribution of object shapes, with objects most likely to be regular in shape, but the smaller possibility of more random shapes still occurring.

3.1.2.1 Weighting of Prior Constraints

Models of cue integration that include the use of biases within the visual system, such as a ‘light from above’ prior (Adams et al., 2004; Mamassian & Goutcher, 2001; Sun & Perona, 1998) or an ‘elevated viewing’ prior (Mamassian & Landy, 1998) must combine this prior information with the depth cue information somehow. Within the Bayesian framework, this process occurs by the multiplication of the ‘prior distribution’ (the known probability of any particular 3D shape/depth) with the ‘likelihood function’ (probability of the image, as a function of the 3D shape/depth) and the two components are entirely independent (see Chapter 8 for more details). Prior constraints are usually assumed to be ‘in-built’ or gathered over long periods.
of time, and thus the prior distribution should not be dependent on the current stimulus. The ‘weight’ of a prior is dependent on how peaked the prior probability distribution and the cue likelihood functions are, relative to each other. A very peaked prior distribution, where a narrow range of depths/shapes are very likely, will affect the predicted percept to a greater extent than a wide distribution where all shapes are similarly likely.

However, in their investigation of the visual system’s use of priors, Mamassian & Landy (2001) found evidence that in some cases priors may not be independent of the stimulus image. They performed a psychophysical experiment using stimuli that contained shading and contour cues to 3D shape, which should be affected by the ‘light from above’ and ‘elevated viewing’ priors. By rotating the stimuli they could place the two priors either in agreement or conflict, and the relative reliability of each cue was dependent on the angle of rotation. They found that the visual system seemed to treat the prior constraints in the same way as the shape cues. The weighting of each prior constraint was dependent on the reliability of the corresponding cue. The authors model these results using an approach based on Bayesian theory, but allow the prior distributions to depend on the reliability of their respective cues. This creates a method of ‘weighting’ the prior information dependent on the stimulus.

3.1.3 What type of cue combination model is best?

The body of literature investigating depth cue integration has shown that the best performing type of model often depends on the exact situation being
investigated. For example, previous work using ambiguous depth cues has shown that Bayesian models can be used to successfully describe observer perception, where the combination of cue information appears to help to disambiguate perception (Adams & Mamassian, 2004). A separate study investigating similar types of depth cues found a ‘maximum likelihood estimate’ cue combination model, that does not account for prior constraints could successfully describe observer perception, since their stimuli did not contain any ambiguities (Hillis et al., 2004). It should be noted however, that in the absence of any prior assumptions or constraints, a Bayesian model is equivalent to a maximum likelihood estimate (reliability-weighted linear integration) if likelihood functions and priors are normally distributed and a maximum a posteriori (MAP) decision rule is used (see Chapter 8 for details of these terms).

It is generally the case for monocular depth cues that, because they are inherently ambiguous, information can only be derived from them with the addition of a prior assumption about, or constraint on, the types of objects and scenes typically found in natural viewing conditions. Thus it can be argued that for the majority of cases involving monocular cues, a Bayesian type model, that incorporates such priors, is the most appropriate. When binocular depth cues are also present, both strong (Bülthoff & Mallot, 1988) and weak fusion (Hillis et al., 2004) models, that do not take account of any prior constraints, have been shown to accurately predict perception. However, the interactions apparent when cues were combined in Bülthoff & Mallot's experiment may be an example of the type of robust behaviour that
could be explained by the modified weak fusion of Landy et al. (1995) - see section 3.1.1. If the integration of depth / 3D shape cues does not always occur by simple averaging, a wide range of non-linear combination functions could occur. Are all types of depth / 3D shape cue integrated in the same way, or do different mechanisms exist to integrate different combinations of cues into a single percept (or at least is different cue integration behaviour evident for different cue types)? In the modified weak fusion model of Landy et al. (1995), interaction between information from different cues is restricted. It is limited to the ‘promotion’ of cues to absolute depth and the estimation of reliability via ancillary information, but this still leaves scope for considerable variation in the precise way particular cues are combined, depending on the types of cues available. Bayesian models of cue integration allow combination on the basis of statistical probability, allowing for factors external to the current image information to influence the final percept. Since this prior information only influences estimates using the related cue type, and priors related to different cues are likely to be disparate in nature, this again leads to the potential for very different perceptual behaviour when different types of cues are combined.

With these ideas in mind, there have been a number of studies in recent years that have investigated the interactions of particular pairs of depth or 3D shape cues. Examples of these types of studies, and the contrasting cue integration behaviour observed when using different stimuli and cue types follow:
Doorschot, Kappers, & Koenderink (2001) conducted an experiment to examine how binocular disparity and shading combine to provide object shape perception. They took stereo photographs of real three-dimensional objects, at three stereo bases and under three different lighting angles. The photographs were presented, using a mirror stereoscope, to observers who were asked to indicate the surface attitude for a large number of points on the object surface. Despite the fact that this was not a perturbation analysis study (see section 3.1.1) and cue adjustments were large, principal component analysis of the results found that the two depth cues (disparity and shading) were independent and combined in a linear fashion, in agreement with a weak fusion model.

Using plaid stimuli similar to Kingdom (2003) (see Figure 2.5), Schofield, Rock, Georgeson, & Yates (2006) studied the interaction of texture and shading depth cues. Their stimuli consisted of a plaid of two mixed cues, one containing shading and texture in phase, the other shading and texture out of phase. Using a haptic matching system, observers estimated depth of the perceived corrugated surface. Schofield et al. (2006) found that when the two cues were aligned, significant depth was perceived. However, when the cues were out of phase, the texture cue was used as information indicating that the shading was due to changes in surface reflectance and the stimuli was perceived as flat. This result is similar to that of Kingdom (2003) who showed chromatic changes could influence the weighting of luminance shading depth cues.
Adams & Mamassian (2004) investigated the effects of a binocular disparity cue on an ambiguous texture cue. They created textured stimuli, which could be interpreted as either concave or convex. Stimuli were presented either monocularly, or with a binocular disparity cue that could indicate either a concave or convex shape. In the monocular case they found a bias towards convex interpretation of the texture stimuli. In the binocular case they found that rather than combining in a simple linear fashion, the cues interacted. When binocular disparity indicated a concave shape, the ambiguous texture cue (which was perceived as convex when viewed monocularly) increased the perceived concavity when compared to a ‘flat’ texture pattern, rather than decreased the concave perception as would be expected by a linear addition. In this case, the binocular disparity cue seems to disambiguate the texture cue, such that they are then combined in a complimentary fashion. Adams & Mamassian (2004) successfully modelled these results using a Bayesian approach with a prior for convex texture interpretation.

3.2 Cue conflicts

In section 3.1.2, I discussed work by Knill (2007b) where he proposes that a Bayesian model is the most sensible approach to situations where large cue conflicts occur. This is because, in natural viewing, cue conflicts nearly always arise due to prior constraints or assumptions made by the visual system; for example, in Knill’s experiment, a contour depth cue in the stimuli is influenced by a prior assumption that ellipses are likely to be tilted circles – an assumption which can conflict with the information provided by visual cues, in this case binocular disparity (see section 3.1.2 for details).
A similar Bayesian approach to that used by Knill was previously used by van Ee et al. (2003) to model perception when multiple visual cues conflict. In this experiment the authors presented wire frame stimuli that could have conflicting surface slant specified by binocular disparity and perspective cues. The authors found that observers' perception of slant was bi-stable, and observers were able to perceive both the surface slants indicated by the binocular disparity cue and the perspective cue separately, and switch between these perceptions (see also van Ee, van Dam, & Erkelens (2002)). This was described successfully by the authors Bayesian model, which included the integration of both strong and weak rectangularity prior assumptions affecting the perspective cue. However it seems likely that the bi-stable perception of these stimuli is apparent and measurable because the cue conflicts used were often very large. Under more moderate cue-conflict conditions, bi-stable perception is not usually observed. Further evidence that shape perception in the presence of cue conflict is potentially complicated and variable dependent on the stimuli and available cues is provided by Bloj & Hurlbert (2000). In this study the authors examined 3D shape perception using luminance and colour gradient cues, together with conflicting contour (outline) cues. They used computer-simulated stimuli that replicated flat cards and cards with a central fold in both convex ‘corner’ and concave ‘roof’ configurations, with varying illumination in order to generate a range of gradient cues (such as mutual illumination within the convex folded card). The stimuli were bounded by outlines that were either congruent with the corner, roof or flat cases. The results indicated that the addition of luminance and colour gradients made the stimuli appear less flat, even if the
gradient cue conflicted with that from the outline. In other words, the addition of a lower weighted, but conflicting, cue (gradient) appeared to increase the perception of depth provided by the stronger weighted cue (outline).

When cue conflicts are small, the situation appears to be comparatively simple. Hillis et al. (2004) investigated interactions of binocular disparity and texture depth cues, again including stimuli with conflicting cues, but in this case using only small conflicts. By using a two interval forced choice paradigm (see Chapter 5, section 5.1.1.5) together with the cue perturbation technique described in Landy et al. (1995), they measured slant discrimination when cues were presented individually and in combination (with small conflicts between cues). The resulting discrimination thresholds provided reliability estimates for a linear cue combination (MLE) model, enabling the cue weightings to be calculated. This method relies on the assumption that conflicting cues are combined linearly to create a single percept. The authors showed that cues seem to be weighted so as to maximise the precision of combined cue estimates (i.e. minimum variance, see section 3.1.1). The success of this modelling suggests that the assumption of linear cue combination, when conflicts are small, is valid, at least for the cues under investigation in this study.

In Chapters 2 and 3 I have provided a review of the scientific literature concerned with human perception of three-dimensional shape, particularly when using cues based on illumination phenomena, and how the visual system might go about combining shape information from disparate sources.
I have discussed some of the areas where our understanding is limited, and noted how this provides the motivation for the experiments described later in this thesis. In the following two chapters I will describe some general techniques that form the psychophysical framework used throughout these experiments, and the analysis of the results, together with details of the specific techniques and apparatus used.
Chapter 4 - Visual Psychophysics

The study of the relationship between physical stimuli and sensation/perception forms one of the oldest and most significant areas of the science of experimental psychology. This study of human response to physical stimuli is referred to as psychophysics, after the work of Fechner in 1860 (English translation: Fechner (1966)). A large range of psychophysical measurement techniques have been developed since Fechner’s work, and many models of human perception have been created based on the findings of psychophysical studies.

Psychophysical techniques are widely used in vision research to make quantitative assessments of the function of the visual system and to study and explain visual perception. The experiments undertaken during the work described in this thesis used psychophysical techniques to investigate shape perception from colour illumination cues. For this reason, I will outline some of the most widely used concepts and techniques in this chapter. Many of these techniques are subsequently referred to in the later chapters that describe my experiments and analysis.

4.1 Measurement of Detection and Discrimination Thresholds

Any visual stimulus has a threshold value for detection (from a background), or discrimination from a similar stimulus with a different intensity. When the value of a stimulus is below the detection threshold the stimulus is invisible. When the difference between two stimuli is less than the discrimination
threshold, the two stimuli are indistinguishable. The smallest difference between two stimuli required to make them distinguishable is often referred to as a Just Noticeable Difference (JND). A threshold is usually considered the level where 50% of stimuli are detected or discriminated successfully, although sometimes more strict criteria are used, such as a 75% threshold. Typically the change in observer performance from below to above threshold forms a sigmoid function, known as the psychometric function (see figure 4.1).

![Figure 4.1. Example psychometric function for detection.](image)

The width of the psychometric function is dependent on the JND. When smaller differences are noticeable, the function will be steeper. In the 1840s Ernst Heinrich Weber discovered that the size of JNDs for the perception of many physical stimuli was proportional to the absolute value of the starting stimulus. Gustav Fechner later went on to show that this is a result of the fact that there is a logarithmic relationship between physical stimulus intensity.
and perceptual effect (Fechner, 1860/1966). Thus this relationship is known as the Weber-Fechner law. The methods used to measure detection and discrimination thresholds are largely the same. Gescheider (1997) outlines a range of methods for measuring thresholds, many of which were originally described by Fechner (1860/1966):

### 4.1.1 Method of Adjustment

An experiment to measure a threshold by the method of adjustment requires the observer to increase the intensity of a stimulus such that it becomes distinguishable from the background or another stimulus, or decrease the intensity of a stimulus until it becomes indistinguishable. This method provides an easy way to find threshold values, but suffers from subjects anticipating the threshold or becoming used to the responding at a particular intensity. Thus this method is not considered particularly reliable.

### 4.1.2 Ascending and Descending Limits

The method of limits is similar to the method of adjustment, but the observer does not control the stimulus intensity, but instead is asked if the stimulus is visible/distinguishable repeatedly as the stimulus intensity is increased or decreased by the experimenter. This method suffers from the same issues of observer anticipation and habitualisation as the method of adjustment.
4.1.3 Staircase Methods

An improvement on the method of limits is to present a stimulus many times in a ‘staircase’ fashion such that the intensity of the stimulus repeatedly passes below and above the threshold. When the observer changes his or her response for ‘seen’ to ‘not seen’ or vice versa, the value is recorded and the direction of stimulus change is reversed (either from increasing to decreasing or vice versa). The mean value of the reversal points, or the ‘most frequently visited’ reversal point, determines the threshold value. The repeated measurements and the way the method ‘zeros in’ on the threshold provides increased accuracy but some of the same problems noted above can still occur. It is possible to randomize staircase step sizes to reduce learning from the observer. See Garcia-Perez (1998) for more details on staircase methods.

Adaptive staircase methods have also been developed, such as the QUEST procedure that uses a Bayesian estimate of the threshold to determine the level of the next stimulus, updating the estimate as the experiment proceeds (Watson & Pelli, 1983). This process is based on the fact that the psychometric function is invariant when expressed as a function of log intensity, due to the Weber-Fechner law (see section 4.1).

4.1.4 Method of Constant Stimuli

The method of constant stimuli is one of the first psychophysical techniques developed, originally described by Friedrich Hegelmaier in 1852 (see Laming & Laming (1992)). This method involves presenting stimuli with a range of
intensities to the observer in a random order. The intensities of the stimuli are chosen to cover the range of the threshold, including some below threshold and some above. In a detection threshold experiment, observers indicate if each stimulus is visible. After many repeated presentations, plotting the percentage of stimuli visible at each stimulus value against intensity, a psychometric function like the one shown in Figure 4.1 can be generated and a threshold value obtained. The randomization of stimuli intensity in this method reduces errors related to stimulus history, such as observer learning or adaptation.

4.1.5 Forced Choice Procedures

In a forced choice paradigm, the observer must either choose if a stimulus was present or not in each trial, or between two or more stimuli per trial, based on some pre-defined selection criteria. Perhaps the most common and useful type of forced choice experiment is known as the Two Alternative Forced Choice (2AFC; Bergmann, 1858; Fechner, 1860/1966). In this type of experiment, the observer is presented with two stimuli, separated either spatially or temporally (if stimuli are separated temporally, then the experiment is sometimes referred to as a Two Interval Forced Choice (2IFC)). This type of experiment is ideally suited to measuring discrimination thresholds using the method of constant stimuli: one presentation contains a fixed stimulus, the other a stimulus that varies in intensity trial to trial. Typically the observer must decide which presentation contains a stimulus with a higher or lower intensity. After a number of trials the resulting data is used to generate a psychometric function from which a threshold can be
determined. Stimulus intensity can also be varied using a staircase procedure (see section 4.1.3), with the direction being determined by the observer’s responses. Correct responses lead to reduced changes in stimulus intensity and vice versa, causing the procedure to ‘zero in’ on the threshold. Using a staircase technique has the advantage of minimising observer bias and any effects of stimulus history, much like the method of constant stimuli, but can provide a more efficient procedure.

In the case of two-alternative forced choice experiments the observer must make a response for each trial, choosing between only two alternatives. The chance score is therefore 50%, and this is known as the Point of Subjective Equality (PSE), where the stimuli in the two intervals are not discriminable. Thus the discrimination threshold is typically considered to be at 75% of correct responses (Gescheider, 1997), although other methods exist (e.g. Ulrich & Miller (2004)).

### 4.2 Supra-threshold and Other Experimental Procedures

In addition to the measurement of thresholds, other properties of physical stimuli may also be quantified. These properties are normally measured in what are known as supra-threshold experiments, where stimuli levels are well above threshold, and some other aspect of the observer’s percept or sensation is to be measured. An example of a psychophysical experiment that is not concerned with measuring stimulus thresholds is a type of experiment often used when assessing how the perception of colours is affected by adaptation, contrast or other mechanisms: a colour naming
experiment. In this type of experiment observers select the appropriate match for a particular colour presentation from a constrained set of named categories. Such experiments may also be used to investigate the linguistic effects of colour groupings, or individual differences. For example a colour may be considered ‘blue’ by one observer by ‘green’ by another (Berlin & Kay, 1969). The following sections describe some more common types of supra-threshold experiments.

4.2.1 Magnitude Estimation

During a magnitude estimation experiment, observers are asked to estimate the level of a visual percept either using numbers or by matching the percept with another, different, physical stimulus (Stevens, 1957). This method has an advantage over the measurement of discrimination thresholds, in that it can reveal non-linearities or biases in the perception of a stimulus. However, magnitude estimation can be prone to errors inherent in the task that observers are set in order to record the results, particularly if this involves a translation into a different type of image or sensory modality. These errors may be difficult to identify and quantify. A magnitude estimation task is used in the first experiment described in this thesis (Chapter 6).

4.2.2 Ranking and Scaling

Other supra-threshold tasks include ranking experiments, where an observer is asked to put stimuli in order of increasing/decreasing sensation or preference (see Marks (1974)). This type of experiment is often used when
subjective observer preference is the parameter to be measured. In other situations experimenters might ask observers to scale a stimulus with a value (normally from a pre-determined range) relative to either a reference stimulus or simply their own interpretation of the ‘average’ stimulus level. Scaling with reference to an observer’s internal average is known as the method of single stimuli and attempts to avoid any uncertainty about observer strategy when comparing stimuli by only using memory for comparison. This method has been shown to be quite reliable in some situations (Woodworth & Schlosberg, 1954).

4.2.3 Reaction Times

Another psychophysical parameter that is frequently measured is reaction time. The time taken by observers to respond to, or make a decision about a stimulus is often recorded in addition to the observer decisions and can be used as an objective measurement of the difficulty of the task (Münsterberg, 1894; Petrusic, 1993), or used to exclude erroneous trials (for example very long response times may indicate the observer was distracted from the experiment). Alternatively, the reaction time itself may be the experimental parameter under investigation, for example when measuring performance in visual search or other time critical tasks (for example: Treisman & Gelade (1980)), or when the observer task is to estimate the duration of a stimulus.
4.3 Noise and Signal Detection Theory

Signal Detection Theory is widely used to analyse discrimination data from psychophysical experiments. This theory is based on the idea that cognitive performance is limited by signal variability due to both internal (neural) and external noise. As such, a statistical decision making process must be employed to select the most likely true state of the signal (see Green & Swets (1966), Tanner, Wilson, & Swets (1954) and Swets (1996)).

If the noise in a signal is assumed to have a probability density function that can be described by a Poisson distribution (a reasonable assumption since the Poisson Distribution describes probability of a given number of independent events occurring in a fixed interval of time), then how easy a signal is to detect (its discriminability) is dependent on the overlap of the probability distributions of the noise alone, and the signal plus noise. The overlap is quantified by the parameter $d'$, which is the normalised difference between the distribution means (see Figure 4.2) and also known as the observer sensitivity index. In order to decide if a signal is present or not, a criterion value is needed. Responses above this value are taken as a positive result for a signal, and a negative result is given for a response below the criterion value. It can be seen from Figure 4.2 that if there is significant overlap between the noise and signal + noise probability distributions (i.e. the signal is not large compared to the noise), then any criterion value that is covered by both distributions will result in both false negatives and false positives (false alarm). The value of the criterion determines this response
bias, and also affects the shape of the psychometric function and associated JNDs.

**Figure 4.2.** Signal Detection Theory: Probability density functions of noise and signal + noise. The chosen criterion value determines the response bias and the resulting levels of false positive and false negative responses.

In the case of probability density functions that take the form of the Poisson distribution, the variance is proportional to the mean. However, this may not always be the case. When the distribution of noise takes some other form, then the discriminability of the signals, $d'$, is given by Equation 4.1 below. In the case of Poisson distributions, the spread of the distributions is divided out.

$$d' = \frac{\text{separation of means}}{\text{spread of distributions}}$$

**Equation 4.1.** Discriminability of noise and signal + noise.
The theory of signal detection also applies to the discrimination of two signals (plus noise in each) as well as the detection of a signal from noise alone. These correspond to the tasks of detection and discrimination described above in section 4.1.

In the case of cue integration (see chapter 3), signal detection theory predicts that discriminabilities from the single cue conditions \((d'_1, d'_2)\) should combine as shown in Equation 4.2, assuming equal and independent noise.

\[
d' = \left( d'_1 \right)^2 + \left( d'_2 \right)^2 \frac{1}{2}
\]

**Equation 4.2.** Discriminability when combined of two signals of independent discriminabilities \(d'_1\) and \(d'_2\).

Many of the traditional psychophysical methods described earlier in this chapter are not able to distinguish the independent factors of sensitivity index (discriminability of a signal) and the decision criterion used by the observer that combine to produce a threshold value. This is because if an observer uses, for example, a very strict criterion, the number of positive observations will be reduced and the 50% detection threshold for a signal (see section 4.1) will be at increased signal level/stimulus intensity. However, it is possible to measure observer sensitivity by using a 2AFC task, since the same decision criterion is necessarily applied to all observers.
4.3.1 Receiver Operating Characteristic (ROC) Curves

In a single forced choice (yes/no) experiment, using two types of trials, one with and one without a stimulus, both sensitivity index and criterion value can be calculated from the hit rates and false alarm rates. The hit rate and false alarm rate for an experiment are typically plotted against each other in what is known as a Receiver Operating Characteristic (ROC) curve (Swets, 1996). As the criterion value used to determine the presence of a signal is changed, so is the response bias (i.e. the relative rates of hits and false alarms). This is represented in the ROC curve (see Figure 4.3); different criterion values are represented by different points along the curve. The criterion value chosen may vary from observer to observer, and can be affected by the experimental conditions and instructions. The area under the ROC curve quantifies the discriminability of the signal, $d'$. If the curve is a straight line (area under curve = 0.5) then observer performance is at the level of chance (signal not discriminable); 50% of signals are detected and 50% of non-signals cause a false alarm. As the area under the ROC curve increases towards 1 observer performance increases, with a greater number of hits than false alarms, with a perfectly discriminable signal having an area of 1 under its ROC curve.

The ROC curve provides independent measures of criterion value and sensitivity index since different criterion values only create different points along the curve. The shape of the curve, and thus the measure of $d'$ is not dependent of the choice of criterion value.
**Figure 4.3.** Example of a Receiver Operating Characteristic (ROC) curve. Data from a single forced choice experiment is plotted, Hit Rate (y-axis) against False Alarm Rate (x-axis). The dashed line represents a chance level of signal discrimination. The solid line represents an improved level of discrimination, $d'$, measured by the area under the curve.

### 4.4 Other Considerations

When performing psychophysical experiments, there are a number of factors that must be considered in order to ensure reliable and accurate results. Important considerations in the design of experimental paradigms include both physical and psychological factors. For example, an experimenter must be careful to control the influence of factors external to the experimental stimuli, such as background light or other undesired stimuli that may influence the observer’s perception. In the case of the experiments described in this thesis for example, care was taken to try to exclude cues to stimulus shape other than those being tested, such as the screen surround and room background.
Another important psychological effect that must be controlled for is adaptation. Observers will tend to adapt to a stimulus to some extent. An obvious demonstration of this is the adaptation of the visual system to different luminance levels, the sensitivity of which can vary by up to 14 log units (Hood & Finkelstein, 1986; Rushton, 1962) over a course of up to 10 minutes. Care must be taken when designing an experiment to allow for (or mitigate the effects of) any observer adaptation. Other temporal effects on observer performance may also need to be taken into account when designing an experiment. For some types of task, the possibility exists that different mechanisms may be used to process information over different lengths of time or in different situations (e.g. Tolhurst (1975)). To mitigate such effects experiments should be designed to use presentation times suited to the process under study.

Additionally, learning effects must also be considered. If an observer repeats a task many times, perhaps with the intention of improving the accuracy of results, they may learn some details of the stimuli or become habitual in their responses, biasing the result. For this reason, observers that are naïve to the purpose and stimuli of an experiment are usually used, and stimuli are usually randomised in presentation order.
Chapter 5 – General Methods

This chapter details the techniques that were used when creating and displaying stimuli for all the experiments described in this thesis. Also described are details of the experimental apparatus that were used throughout the experiments.

5.1 Computer Rendering

A range of techniques exist to create two dimensional images of a three dimensional scene using computational techniques. This process is often referred to as 3D image rendering. In the majority of cases, a numerically defined three-dimensional geometrical model is combined with lighting and surface reflectance information to create a two-dimensional image, as it would appear seen from a particular view point (see Appel (1968)). The scene model must contain information about the three-dimensional geometry of the visible objects, along with information about the viewing position and direction, in order to create a geometrically accurate rendition of the scene from a given view point. Additionally, information about light sources, such as their position, intensity and colour must be present so the scene can be lit correctly. Finally, information of the surface reflectance of the objects within the scene must be present, such that surface appearance can be calculated correctly. With these three data sets, along with a suitable model of physical light /surface interactions, a physically accurate image of the scene can be produced.
In order to transform the information in the model into a plausible image, rendering software typically simulates the physical interactions of light and surfaces that give rise to object appearance in the real world. The light arriving at the viewing position is calculated from the knowledge given in the model of the light sources and the surface reflectance and position of the objects in the scene, using the known properties of visible light and typical surfaces.

Unfortunately, full simulation of the physical world is a very computationally intensive task and is not generally feasible. Many techniques are used in computer graphics to reduce the time and computational overhead required to render scenes, often using approximations and simplifications of the real world. For example, the true physics of light reflection will often be ignored, such that light originating from an object surface may be given a fixed colour, independent of the light source colour, or reflections may be given an equal intensity in all directions (Lambertian reflection). Whole surfaces may also be grouped and assigned the same pixel values in the final image so the value only needs to be calculated once.

### 5.1.1 Ray Tracing and Monte Carlo methods

If very high accuracy is required in the final image, approximations such as those described above are clearly not suitable and more sophisticated techniques such as *ray tracing* (Whitted, 1980) must be used. The ray tracing process aims to simulate the physical world by tracing the direction and intensity of light rays in the scene and modifying them as they interact with
object surfaces, using physically accurate models of surface reflectance. Ray tracing typically traces rays from the viewpoint back towards the light source. The reason for this is that the majority of light emitted from a source never reaches the viewpoint, so calculation of these rays would be wasted. Tracing all rays (or at least sufficient for the required resolution of the final image) from light sources, in a process (confusingly) called ‘backwards’ ray tracing, is computationally intensive process and often not feasible, due to processing time constraints. However, backwards ray tracing can provide inter-object reflection (global illumination) effects that are missing in standard ray tracing and can therefore provide improved accuracy. The large number of ray paths that must be calculated to provide global illumination can be prohibitive. However, techniques exist to reduce the number of rays that must be calculated. A common technique used is Monte Carlo sampling (Metropolis & Ulam, 1949). This statistical sampling method computes a relatively small number of sample rays and combines these data with algorithms designed around statistical knowledge of likely solutions. This results in good approximations to the illumination without having to calculate all ray paths, allowing faster rendering.

5.1.2 RADIANCE Synthetic Imaging System

The RADIANCE lighting simulation software (Ward, 1994) is a suite of tools for rendering physically accurate images of virtual scenes. It is freely available, and the software and its use are detailed in the book Rendering with Radiance (Larson & Shakespeare, 1998). A modified version of RADIANCE (see section 5.3.1 - ‘Hyperspectral’ rendering) was used to
create all the rendered stimuli used in the experiments described in this thesis. RADIANCE uses a combination of deterministic and Monte Carlo ray tracing approaches to achieve a compromise between accuracy and rendering time. Rays are traced from viewpoint to source, and direct illumination, specular and diffuse inter-reflections are calculated separately. Surface reflectances can be defined in a number of ways, ranging from a perfectly diffuse (Lambertian) model to a full Bi-Directional Reflectance Distribution (BRDF) model (see for example Nicodemus (1965) for more details on BRDFs). The RADIANCE software and reference manual are available for download at http://radsite.lbl.gov/radiance/.

5.1.3 ‘Hyperspectral’ rendering

RADIANCE, along with most other rendering software, only calculates three ‘colours’ or spectral wavebands, typically red, green and blue (RGB triplet). While only three primaries are required for producing a wide range of chromaticities (see section 5.3.1 on colour spaces and section 5.3.3 on display devices) this approach at a rendering stage can lead to inaccuracies, since the spectral power distribution of light sources and spectral reflection curves of surfaces cannot be accurately described by only three wavebands. This problem has been addressed by a number of researchers (e.g. Johnson & Fairchild (1999) and Ruppertsberg & Bloj (2006a, 2008)). Ruppertsberg and Bloj extended the usual RADIANCE rendering process to include many more than 3 wavebands. By rendering multiple images of the same scene, and assigning light source and surface reflectance values for narrow wavebands (potentially only a few nm in width) to the ‘red’ ‘green’ and ‘blue’
channels for each sub-image, a ‘hyperspectral’ image can be created. The multiple renderings, each of which covers only a small part of the visible light spectrum, are combined into a single image in CIE XYZ colour space, using the CIE colour matching functions (see section 5.3.1 and Figure 5.2) and from there to calibrated RGB values for display (see section 5.3.3.1 for technical details on this conversion). The improvement in representation of a colour signal by an increased number of sample wavebands is shown in Figure 5.1.

![Figure 5.1](image.png)

**Figure 5.1.** Example of the improved representation of a colour signal with an increased number of sample waveband (N). From Ruppertsberg (2006).

If the number of wavebands (N) is increased significantly, for example to 81 (5nm wide wavebands), a very accurate signal can be obtained, but a correspondingly longer time to render images is required. For 81 wavebands,
27 images must be rendered to produce the final hyperspectral image (3 wavebands are rendered per image).

RADIANCE, using the hyperspectral rendering technique of Ruppertsberg & Bloj (2008) was used to render stimuli for the experiments described in this thesis. Further details of how the stimuli were created for the experiments are given in the section 5.5.

After discussion with the developer of RADIANCE, Greg Ward, it became apparent that there are some small differences between the way RADIANCE handles each of its three standard colour bands (R, G and B), which under normal (not hyperspectral rendering) improves performance. This is not ideal for hyperspectral rendering as described above, since the wave bands were assigned to sub-image renderings sequentially (i.e. the first three wave bands with the shortest wavelengths were assigned to the first sub-image rendering and so on), with the assumption that no wavelength specific differences occur between waveband calculations. It was suggested that this issue would make little difference, but can be mitigated by interleaving the wave bands such that each of the 27 rendered images contains one band from the long, one from the medium and one from the short wavelength parts of the spectrum. This improves accuracy, although it was found that it also increases the processing time needed to create the final display image from the multiple renderings.
To test the effect of this modification to the rendering process, two images similar to the stimuli used in Experiment 1 were rendered: one image where adjacent wavebands were used for each sub-rendering, and one where wavebands were interleaved. Comparing these renderings pixel by pixel, it was found that 0.64% of pixel values differed between the images. This represents a small difference, but of a similar magnitude to the difference between images of stimuli in which the object shape varies by 1 degree (see section 5.5 for details of the stimuli), so potentially significant. For this reason, interleaved wavebands were used when rendering all the experimental stimuli.

5.2 Stimuli Display

An important part of any psychophysical experiment is the accurate display of stimuli and the control of display conditions to ensure valid and accurate results. This section describes the equipment and methods used in this study to achieve these aims.

5.2.1 Colour Spaces

This section provides background information on colour science that is useful for understanding the following sections related to the generation of colour-accurate images and their presentation on computer displays.

The human visual system is not capable of determining the exact spectral power distribution of incident light, since it is reliant on only four types of
receptors, of which three (the three cone types) are generally considered to contribute to the colour system (Svaetichin, 1956). Because these receptors must be broadly tuned in their response to cover a usefully wide range of wavelengths, a signal built from the responses of these three receptors cannot fully represent the spectral distribution of the stimulating light. The normalised responses of the three colour receptors (cones) in the human retina are show in Figure 5.2.

![Figure 5.2. Normalised cone responses (Stockman & Sharpe, 2000).](image)

It is possible to calculate the approximate colour percept of any particular spectral power distribution via the CIE colour matching functions (CIE, 1931), shown in Figure 5.3.
Figure 5.3. CIE XYZ colour matching functions (CIE, 1931).

These functions when integrated with the spectral power distribution \( I(\lambda) \) provide the ‘tristimulus’ values of the colour signal in the CIE XYZ colour space (for a standard observer):

\[
X = \int_{380}^{780} I(\lambda) \bar{x}(\lambda) d\lambda \quad \text{Equation 5.1}
\]

\[
Y = \int_{380}^{780} I(\lambda) \bar{y}(\lambda) d\lambda \quad \text{Equation 5.2}
\]

\[
Z = \int_{380}^{780} I(\lambda) \bar{z}(\lambda) d\lambda \quad \text{Equation 5.3}
\]
X, Y and Z are the three ‘primaries’ of an additive colour model required to provide the same colour percept as the original signal $I(\lambda)$. The XYZ colour space was specifically designed such that the Y primary corresponds to the luminance of the signal (CIE, 1931). The chromaticity of a colour signal is given by two parameters derived from the X, Y and Z values, called $x$ and $y$:

$$
x = \frac{X}{X + Y + Z} \quad \quad y = \frac{Y}{X + Y + Z}
$$

Equations 5.4 and 5.5

The derived colour space formed by these parameters is known as the CIE xyY colour space (CIE, 1931). This is perhaps the most widely used colour space and is shown in Figure 5.4.
Figure 5.4. An equiluminant plane in the CIE xyY colour space (CIE, 1931). Chromaticity is given by $x$ and $y$, on the horizontal and vertical axis. The luminance axis ($Y$) is perpendicular to the plane of the page. The chromaticity of monochromatic (single wavelength) light forms the curved edge of the coloured area; Wavelength in nm is shown on this curve.

Figure 5.4 shows the xyY colour space for a fixed luminance. The curved boundary describes the monochromatic locus. That is, the chromaticities of light of a single wavelength. The straight boundary along the bottom is known as the ‘line of purples’ and the chromaticities here cannot be derived from monochromatic light. The bounded area comprised the gamut of typical human vision and covers all visible chromaticities (Wyszecki & Stiles, 1982). It should be noted that due to the limitations of computer displays and
printing technologies, not all colours are accurately reproduced in the diagram.

Other colour spaces have also been developed, such as attempts at ‘perceptually uniform’ colour spaces (the CIE xyY colour space is not perceptually uniform: chromaticities in different areas of the space are not equally discriminable to observers (MacAdam, 1942)), such as the CIELUV and those built directly from perceptual parameters, like the HSL (Hue, Saturation, Lightness) and HSV (Hue, Saturation Value/Brightness) spaces. Other colour spaces have been developed based on the physiology of colour processing. For example the CIE La*b* (CIE 1976) and ‘DKL’ (Derrington, Krauskopf, & Lennie, 1984) colour spaces that represent colour using axes analogous to the colour opponent pathways of the visual system (for details of the colour opponent mechanisms of the human visual system see, for example: Dacey & Lee (1994), Derrington & Lennie (1984), DeValois & DeValois (1993), Hendry & Yoshioka (1994) and Hering (1964)).

Several colour spaces have also been developed based on colour additive models that are particularly suited to defining colours for printing and computer display, such as AdobeRGB (http://www.adobe.com/digitalimag/pdfs/AdobeRGB1998.pdf) and sRGB (http://www.w3.org/Graphics/Color/sRGB.html). It is possible to calculate R, G and B primary values for these spaces (calculated for the spectral outputs of the red, green and blue phosphors/filters in typical computer monitors) using a set of colour matching functions similar to those described above and
shown in Figure 5.3. Alternatively, conversion from XYZ to RGB colour space is possible via a matrix multiplication. This matrix can either be for a standard monitor, or preferably calculated from measurements of the monitor’s primaries (see section 5.3.3.1 for practical details on this type of monitor calibration process).

The calculation of colours in an RGB colour space in this way allows a relatively precise representation of any colour signal on a computer monitor (within the gamut of the monitor), providing the same perceived colour and brightness, despite the inability of such displays to necessarily reproduce the same spectral power distribution as the original signal. However, the range and number of colours available is limited by the technology used (see section 5.3.3).

### 5.2.2 Cambridge Research Systems ViSaGe

When attempting to perform studies involving coloured images displayed via computers, it is important that as well as creating stimuli that are as colour-accurate as possible, the method of presentation/display must preserve this accuracy as far as possible. Standard computer graphics hardware is currently typically capable of 8 bits per channel resolution (or 24 bits in total over the red, green and blue channels). It has been shown that while 24 bit systems may be suitable for some experiments varying in luminance only, in general studies involving chromatic stimuli require higher resolution (Garcia-Suarez & Ruppertsberg, 2010). The inadequacy of a 24bit colour system results from the fact that in some areas of the RGB colour space the steps between adjacent colours in a 24 bit system are discriminable to human
observers, being larger that the ‘MacAdam ellipses’ (MacAdam, 1942) at those chromaticities. To improve on this, systems that provide a greater ‘colour depth’ are available, such as the Cambridge Research Systems (CRS) ViSaGe, which can provide calibrated 42 bit (14 bits per channel) colour resolution (http://www.crsltd.com/catalog/visage/overview.html). Cambridge Research Systems also provide a Matlab (http://www.mathworks.com/products/matlab/) Toolbox for use with the ViSaGe system. This toolbox contains a variety of functions to generate and control the display of stimuli. All the software developed for stimuli presentation and collection of observer responses during the experiments described in this thesis was written using Matlab and the CRS Toolbox.

5.2.3 Display Devices

A number of different technologies are currently available for the display of computer images. The most common display types are Cathode Ray Tube (CRT), Plasma and Liquid Crystal Displays (LCD). Although Plasma and LCD display are widely used commercially and domestically, they have some significant limitations when it comes to performing psychophysical experiments. Both suffer from high ‘black levels’, i.e. they still produce a significant amount of light even when there is no signal. LCD displays also suffer from a relatively poor response time and thus exhibit noticeable motion artefacts when moving images are displayed. While CRT technology is much older, modern CRT devices have very good characteristics, with a fast response rate, high spatial resolution, wide colour gamut and high contrast. For these reasons, CRT monitors were used to display stimuli in all of the
experiments described in this thesis. For more details on the comparison of different display technologies and their use for psychophysical experiments see Brainard, Pelli, & Robson (2002), Chen, Cheng, & Shieh (2005) and Elze, Lochmann, & Tanner (2007).

The array of pixels used in CRT monitors consists of three different phosphor types, creating the three display primaries (other display types also tend to use three types of colour pixel, but use different materials). The chromaticity of these primaries determines the gamut of the monitor, as demonstrated in Figure 5.5.
Figure 5.5. The sRGB gamut typically used for computer display, represented in CIE xyY colour space (http://www.w3.org/Graphics/Color/sRGB.html).

The corners of the coloured triangle in Figure 5.5 are at the chromaticites of the sRGB primaries, and are similar to the primaries used in many monitors. All the colours within the bounding triangle can be produced by a combination of the three primaries, while colours outside the triangle cannot be reproduced.

5.2.3.1 Display Calibration

When attempting to display stimuli of known chromaticities on a computer monitor the conversion from xyY to RGB colour space (see section 5.3.1)
requires accurate knowledge of the RGB primaries of the monitor (i.e. the output spectra of the red, blue and green pixels in the display matrix). This conversion can be done using a display colour space such as sRGB (shown in Figure 5.5). However, the chromaticities of the primaries of any real monitor used will inevitably be different those of a generic colour space. In order to perform the conversion correctly, one must measure the primaries of the monitor in use. This can be done using a colorimeter such as the Cambridge Research Systems ColourCAL
(http://www.crsltd.com/catalog/colorcal/index.html). The colorimeter is used to measure the chromaticities of the primaries of the monitor, and their full range of luminance. The measurement of the chromaticity of each primary enables accurate calculation of the RGB input to the monitor required for a desired output chromaticity. Measurement of the luminance of each primary as a function of electrical input level enables correct ‘gamma’ correction, (‘gamma’ describes the relationship of output luminance to numerical monitor input; typically this is an exponential function, the exponent of which is referred to the gamma; (Poynton, 2003)). The measured information about the colour and luminance functions of the monitor primaries allows the correct conversion from xyY to RGB (and vice versa). By use of this calibration process, which was performed for the monitor used in the experiments described here, it is possible to ensure stimuli are presented accurately at the required chromaticity and luminance.

When performing such a calibration, there are some assumptions that are still made regarding the display, namely phosphor/pixel constancy across screen (since measurements are only usually made in the central area of the
screen) and phosphor independence (see Brainard (1989) for details). To mitigate problems that might be caused by a lack of phosphor/pixel constancy across the screen, a Photo Research Inc. PR650 spectrophotometer was used. Luminance and chromaticity measurements were made of each primary at different areas of the monitor screen used for the experiments, and the monitor adjusted to provide the most spatially uniform output possible.

5.3 Use of rendering in psychophysical studies

Computer rendering has been used to provide a realistic representation of three-dimensional objects or scenes in a number of psychophysical studies of object perception (for example, see Delahunt & Brainard (2004b), Nefs & Harris (2008) and Ruppertsberg & Bloj (2007)), or to replicate specific illumination effects only present in three-dimensional scenes, i.e. inter-reflections (Delahunt & Brainard, 2004a; Ruppertsberg et al., 2008). Of course the use of such stimuli is only useful if the rendered stimuli are sufficiently accurate to represent the phenomena under test. Researchers often assume this accuracy, and as noted previously, there is some evidence that basic image presentation techniques may not be accurate enough for some psychophysical studies (Garcia-Suarez & Ruppertsberg, 2010). A number of recent studies such as McNamara (2001), McNamara, Chalmers, Trosclair, & Reinhard (1998) and Ruppertsberg & Bloj (2006a, 2006b) have helped to develop techniques for assessing the realism of rendered scenes in comparison to the real world. While these studies have shown that the perceptual accuracy of rendered stimuli can be sufficient some types of
experiments, they also suggest that errors, such as shifts in colour space, are typically present in rendered scenes and may be detectable by observers. Other authors have attempted to address the issue of accuracy by providing methods of increasing the precision of the rendering process. For example, Ruppertsberg & Bloj (2008) demonstrated a technique for improving the colour accuracy of images rendered with Radiance, as described above in section 5.2.3.

Since the aims of the studies described in this thesis are to investigate the effects of colour gradients on shape perception, the fact that colour inaccuracies are often present in rendered 3D objects could be problematic. It is for this reason that great care has been taken when creating the rendered stimuli used in the experiments so as to produce stimuli with the most accurate colour properties possible. Details of the specific stimuli and how these were created for the experiments are given in the following section.

5.4 Details of the Stimuli used in the Experiments

Stimuli were specifically designed to allow an investigation of the effects of colour shading, particularly from mutual illuminations, on the perception of 3D object shape. Therefore, stimuli were based on the ‘Chromatic Mach card’ of Bloj et al. (1999), a stimulus where one can vary luminance and chromatic shading gradients dependent on a single shape parameter (the angle of the central fold). In order to easily control the different cues to 3D shape present in the stimuli, they were computer rendered rather than real objects.
This type of stimulus consists of a rectangular card folded at a colour border creating either a concave ('corner shape') or convex ('roof shape') dihedral angle. Using this type of stimulus also allows for a wide range of shapes.

With careful lighting, the full range of physically possible folded card stimuli can be used without introducing other shape cues such as occlusions and shadows, or artefacts that can often be a limiting factor in shape from shading studies.

To ensure that the stimuli were as accurate as possible, they were rendered using RADIANCE and the hyperspectral rendering technique described in section 5.2.3 and displayed using a Cambridge Research Systems ViSaGe system running in 42bit colour mode (see section 5.3.2), attached to a calibrated Mitsubishi Diamond Pro 2070 monitor. Initially, three sets of stimuli were created, one for each condition of Experiment 1:

1. Gradient-cue-only condition: Folded cards rendered at 1000x1000 pixels to cover the full screen, displayed at a scaled resolution of 632 x 949 (the horizontal resolution available is halved when using the CRS ViSaGe in 42bit colour mode).

2. Outline-cue-only condition: Wire frame stimuli, with veridical outlines as for a card that had sides 10x10cm in size and viewed from 1m distance (the same as the actual observer viewing distance), but with no surface shading. For the flat card this corresponded to a rectangle of 11.4 by 5.7 degrees. These stimuli were not rendered but drawn in real-time by the CRS ViSaGe system.
3. Gradient+Outline condition: Rendered folded card stimuli, identical to the Gradient-cue-only condition, but with the addition of a mask to create an object outline that was identical to that of the Outline-cue-only condition.

Examples of the threes stimulus types are shown in Figure 5.6:

![Figure 5.6 Example stimuli. Top: 'corner configuration'. Bottom: 'roof configuration'. From left to right: Outline-cue-only, Gradient-cue-only, Gradient+Outline. In all cases the card angle is 50 degrees (internal angle of 80 degrees). Note: in the Gradient-cue-only condition, the edges of the stimulus images were not visible to observers, due to the use of a viewing aperture (see section 5.6).](image)

Stimuli were created with a range of card angles (defined as the angle of the card surfaces to the horizontal axis perpendicular to the viewing direction) ranging from -70 (‘roof’ shape) to +70 (‘corner’ shape) degrees, in steps of 10 degrees and were displayed on a grey background, luminance 10 cd/m². Figure 5.7 shows a schematic top-down view of the stimulus configuration.
In the case of the Gradient+Outline stimuli, the shading on the card surfaces is not entirely veridical. This is due to the method used to render the stimuli, whereby large cards that extended beyond the field of view were first rendered and a ‘mask’ overlaid to give the required outline. This method was chosen to make sure that, for a given card angle, the shading in both the Gradient-cue-only condition and the Gradient+Outline condition was identical.

A downside of using this method for creating Gradient+Outline stimuli is that a slightly greater level of mutual illumination is present on the Gradient+Outline stimuli than would be strictly correct if the cards were the size they appeared, but such an inconsistency is unavoidable if identical...
shading gradients are to be present in stimuli with and without visible edges.

A pixel by pixel comparison of rendered chromatic Mach cards with either correctly sized sides, or larger sides and a mask, revealed that only around 5% of pixels were different between the two approaches for a concave card at 50 degrees (80 degree internal angle). This suggests that even at angles where significant mutual illumination is present (e.g. the 50 degree card tested), the inaccuracies in the images created by the masking method of creating the object outline, as described above, are not very large. This issue is also present to some extent in the Gradient-cue-only condition, because all stimuli were viewed through an aperture that restricted the viewing angle (see section 5.6).

The material properties used for rendering the card surfaces were chosen from the Natural Color System (NCS) papers (http://www.ncscolour.com/). The use of real materials, with existing spectral surface reflectance data, allowed surface colours to be easily defined for physically accurate hyperspectral rendering (Ruppertsberg & Bloj (2008); see section 5.2.3). NCS_S0300N was used for the white card side and a red coloured paper (NCS_S0580Y90R) for the coloured card side. This particular pair of colours was chosen due to the high reflectance of the white and the high colour saturation of the red, resulting in a relatively large amount of chromatic mutual illumination reflected from the white side of the Mach card for highly convex shapes. Materials that created a significant amount of mutual illumination were chosen with the intention of maximising the chances of observing any effects of mutual illumination on perceived shape: In the case
of materials that are less reflective, or card sides that are more similar in colour, the contribution of mutual illumination to the resulting image of the stimulus would be lower. Any shape cue resulting from the complex light inter-reflections would therefore most likely be weaker and may not contribute significantly to shape perception.

5.4.1 Rendering Details

Stimuli were rendered using RADIANCE installed on computer running a 64-bit Linux operating system and featuring an Intel Core 2 Quad processor with 8 GB of RAM. Batch processing was used via shell scripts written specifically for this task.

5.4.1.1 Scene Geometry

As noted above, large card planes were used that extended beyond the rendered field of view (i.e. full screen). Surfaces were rendered under a D65 spectrum light source that consisted of a spherical point source positioned in front, above and slightly to the right of the stimulus (x,y,z co-ordinates of 13.33, 16.66 and –66.66cm respectively) such that significant shading effects from mutual illumination and luminance fall off towards the edges were present in the stimuli, and this shading varied noticeably with card angle. As noted above, the mask to provide object outline was created and placed over the image at the time of display, using CRS Toolbox functions, rather than during rendering. This resulted in fewer stimuli needing to be rendered since the mask could be changed without re-rendering.
5.4.1.2 Running Radiance on Multiple Processors

In order to reduce rendering time for stimuli, several RADIANCE threads can run independently. The RADIANCE function `rpiece` is used to split rendering into multiple parts that can then be rendered individually on multiple processors. Since the computer used for rendering had four processing cores, rendering was split into four parts.

5.4.1.3 Creating Hyperspectral and 42-bit Images

The process of hyperspectral rendering involves producing binary image files from each three-waveband rendering and combining these binaries into a single hyperspectral image (see section 5.2.3 for details of the hyperspectral rendering technique). The hyperspectral image must be converted to RBG format for display (42-bit resolution in this case for use with the Cambridge Research Systems ViSaGe). This is done by multiplying the spectral image by the CIE colour matching function to convert first to CIE XYZ space, and then converting from XYZ to RGB space (see section 5.2.1 for details).

5.4.1.4 Controlling for Variability in Rendering

Due to the Monte Carlo sampling methods employed in RADIANCE (see section 5.2.2) there likely to be small differences between separate renderings of the same image. In order to quantify these differences, many copies of the same scene were rendered and the results compared pixel by pixel. Comparison was done on filtered images 1000x1000 pixels (filtering is
performed by the RADIANCE function pfilt, which reduces image size and performs anti-aliasing). The scene chosen was of a chromatic Mach card with significant mutual illumination (70 degrees plane angle, or 40 degrees internal angle). Such a tightly folded card shape was chosen because the multiple light bounces needed to produce mutual illumination are affected to a greater extent by the stochastic nature of the Monte Carlo rendering process. It was found that there was no difference between a set of images rendered in close succession on the same machine, indicating that the same seed for the pseudo-random Monte Carlo sampling was used for each rendering. Images rendered in different batches either on the same or different machines were found to have between 0% and 0.061% of pixels different. Images rendered on the quad core desktop machine, using each processor core to render \( \frac{1}{4} \) of each image concurrently (see section 5.5.1.2), were found to have a similar variation between images rendered in the same batch as images rendered in separate batches but in a single process. Presumably each rpiece process uses a different seed for the Monte Carlo sampling.

As a means of providing objective comparison for the image differences quantified above, two images with planes with a 1-degree difference were compared (69 and 70 degree plane angles). The difference was 0.62%, around 10 times that due to the random sampling. Since the variation due to random sampling is one to two orders of magnitude smaller than the minimum changes between stimulus shapes (2 or 10 degree steps were
used, depending on the experiment), it should not have a significant impact
of the accuracy of results using these stimuli.

5.5 Experimental Apparatus

The experiments were controlled by Matlab software, using the CRS Toolbox
and the Psychophysics Toolbox (Brainard, 1997). Stimuli were displayed
using the CRS ViSaGe (see section 5.3.2). For experiments 1, 2 and 3
observers made estimates of the angle of the card stimuli by adjusting the
angle between two lines in a ‘view-from-above’ configuration, displayed on a
separate monitor (see Figure 5.8), drawn using Psychophysics Toolbox
functions. Observers were able to adjust the angle of the two lines to match
the stimulus using a rotating response dial. In experiment 4, observers
performed a 2IFC experiment (see section 4.1.5) and responded by pressing
their choice of two buttons on a Cambridge Research Systems response box.

Figure 5.8 Monocular viewing apparatus. On the left is the screen displaying the matching
task lines. On the right is the box with viewing aperture that surrounds the stimulus monitor.
Two types of viewing apparatus were used for the experiments: monocular and binocular versions. In the monocular case, the experimental setup was designed to eliminate cues to shape that were not under investigation, such as pictorial cues from the screen surround and binocular disparity. To achieve this, a box was built that surrounded the stimulus monitor with a small circular aperture at a distance of 1m from the screen (see Figure 5.8). The box had matt black internal sides to avoid reflections of light from the screen. The aperture provided a field of view that ranged from around 14 to 15.5 degrees in diameter, dependent how close the observer’s eye was positioned to the end of the viewing aperture (the distance was not fixed and could vary, dependent on factors such as whether observers were wearing spectacles or not). This range assumes a variation in eye position of up to 20mm. The aperture ensured that viewing must be performed monocularly and the majority of the screen was visible, but not the edges or surround – this was the case for the full field of view range noted above.

The binocular version of the experimental apparatus replaced the viewing box and aperture with a mirror stereoscope that was configured to have the same optical path distance as in the monocular set up. For experiments that included binocular cues, stimuli were rendered in stereo pairs, with a separate image for each eye’s viewpoint. The images were rendered with viewpoints 1m from the stimulus and separated by a typical interpupillary distance of 65mm. Figure 5.9 shows a schematic of the stereoscope arrangement.
Figure 5.9 Arrangement of the mirror stereoscope used for Experiment 1. Labels M1 and M2 indicate the first and second mirrors respectively, for left (l) and right (r) eye images.

The techniques and apparatus described in this chapter were used throughout the experiments detailed in this thesis. These experiments were designed to investigate the contribution of colour illumination gradients to three-dimensional shape perception, and all the experiments used variations of the chromatic Mach card stimuli described above. The first experiment undertaken is the subject of the next chapter.
Chapter 6 - Experiment 1: Colour illumination gradients as a cue to 3D shape.

The experiment described in this chapter was designed to investigate how the perceived 3D shape of the ‘chromatic Mach card’ stimuli (described in section 5.5) depends on the available shape cues. Shape cues present in the stimuli are: colour gradients from shading and mutual illumination, object outline (perspective), and in the case of binocular viewing, stereo gradients (the disparity between the gradients seen by each eye) and stereo outline. While a number of studies have investigated the contributions of surface shading (e.g. Bülthoff & Mallot (1988), Langer & Bulthoff (1999), Mamassian & Kersten (1996), Mingolla & Todd (1986) and Nefs (2008)) and perspective (e.g. Knill (2007a, 2007b) and van Ee et al. (2002)) to 3D shape perception, most have used achromatic stimuli and only a limited number of studies have looked at the specific combination of shading and perspective outline cues (e.g. Humphrey, Symons, Herbert, & Goodale (1996) and Wagemans et al. (2010)). In order to investigate how colour illumination gradients may contribute to perceived object shape, either separately or in combination with other cues, Experiment 1 investigated the perceived 3D shape of the card stimuli under a number of conditions: First, shape perception was measured when either realistic complex colour shading or perspective outline cues were available individually (Gradient-cue-only and Outline-cue-only conditions). Secondly, perception was tested when both these monocular cues were present (Gradient+Outline condition). Finally, perceived shape was also
measured for binocular versions of each of the conditions described above. See section 5.5 for details of the stimuli for each condition.

Some previous work has suggested that when stereo cues to depth are present, shape from shading is vetoed (Bülthoff & Mallot, 1988). If this were the case, then we would expect observers to make the same settings (and have the same variability in settings) when viewing binocular versions of the Outline-cue-only stimuli and Gradient+Outline stimuli. Alternatively, cue integration models of the ‘weak-fusion’ class (see Chapter 3 for details) predict that, while the mean shape settings in monocular and stereo conditions should be the same (because, in this case, all the available cues are congruent), variance will be lower in the stereo case, because more information, from a greater number of cues, is available.

No ‘stereo-gradient’ condition was tested because the stereo viewing apparatus (Figure 5.9) could not be integrated with the viewing aperture arrangement used in the monocular conditions. This arrangement is particularly important for stimuli without any outline cue because without the viewing aperture the edge of the screen provides an erroneous outline shape cue. Additionally, no ‘disparity-cue-only’ condition is possible since disparate shading cannot be created without shading, and while a disparity-defined outline could be created using random-dot stereograms, such a stimulus would also still contain an additional perspective cue.
Of particular interest in the case of the shading cue is whether colour gradients due to mutual illumination confer any additional information to the visual system over the luminance gradients due to surface orientation (see section 2.1.3 – Shape from Shading). As noted in chapter 2 (see section 2.1.4), previous work has shown that perception of 3D shape can significantly influence perceived surface colour (Bloj et al., 1999), indicating that colour and shape perception are linked via knowledge of mutual illumination. In other words, potential reflections of light between objects appear to be taken into account by the human visual system when making judgements of surface colour. The fact that the visual system is sensitive to mutual illuminations (Ruppertsberg et al., 2008) suggests that the presence of mutual illumination, and the additional shape information it carries (Forsyth & Zisserman, 1990), may influence 3D shape perception. Experiment 1 was therefore specifically designed to enable a comparison of perception when shading contains only luminance gradients, and when colour gradients due to mutual illumination are also available. This is possible with the ‘chromatic Mach card’ stimuli because mutual illuminations are only present when the card is concave and not when it is convex.

6.1 Method

A magnitude estimation task (see section 4.2.1) was chosen for the 3D shape judgement task of Experiment 1. In each condition observers were asked to make shape estimates by adjusting a separate representation of the stimulus to match the perceived stimulus shape. Observers made estimates of the angle of the card stimuli by adjustment of a rotating response dial,
changing the angle between two lines representing the stimulus in a ‘view-
from-above’ configuration (similar to Figure 6.1, right). These lines were
displayed on a separate monitor and observers adjusted them so that the
angle separating one line from the other was the same as the perceived
shape of the stimulus. In the top-down view, lines were 2mm wide and each
line 10cm long (matching the size of the Gradient+Outline stimulus, which
was 20cm wide when flat – see Figure 6.1c). Line and background colour
and luminance were matched to that of the Outline-cue-only stimuli. The
second monitor was also viewed from 1m. This is a similar task to that used
by other authors (Adams & Mamassian, 2004; Todd et al., 2005). The
magnitude estimation method has the advantage of allowing measurement of
biases in observers’ perception, unlike alternatives such as the two-interval
forced-choice paradigm. This is essential in order to explore any prior
constraints that may bias perception (see Chapter 2 for more information on
priors).
Figure 6.1 Example stimuli. Top: ‘corner configuration’. Bottom: ‘roof configuration’. (a) Outline-cue-only. (b) Gradient-cue-only. (c) Gradient+Outline. Note: in all conditions the edges of the screen were not visible to observers, due to the use of a viewing aperture (see section 5.6). Thus in the Gradient-cue-only condition, the edges of the stimuli were also not visible. On the right the diagram shows a schematic of the stimuli configuration, as viewed from above. Observers set a similar pair of lines on a separate monitor to estimate the stimulus shape.

Observers performed the experiment over several sessions and were shown a brief demonstration of how the task should be performed before beginning the first session. This consisted of a short video presentation containing images of the stimulus, with both gradient and outline cues presented together (Figure 6.1c) and also the matching lines. The video showed the stimulus and corresponding line configuration vary throughout the full range of stimuli present in the experiment (two cycles over 30 seconds, showing from the steepest convex to the steepest concave angle and vice versa). The ‘view from above’ task was explained verbally at the same time, using the video demonstration for reference.
First, observers performed matches to the single cue conditions (Outline-cue-only and Gradient-cue-only), using the monocular viewing apparatus. The order of these conditions was randomised for each observer. Each condition required 150 observer estimates, ten for each of the available card angles. Observers were allowed as much time as they wished to make each match, but were encouraged to perform the task as quickly as possible, making estimates from their initial impressions of object shape. Typically each session was completed in around 25 minutes. After observers made estimates for stimuli consisting of single cues, they then performed matches for the Gradient+Outline condition, again using the monocular set up. After making settings for monocular stimuli, observers performed shape matches for two further conditions where stimuli consisted of stereo pairs: 1) a ‘stereo-outline’ condition, where stimuli consisted of stereo pairs of Outline-cue-only images (Figure 6.1a); 2) a ‘Stereo Gradient+Outline’ condition, where stimuli consisted of stereo pairs of Gradient+Outline images (Figure 6.1c).

The order of experimental conditions was intended to reduce the chance of observers applying any information learnt from more informative conditions with a greater number of cues to the conditions with more simple stimuli, particularly the ‘single-cue’ conditions (monocular Gradient-cue-only and Outline-cue-only).

6.1.1 Observers

Six naive participants took part in the monocular conditions of the experiment. The observers had a mean age of 42 and two were male. All had
normal colour and stereoscopic vision, as well as normal or corrected to normal visual acuity. Normal colour vision was verified using the Farnsworth-Munsell 100 hue test (score below 100). Normal stereo acuity was verified with the TNO stereo test (120 seconds of arc or less). Five of these six observers took part in the binocular conditions of the experiment (one observer was unavailable for the binocular conditions).

6.2 Results and Analysis

6.2.1 Discriminability of the stimuli

Before analysing the differences between the settings observers made for stimuli at different angles, it is important to consider how discriminable the stimuli are. In the gradient-only condition, the stimuli consist of both luminance and colour gradients and previous literature surrounding discrimination of such gradients is limited. Ruppertsberg et al. (2008) investigated similar gradients and characterise them by total (pooled) cone contrast:

$$C = \sqrt{(C_L^2 + C_M^2 + C_S^2)}$$

Equation 6.1 Total (pooled) cone contrast. $C_L$, $C_M$ and $C_S$ are long, medium and short cone contrasts respectively (see section 5.3.1; Chaparro, Stromeyer III, Huang, Kronauer, & Eskew (1993)).

Cone contrasts can be calculated for the gradients in the stimuli by conversion of the original renderings from CIE XYZ colour space to the LMS
colour space (see section 5.3.1), via the Bradford Transformation Matrix (see Westland & Ripamonti (2004)). The long (L), medium (M) and short (S) cone contrasts can then be calculated as the ratio of the change in L, M and S values across the gradient to the average L, M and S values. Total cone contrast is then calculated via equation 6.1.

Figure 6.2 shows calculated total cone contrasts for the gradients present on each side of each stimulus used in Experiment 1. Contrasts were calculated for horizontal gradient samples across the stimuli at half height. In the previous study by Ruppertsberg et al. (2008), changes in gradients that resulted in changes in total cone contrast of above 4% (stimuli contained gradients ranging from 4.6% to 13.6% total cone contrast) were found to be discriminable by observers. The smallest magnitude gradient in the stimuli used in Experiment 1 was present on the coloured side of the card for an angle of +20 degrees (concave) and resulted in a cone contrast of 5%. The minimum difference in total cone contrast between two consecutive angles was 12% (+20 and +30 degrees, on the coloured side of the card). The gradients in the stimuli used in Experiment 1 are outside the range studied by Ruppertsberg et al. (2008), but in general these gradients, and the differences between those in adjacent stimuli, are large compared to those in Ruppertsberg’s study (see Figure 6.2). For this reason it seems likely that the gradients should be visible in all the Gradient-cue-only stimuli and that the stimuli may be discriminable from each other.
Figure 6.2 Total cone contrasts of the horizontal gradients (at half height) present on each side of the chromatic Mach card stimuli used in Experiment 1. Top left: total cone contrast for the left (red) side of the stimulus. Top right: total cone contrast for the right (white) side of the stimulus. The bottom row shows an example stimulus (concave, 50 degrees) with a dashed line indicating the contour along which gradients were measured.

In the Outline-cue-only condition the stimuli consist of a pair of trapezoids, defined by lines. Previous work has shown that people perceive such shapes as slanted rectangles (Ames, 1951; W. C. Clark, Smith, & Rabe, 1955; Olson, 1974; K. A. Stevens, 1981; K. A. Stevens & Brookes, 1987; Zimmerman, Legge, & Cavanagh, 1995). The detection threshold for this type of stimuli depends on size, but for stimuli of a similar size to those used in Experiment 1, the threshold has been shown to be around 10 degrees (Freeman, 1966) – the same as the smallest slant angle used. Discrimination thresholds for such stimuli have been shown to be around 10% (Saunders & Backus, 2006). In the stimuli used in Experiment 1, the smallest change in
slant angle between consecutive stimuli was around 15% (between 60 and 70 degree stimuli), so all the Outline-cue-only stimuli should be discriminable from each other. Gradient+Outline stimuli should also be expected to be discriminable from each other as they consist only of a combination of the discriminable gradient and outline stimuli.

Because the stimuli are likely to be discriminable from each other in all the conditions (certainly those that contain an outline cue), measureable differences in observer settings between stimuli are expected if the cue information elicits a consistent percept of shape.

6.2.2 Results of the Monocular Conditions

If observers can use the information in the stimulus image to successfully determine 3D object shape, we would expect them to be able to reliably assign convex and concave shapes to the correct category. This is particularly important in the monocular conditions because the monocular cues present in the stimuli used in this experiment are inherently ambiguous, but the use of assumptions and/or prior knowledge could be used to disambiguate the cue information. We would also expect the angle set by observers to increase with stimulus angle, and might also expect observers to make settings with lower variance when the cue (or combination of cues) is more reliable (see Chapter 3).

Figure 6.3 shows the mean settings of all observers for the three monocular conditions of Experiment 1 (Gradient-cue-only, Outline-cue-only and
Gradient+Outline conditions respectively). Angles shown are in degrees and represent the angle of the card surfaces from the horizontal axis perpendicular to the observer (see Figure 5.7 – page 84). Positive angles represent a concave object (corner shape) and negative angles a convex object (roof shape).
Figure 6.3 Results of the monocular conditions of Experiment 1. Top: Gradient-cue-only; Centre: Outline-cue-only; Bottom: Gradient+Outline. Mean settings are shown for all 6 observers. Vertical axis: observer setting (degrees); Horizontal axis: stimulus angle (degrees). Error bars indicate +/- one standard deviation. The red line represents expected mean settings for veridical perception.
In all three conditions observers appear to be able to use the available cues to consistently assign 3D object shape: Perceived card angle typically increases with physical angle for all conditions; In the Outline-cue-only and Gradient+Outline conditions, data from all observers showed similar trends and all observers could assign shapes to the correct category (concave or convex; Fisher’s exact test p<0.05). Fisher’s exact test provides a means of examining the significance of the association between two categories, and is suitable for the small sample sizes used here. A p-value of less than 0.05 indicates that the two categories tested are not associated – in this case that the observers can tell the difference between concave and convex shapes. In the Gradient-cue-only condition, only four of the six observers could assign concave and convex shapes to the correct category (Fisher’s exact test; AC: p= 0.01, BM: p=0.01, LM: p=0.002, SC: p=0.01, LL: p=0.3, MD: p=0.54). The four who could use the gradient cue made settings that increased with absolute card angle.

Plots of individual observer data for the Gradient-cue-only condition are shown in Figure 6.4, and demonstrate the difference in behaviour between the four observers who could make consistent settings (AC, BM, LM, SC) and the two who could not (LL, MD).
Figure 6.4 Gradient-cue-only condition results. Data are shown for each observer separately. For each observer, data are shown for each trial. Vertical axis: observer setting (degrees); Horizontal axis: stimulus angle (degrees). The red line represents expected mean settings for veridical perception.

Observer data in all conditions show a noticeable bias towards underestimating card angle, or a perceiving the card surfaces as more frontoparallel than is the case. It can be seen from Figure 6.4 that for the four
observers who use the gradient cue well, ambiguous perception is rare: observers rarely make settings of the opposite sign to the stimulus. Settings that overestimate the card angle are also very uncommon. The data for all observers in the Outline-cue-only and Gradient+Outline conditions show the same characteristics, suggesting that the underestimation bias seen in the mean data is not an artefact created by averaging positive and negative settings. i.e. in general observers did not demonstrate ambiguity in their perception of the stimuli used in Experiment 1. This suggests that the underestimation of card angle present in the data is due to a bias in observer’s perception of the stimulus shape; a result that is consistent with other literature on slant perception, which shows that perception of slant is typically biased and flatter (more frontoparallel) surfaces are reported than those physically presented (e.g. Adams & Mamassian (2004), Bülthoff & Mallot (1988), Mitchison & Westheimer (1984), Todd et al. (2005) and van Ee et al. (2003)).
Figure 6.5 Mean observer variance at each stimulus angle. Vertical axis: variance (degrees$^2$); Horizontal axis: stimulus angle (degrees). Green squares: Gradient-cue-only condition; Red circles: Outline-cue-only condition; Blue diamonds: Gradient+Outline condition.

Figure 6.5 shows the variance of settings made by observers at each stimulus angle, and in all three conditions. Mean variance across all observers is shown, but all observers showed similar trends. Comparing variance for the two ‘single-cue’ conditions (Gradient-cue-only and Outline-cue-only), it is clear that the outline cue elicits more reliable settings, with lower variance at all stimulus angles. Mean variance for all observers and all card angles is higher for the gradient cue (315 deg$^2$) than the outline (27 deg$^2$). Variance in the Gradient+Outline condition was very similar to the
Outline-cue-only condition and Levene's test for equality of variance revealed no significant difference between the two conditions for the majority of stimulus angles. A lower variance would be expected when the two monocular cues are combined if observers integrate cue information in a statistically optimal way (see Chapter 3). However, the addition of the relatively weak gradient cue (evidenced by the high setting variance in the Gradient-cue-only condition) to the outline cue would not be expected to produce a large reduction in variance. Because this potential reduction in variance is likely too small to be seen in the data, we cannot be sure if a statistically optimal combination of cues occurs, or whether observers use a non-optimal strategy such as cue vetoing.

6.2.2.1 Comparison of perceived shape across monocular conditions

As well as comparing setting variance, it is also of interest to compare the mean settings made by observers in each condition. Figure 6.6 shows mean settings for all observers for each condition. Some subtle differences between the conditions can be seen. First, when comparing Gradient-cue-only and Outline-cue-only settings, for nearly all angles Gradient-cue-only settings are lower (more convex) than settings made using the outline cue. A two-way repeated measures ANOVA shows this to be a significant effect for the average settings made by all observers (F(1,5)=7.65; p=0.04). In the Gradient+Outline condition observers typically made settings at a greater angle than in either the Outline-cue-only or Gradient-cue-only conditions. Comparing Gradient+Outline and Outline-cue-only settings, 2-way repeated measures ANOVAs show this effect to be significant at all stimulus angles for
3 out of 6 observers, for both convex ‘roof’ and concave ‘corner’ shapes (BM (roof): F(1,9)=58.71, p<0.001; BM(corner): F(1,9)=49.87, p<0.001; LM (roof): F(1,9)=293.4, p<0.001; LM(corner): F(1,9)=348.8, p<0.001; SC (roof): F(1,9)=36.0, p<0.001; SC(corner): F(1,9)=14.41, p=0.004). While the effect was not significant for the other three observers, mean observed angle was never lower in the Gradient+Outline condition than the Outline-cue-only condition. An improvement in shape estimates (closer to veridical) when more cues are available has also been noted by other authors (e.g. Bülthoff & Mallot (1988)).

Figure 6.6 Mean observer settings at each stimulus angle. Vertical axis: observer setting (degrees); Horizontal axis: stimulus angle (degrees). Green squares: Gradient-cue-only condition; Red circles: Outline-cue-only condition; Blue diamonds: Gradient+Outline condition.
It is also interesting to compare settings made for convex and concave shapes within each condition. Figure 6.7 shows mean settings for all observers as a function of the magnitude of the stimulus angle. In the Outline-cue-only condition and the Gradient+Outline condition, no significant difference exists between the magnitude of perceived concave and convex shape (two-way repeated measures ANOVA: Outline-cue-only, $F(1,5)=1.938$, $p=0.223$; Gradient+Outline, $F(1,5)=0.280$, $p=0.619$). However, in the Gradient-cue-only case, the roof was consistently perceived as having a larger angle (steeper card with more depth) than the corner and a two-way repeated measures ANOVA revealed that this difference was significant ($F(1,5)=7.007$, $p=0.046$).
Figure 6.7 Comparison of observer settings for convex and concave versions of the stimuli.

Top: Gradient-cue-only condition; Centre: Outline-cue-only condition; Bottom: Gradient+Outline condition. Green triangles: concave ‘corner’ stimuli; Blue squares: convex ‘roof’ stimuli.
6.2.3 Results of the Binocular Conditions

The previous sections have considered settings made by observers when viewing monocular versions of the stimuli. Binocular disparity is an important cue to 3D shape and may have a significant influence on the perception of shape for our stimuli. In particular, there is a large bias towards under-estimating the angle of the stimuli when viewed monocularly, which could be reduced when stereo cues are also available. All the cues to 3D shape in the stimuli used in Experiment 1, including the stereo viewing conditions, were congruent. Therefore, if observers’ performance is close to that predicted by ‘weak-fusion’ type cue combination models (see Chapter 3 – Cue Combination), then it would be expected that shape settings would be of similar magnitude, and the underestimation bias still present, but more reliable (lower variance) when binocular disparity cues are also present.

6.2.3.1 ‘Stereo-outline’ Condition Results

Figure 6.8 shows the mean shape settings made by all observers for the binocular ‘stereo-outline’ condition, where stimuli consisted of stereo pairs of ‘outline-only’ images (i.e. stimuli contained only binocular disparity and outline perspective cues to shape/depth). Also shown in Figure 6.8, for comparison, are the mean shape settings of all observers in the ‘Outline-cue-only’ condition of experiment 1, which used monocular versions of the same stimuli.

Typically, greater stimulus angles were set for the stereo viewing arrangement than when viewing was monocular. This effect was significant for both concave and convex shapes for only 2 of the 5 observers that took
part in both experiments (observers LM and MD, two-way repeated measures ANOVA, p<0.001). However, none of the observers made lower stimulus angle settings (for either convex or concave shapes) when viewing the stereo stimuli, compared to the monocular stimuli.

Figure 6.8 Mean shape settings from all observers in the binocular ‘stereo-outline’ (blue diamonds) and the monocular ‘Outline-cue-only’ (red squares) conditions of Experiment 1.

Figure 6.9 shows mean observer variance for all observers in both the ‘stereo-outline’ condition and the ‘Outline-cue-only’ conditions. Variance is similar for settings made using the stereo and monocular versions of the stimuli, and Levene’s test for equality of variance revealed that while variance was different at some angles, no clear pattern of difference in variance between the two conditions is present.
6.2.3.2 ‘Stereo Gradient+Outline’ Condition Results

In the case of the ‘Stereo Gradient+Outline’ condition it is again interesting to compare observer settings to the monocular equivalent (the ‘Gradient+Outline’ condition). Mean shape settings made by all observers for both ‘Stereo Gradient+Outline’ and ‘Gradient+Outline’ stimuli are shown in Figure 6.10. As in the case of the ‘stereo-outline’ condition (section 6.2.3.1), the addition of the stereo information elicits a greater sense of depth in the stimuli, and correspondingly larger angle settings from observers. In the ‘Stereo Gradient+Outline’ condition this effect is significant for both convex
and concave shapes for three of the five observers, two of whom are the same observers who made significantly higher settings in the ‘stereo-outline’ condition compared to the ‘Outline-cue-only’ condition (LM, MD, SC; two-way repeated measures ANOVA, p<0.05).

Figure 6.10 Mean shape settings from all observers in the binocular ‘Stereo Gradient+Outline’ (blue diamonds) and monocular ‘Gradient+Outline’ (red squares) conditions of Experiment 1.

Comparing the variance of settings made by observers when viewing stereo and monocular versions of the Gradient+Outline stimuli (Figure 5.6, right column), again the addition of the binocular disparity cue does not appear to elicit any reduction in the variance of settings made by observers, with
Levene’s test for equality of variance revealing no general difference between the two conditions. These variance data are shown in Figure 6.11. Essentially the same effects are seen in the ‘stereo-outline’ condition results (section 7.2.1) and the ‘Stereo Gradient+Outline’ condition results, when the data are compared to that from equivalent monocular stimuli: The addition of the binocular stereo cue to the stimuli results in more veridical shape settings (greater perceived depth), but no change in the variance of settings. How these results compare with predictions made by cue combination models is discussed in section 6.3.4 (Cue Combination).

Figure 6.11 Mean observation variance for all observers in the binocular ‘Stereo Gradient+Outline’ (blue diamonds) and the monocular ‘Gradient+Outline’ (red squares) conditions of Experiment 1.
6.2.3.3 Effects of the Addition of the Gradient Cue to Stereo Stimuli.

A further interesting analysis of the data from the binocular conditions of Experiment 1 is the comparison of the settings made by observers when viewing stimuli with and without the shading cue, similar to that made in section 6.2.2.1 for results of the monocular conditions. The mean shape settings for all observers from both the ‘stereo-outline’ and ‘Stereo Gradient+Outline’ conditions are therefore plotted together for comparison in Figure 6.12, and the mean observation variance for these two conditions is plotted in Figure 6.13.

Figure 6.12 Mean shape settings from all observers in the binocular ‘Stereo Gradient+Outline’ (purple diamonds) and ‘stereo-outline’ (orange squares) conditions of Experiment 1.
As with the addition of the gradient cue in the monocular case, and the addition of the stereo cue, when a greater number of cues are present (more information about 3D structure in the stimuli), the closer to veridical the angle settings made by observers. The addition of the gradient cue when viewing is binocular results in significantly greater angle settings for observers AC (convex shapes only), LM (both concave and convex shapes), BM (concave shapes only) and MD (concave shapes only); two way repeated measures ANOVA p<0.05.

This is a similar finding to that presented in the analysis of the monocular conditions: although the increase in set angle with a greater number of cues is not always significant for both concave and convex shapes for all observers, no reductions set angle occur for any observers. Additionally, observation variance again seems to be largely unaffected by the addition of further cues, with Levene's test for equality of variance revealing no difference between variance in the ‘stereo-outline’ and ‘Stereo Gradient+Outline’ conditions for most stimulus angles.
6.3 Discussion

6.3.1 Do observer settings need correcting because of the task?

The task used in Experiment 1 requires observers to set a pair of lines in a ‘V’ configuration to match their perception of the stimulus shape. It has been shown that errors are typically present when estimating the angle between a pair of lines in this configuration, a factor that may affect the results of any experiments using this task. A number of studies have been performed to quantify these errors (Chen & Levi, 1996; Kennedy, Orbach, & Loffler, 2006; Loffler, Kennedy, Orbach, & Gordon, 2003; Nundy, Lotto, Coppola, Shimpi, &
and it is often quoted that obtuse angles are underestimated and acute angles overestimated. Because errors in estimating the angle between two lines are known to occur, it may be sensible to correct any experimental results obtained when the task requires such an estimate. The majority of the studies referenced above measured errors estimating the angles between lines in the same quadrant (typically one of the upper quadrants) or around axis of random orientation. However, one study has considered angle estimation errors for lines about a vertical axis, in an identical configuration to those used in Experiment 1 (Fisher, 1969). A re-plot of the data from this study is shown in Figure 6.14. Observers were generally quite accurate in estimating the angle of vertically orientated ‘V’ angles and the maximum errors present are around 3 degrees. While it is possible, using Fisher’s data, to correct observer settings from Experiment 1 for the additional errors due to the task, the relatively small size of the perpetual errors in comparison to the observed underestimation bias (see Figure 6.3) suggest that this correction is unnecessary in the analysis. The biases in observer settings are typically much larger than those that would be expected to arise solely from the ‘V’ angle setting task and these biases must be largely due to other factors.
Figure 6.14 Observer errors when estimating the angle between two lines in a ‘V’ formation about a vertical axis (Re-plotted from Fisher (1969)).

6.3.2 How effectively are the individual cues used?

As noted above, previous work has shown that trapezoid shapes are perceived as slanted rectangles (Ames, 1951; Clark et al., 1955; Olson, 1974; Stevens, 1981; Stevens & Brookes, 1987; Zimmerman et al., 1995), so it is not surprising that observers make consistent shape settings for the Outline-cue-only condition and can reliably distinguish concave and convex versions of the stimuli (Figures 6.2 & 6.6 centre). The Gradient-cue-only condition is, however, more complicated. Luminance and colour gradients are ambiguous cues to shape, dependent on a complex combination of object shape, material, and the lighting environment. Horizontal luminance and CIE chroma (CIE, 1976) profiles for +50 and -50 degree example Gradient-cue-only stimuli are shown in Figure 6.15. Of note is that the
gradients across the surfaces of the cards are complex and very different for concave and convex versions of the object. It is therefore interesting that observer settings suggest generally similar (absolute) perceived object shape for concave and convex versions of the card (Figure 6.7 top). This result suggests that the visual system is correctly interpreting the gradient cue as a combination of surface reflectance, lighting and object shape. Further evidence of this is provided by the similarity of perceived shape across different cues. Shape settings for the Gradient-cue-only condition and the Outline-cue-only condition were quite similar (Figure 6.6).
Figure 6.15 Example luminance and Chroma profiles for the Gradient-cue-only stimuli (horizontal profiles at half height). Top row: -50 degree (convex) stimulus; Bottom row: +50 degree (concave) stimulus. Left column: CIE 1967 \((L^*,u^*,v^*)\) Chroma; Right column: Luminance.

6.3.3 What are the Sources of Bias in Observer Settings?

One of the most striking aspects of the settings made by observers in all conditions is the tendency to underestimate the angle of the card. As noted in the results section of this chapter, this is an effect that is consistent with existing literature. An aspect of this bias that is apparent because of the large range of stimulus angles used is that observers underestimate card angle more (as a fraction of the stimulus angle) when the card is closer to flat (low
angle), resulting in the ‘reverse sigmoid’ shape of the data plots (for example Figure 6.3). The work by Fisher (1969) described above (6.3.1) suggests that perceptual errors in estimating the ‘V’ angle of two lines, as observers are required to do in the setting task, cannot account for the large underestimation of stimulus angle observed. Some previous studies have suggested that the type of task can affect estimates of surface slant because slant is only calculated locally (Zimmerman et al., 1995). The authors of this study argue that tasks that involve estimating slant for non-local surface areas are the cause for the underestimation of slant typically seen in other studies, but it is unclear how this would result in the reduction in bias seen with increasing stimulus angle here.

Another possible source of bias in the setting task could arise because the setting lines always returned to zero angle (flat) at the start of each trial. This could bias observer responses towards lower values. To mitigate this concern, an additional observer was tested using a version of the experiment where the setting lines started at a random position for each trial. The settings made by this observer for the Outline-cue-only condition are shown in Figure 6.16, and are very similar to those made by the other observers. It is clear that the underestimation bias is still present in these data, suggesting that the bias is not due to the starting position of the setting lines.
Figure 6.16 Outline-cue-only condition; mean settings for a single naive observer made with setting lines starting at a random position for each trial. Vertical axis: observer setting (degrees); Horizontal axis: stimulus angle (degrees). Error bars indicate +/- one standard deviation. The red line represents expected mean settings for veridical perception.

A further possible source of underestimation bias could be residual shape cues that indicate the true (zero) depth of the stimulus. In the monocular conditions, the effects of other cues to 3D shape were minimised by use of the viewing aperture, but some cues such as ocular accommodation cannot easily be removed. However, the tendency of observers to underestimate depth, particularly at low stimulus angles, is large and it seems unlikely, given the design of the viewing apparatus, that this can be entirely attributed to the limited residual shape cues available. In fact, because variance is typically higher at higher stimulus angles, suggesting less reliable cue information, we might expect the any underestimation bias due to residual cues to be more
apparent at higher angles, rather than when stimulus angle is small, as is the pattern in the data. An alternative explanation for this underestimation of stimulus depth is that observers integrate prior knowledge about object shapes, or surface orientations into their estimations, using a prior distribution that includes a greater frequency of flat objects than tightly folded ones. As noted in Chapter 2, it is expected that prior assumptions are influential when making depth estimates from monocular cues, since in general such assumptions are required to make use of ambiguous cue information.

Observer estimates also show biases in variance, with much higher variance at higher stimulus angles. Again, it does not seem that this can be attributed to the setting task, as reliability in estimating such ‘V’ angles has been shown to be relatively constant across the range of angles used (Fisher, 1969). This pattern of increasing variance with stimulus angle is interesting because the perspective cues in the stimuli seem potentially more reliable at higher angles. For example, it could be that the source of information in the outline cue is the angle between the left and right sides of the top (or bottom) edge - an image property that changes more rapidly with stimulus angle when the angle is larger, resulting in a reduced susceptibility to (uncorrelated) noise and thus greater reliability when the stimulus angle is larger (see Hogervorst & Eagle (1998)). This effect is demonstrated for similar stimuli to those used in Experiment 1 by van Ee et al. (2002), who present calculated likelihood distributions for slant from linear perspective that are narrower for higher surface slants. We would also expect the gradient cue to become more
reliable as the stimulus angle increases, since the rate of change (as stimulus angle changes) of contrast across both the luminance and chromatic gradients is greater when the stimulus angle is larger (see Figure 6.2, this is discussed in more detail in section 9.4.1).

The observer variance data for most of the conditions of Experiment 1 in fact show the opposite trend (no strong trend is present in the Gradient-cue-only condition), increasing at higher stimulus angles, and cue reliability appears to reduce at higher stimulus angles.

One further bias that is specific to the gradient cue is the difference between settings made for convex and concave shapes (Figure 6.7, top). As noted in the introduction to this chapter, differences between perception for concave and convex versions of the stimuli might occur because mutual illuminations are present in the concave versions only. However examination of the Gradient-cue-only condition data in Figure 6.3 (top) reveals that the difference seen in Figure 6.7 (top) is due to a negative offset in the settings, that is most strongly present at low card angles when levels of mutual illumination are very low (the card is close to a flat shape, so inter-reflections of light between the two sides are minimal). Instead, it seems likely that this effect may have occurred because of the choice of light position when rendering the stimuli, which was relatively close to the stimulus (see section 5.4.1.1). For small corner (concave) card angles, the choice of the light direction and distance resulted in a slight fall off of luminance from the centre to the edges of the card. If the visual system were using the additional assumption that darker regions of an object were further away (Langer &
Zucker, 1994; Schwartz & Sperling, 1983), this effect could account for the pattern of responses seen.

Statistical testing for equality of variance between the settings made for concave and convex versions of the Gradient-cue-only stimuli did reveal that for the two highest stimulus angles used (60 and 70 degrees), variance was significantly lower for concave shapes (Levene’s test for equality of variance, p<0.05). This suggests that, at least when stimulus angle and therefore levels of mutual illumination are high, the extra shape information provided by mutual illuminations may be used by the visual system to improve the reliability of shape estimates.

6.3.3.1 Assessing Biases Inherent in the Task

Although perceptual errors in estimating the angle of a pair of lines similar to those in the task used here appear to be too small to account for much of the underestimation bias in the shape settings made by observers (see section 6.3.1) and the bias is not introduced by the starting position of the adjustable lines of the task (see above; Figure 6.16), it is still possible that some of the shape bias is introduced by the task, perhaps in the cognitive translation that must be made between the perceived 3D shape and the line representation of the ‘overhead view’. To investigate if such an effect is a factor in the results of Experiment 1, a control experiment was undertaken that used real physical cards rather than computer rendered cards as the stimuli. These real cards were the same size and shape as the rendered cards, again one half white and one red, but also included a black grid pattern (see Figure 6.17) and the cards were viewed binocularly under similar lighting conditions.
(overhead spot lamp) to the rendered cards. Therefore, a range of good cues to the three-dimensional shape of the cards was available to observers: stimulus outline (perspective), shading (including mutual illumination), texture and binocular disparity. Observers viewed the cards from a distance of 1m, as with the rendered versions, but the position of their heads was not tightly restricted, also allowing motion parallax shape cues.

![Figure 6.17 A photograph of one of the cards used for the 'Real Card' control experiment.](image)

In this control experiment, observers made settings for the same range of card angles (3 repeats at each angle) as in the main rendered card experiments. Cards were placed in a grooved wooden block to ensure the correct angle. The experimental procedure was otherwise identical to that of
Experiment 1. Importantly, observers also used an identical ‘V’ angle adjustment task to make their shape settings (see section 6.1). 5 naïve observers took part in this control experiment (2 female). All subjects had normal colour vision and stereo acuity, and normal or corrected to normal acuity. Normal colour vision was verified using the Farnsworth-Munsell 100 hue test (score below 100). Normal stereo acuity was verified with the TNO stereo test (120 seconds of arc or less).

Because good shape information is available, we should expect the effects of any priors (for example a prior for flatter objects, or fronto-parallel surface orientation as suggested above) to be minimised. A prediction can therefore be made about the results of this ‘real card’ control experiment: if priors account for the majority of the underestimation of stimulus angle, then observers should make considerably more veridical settings with the real cards than the rendered ones, since more cue information is available in the real cards. However, if alternatively the majority of the bias is due to the task used then we should instead expect observers to make shape settings of similar magnitude with the real cards to when stimuli were rendered cards.

The shape settings made in the ‘real card’ control experiment were very similar for all observers. Therefore mean settings for all observers are shown in Figure 6.18, together for comparison with those made by observers in the ‘Gradient+Outline’ condition Experiment 1 (shown previously in Figure 6.3).
Mean shape settings made by all observers in the ‘Real Cards’ control experiment (red diamonds) and in the ‘Gradient+Outline’ condition of Experiment 1 (blue squares). The red line indicates veridical performance.

Comparison of the shape settings made in ‘real card’ control experiment to those from the ‘Gradient+Outline’ condition of Experiment 1 reveals that when more cue information is available (when using the real cards) observers make shape settings that are in general much closer to veridical. A simple linear fit to the data produces a slope of 0.82 for the ‘real card’ data vs. 0.50 for the rendered cards, and an analysis of variance confirms that the data are indeed significantly different for both concave (F(1,6)=287; p<0.01) and convex shapes (F(1,6)=198; p<0.01).
The results of the ‘real card’ control experiment show that shape perception is significantly more veridical for the ‘cue rich’ real cards than the more impoverished rendered cards, suggesting that the underestimation of stimulus angle seen in all conditions of Experiment 1 is not an artefact of the task used, but a real perceptual effect, perhaps caused by a prior for flatter shapes or fronto-parallel surface orientation. This conclusion is also supported by the similar pattern of underestimation of surface slant seen in recent work that used a very different stimulus and task (Ivanov, Kramer & Mullen, 2013). The shape settings made in the ‘real card’ control still exhibit some underestimation of stimulus angle, but if a strong prior is present, as appears to be the case, this result should be expected – while the good cue information present minimises the effect of the prior it does not remove its influence entirely, and if the prior is strong enough settings would still be biased even without any other sources of error (e.g. the task).

As a means of assessing the relative influence of the cues and a potential prior for flatter shape / fronto-parallel surface orientation, a simple model can be fitted to the data that includes a linear component representing the cue information and a Gaussian component centred on zero stimulus angle (flat), representing a prior:

\[
y = Cx \cdot \frac{e^{-\frac{x^2}{2k^2}}}{\sqrt{2\pi}}
\]

**Equation 6.2** A simple model used to compare shape settings for rendered and real cards.

- \(C\) = slope of linear component.
- \(k\) = width of Gaussian component.
A least squares fit of the simple model shown in equation 6.2 to the mean shape settings of the ‘real cards’ control and the ‘Gradient+Outline’ condition of Experiment 1 reveals a stronger linear component and weaker Gaussian in the ‘real card’ case (real cards: $C=0.96$, $k=34.9$; rendered cards: $C=0.81$, $k=57.7$), suggesting that the cue information is relatively more influential than the prior in for the real card stimuli than the rendered stimuli.

### 6.3.4 Cue Combination

As previously noted, many cue-combination theories suggest that the visual system should combine cue information optimally and that therefore a combination of two cues will result in a lower variance in settings than with either cue alone (see Chapter 3). If the visual system performs a ‘weak fusion’ of cues (Clark & Yuille, 1990), and this is done in a statistically optimal way, producing the lowest variance in the resulting estimate, then we would expect to see a reduction in the variance of observer settings when the gradient cue is also available, compared to when the outline cue is present alone (Hillis et al., 2004; Oruc et al., 2003). In the monocular conditions of Experiment 1 variance was typically very similar in the Gradient+Outline and Outline-cue-only conditions. However, it is clear from the high variance when making settings using the gradient cue alone (Figure 6.5), that the gradient cue is relatively weak in comparison to the outline and optimal cue-combination models would only predict a very small reduction in variance when the gradient cue is available in addition to the outline cue. However, because there is no significant difference in variance between the Outline-
cue-only and Gradient+Outline data, it is not clear if statistically optimal cue combination or some other form of cue-combination, such as cue vetoing occurs – i.e. observers may only use the more reliable outline cue in the Gradient+Outline condition. Statistical examination of the monocular condition data reveal that the addition of the gradient cue to the outline does have some effect on shape perception when cues are combined, and is not ignored: perceived depth is greater in the Gradient+Outline condition than either of the 'single-cue' conditions (Figure 6.6). This result cannot be explained by typical weak fusion cue-integration models that provide weighted averages of ‘single-cue’ settings, which would predict that mean shape estimates in the Gradient+Outline condition would lie in between those made in the single cue conditions. However, some previous work has found similar effects (Bülthoff, 1991; Bülthoff & Mallot, 1988). The results are also in general agreement with another previous study that has suggested that colour gradients must be combined with other cues to provide a significant contribution to depth and shape perception (Troscianko, Montagnon, Leclerc, Malbert, & Chanteau, 1991).

In the binocular conditions of Experiment 1 additional cues are available to the observers due to the binocular disparity between the left and right eye images of the stimuli. Comparing the results of the binocular conditions to those of the monocular conditions reveals that when stereo pairs of stimuli are presented to observers, rather than monocular stimuli, the perception of the 3D shape is in general quite similar (Figures 6.9, 6.10, 6.11 and 6.12). The same bias for underestimation of stimulus angle (more so for smaller
stimulus angles) and increasing variance with increasing stimulus angle that are present in the monocular condition data are also present in the binocular condition data.

In general, the situation with regard to cue-combination is very similar to that seen in the monocular conditions: when additional cues are available, larger (more veridical) angles are set by observers, while variance does not necessarily change. As noted above, this is not a result that is predicted by commonly used ‘weak fusion’ type models, and contradicts the findings of some other studies (e.g. Doorshot et al. (2001), Landy et al. (1995)). However, it should be noted that the increase in set angle effect is small and only present for some observers, while the sample size is potentially too small for changes in variance to be seen. Figure 6.19 provides a summary of the changes in mean angle set by observers in Experiment 1 as the number of cues in the stimuli is increased.
Figure 6.19 Summary of statistically significant changes in the mean angle set by observers between the various conditions of Experiment 1. Note that no reductions in mean angle set are seen for any observers when the number of cues available in the stimuli is increased.

It is perhaps surprising that variance is not reduced with the addition of the binocular disparity cues because, unlike the gradient cue, disparity is probably a comparatively reliable cue for our stimuli - other studies have shown that binocular disparity is often a reliable cue to depth (e.g. Bradshaw & Glennerster (2006), Lovell et al. (2012)). However, other authors have suggested that pictorial cues such as shading and perspective are more useful than binocular disparity for the perception of surface slant (Zimmerman et al., 1995). Partial or intermittent binocular fusion during the stereo conditions could have reduced or removed binocular cue information for some observers, resulting in little effect from adding stereo cues. The stimuli did not contain Nonius markers to aid fusion, but the sharp edges and
central fold of the stimuli are likely to provide reasonable landmarks for fusion, and the typically more veridical shape settings made when binocular disparity was present suggest that observers did successfully fuse the stereo images.

There are a number of reasons that might explain why observation variance remains unchanged when additional cues are added to the stimuli:

First, it may be that the level of variance is at a threshold determined by the task. As discussed in section 6.3.3, a minimum variance will be expected due to variance in the task of estimating the ‘V’ angle of the setting lines, but importantly, this threshold would not be expected to be very different across the range of angle settings (Fisher, 1969). In all the conditions of Experiment 1 (except the Gradient-cue-only condition) variance is much larger at high stimulus angles, suggesting that the task does not account for all the variance in the data.

Second, as noted previously, the number of observers and the number of settings made by each might be too low to enable differences in the variance between conditions to become apparent. This seems quite likely when comparing the Outline-cue-only and Gradient+Outline monocular conditions, because the changes in variance are expected to be small, but less so when adding the binocular disparity cues.

Finally, if a Bayesian approach is used to explain the results (see chapter 3 for details of Bayesian models of shape perception) then the underestimation of the stimulus angle seen in the results of all the conditions of Experiment 1
could be explained by a prior that biases perception towards flatter than veridical shapes, or surfaces that are orientated frontoparallel. If a single prior of this type is in use, its effect would be reduced the greater the reliability or number of cues available, leading to more veridical angle settings (as seen at greater stimulus angles). However, such a model would also predict lower variance with a greater number of cues – and effect that, as noted above, is not clear in the data of Experiment 1. A Bayesian approach to modelling the Experiment 1 data is investigated in more detail in Chapter 7.

6.4 Experiment 1 Summary

Experiment 1 investigated how well the human visual system is able to use the complex colour shading gradients present in realistic scenes as a cue to 3D object shape. The experiment explored how gradients are used in isolation and also in combination with outline (perspective) and binocular disparity shape cues. The results show that the visual system is able to use realistic shading gradients as a consistent cue to shape, despite the complex nature of such gradients. However, gradients were found to be a less reliable cue to shape than object outline. There is also some evidence from the difference in variance between settings made for concave and convex shapes (Gradient-cue-only condition) that mutual illuminations are taken into account when assessing object shape.

An important bias for observers to set angles depicting considerably flatter shapes than presented was seen in all conditions. However, when the
gradient cue is combined with either the outline or binocular disparity cue, or both, observers make more veridical settings and this bias is reduced. Examination of the variance in observer settings shows that the reliability of settings does not improve when more shape information from multiple cues is present in the stimuli, contradicting the predictions of some types of cue-combination model. However, it is possible that a lack of a sufficient quantity of data may make some variance effects unobservable, and more complex models may be able to explain these results.
Chapter 7- Modelling the Integration of the Gradient and Outline Cues

In order to further investigate how the visual system makes use of the 3D shape cues available in the stimuli of Experiment 1 it is useful to try to model the observer data. By testing how well models can predict the patterns and biases of the data it is possible that a greater understanding of the underlying processes involved may be achieved. Chapter 3 discusses some of the frequently used approaches to modelling depth and 3D shape perception, in particular how perception is affected when information from multiple cues is combined. The popular ‘weak fusion’ type of cue combination model assumes that there is no interaction in the visual system between separate modules processing each cue. This means that, at least in models that do not take into account any information outside of the cues themselves (e.g. Prior information – see Chapter 3 for more details), this type of model cannot predict some of the effects seen in the results of Experiment 1, such as the more veridical shape settings made in the Gradient+Outline condition. Predictions of combined-cue perception made by the often used ‘weighted average’ type of weak-fusion model arise from a relatively simple process: averaging the settings made when cues are available separately in order to predict combined-cue perception. Such models therefore make no attempt to account for any biases evident during ‘single-cue’ perception. For these reasons it is clear that modelling the results of Experiment 1 using a ‘weighted average’ weak-fusion model will not provide much insight into the
sources of perceptual bias seen in the data, or the mechanisms of cue combination in the Gradient+Outline condition. However, in the first section of this Chapter a brief account is given of the application of a ‘weighted average’ weak fusion model to the data from the monocular conditions of Experiment 1. This is provided because, while the usefulness of this model is limited, as discussed above, the apparent weighting of the cues in the Gradient+Outline condition is of some interest.

The second section of this chapter provides a more in depth modelling analysis of the Experiment 1 data, using a Bayesian approach that considers the effects of prior knowledge on perception. A Bayesian model is potentially a greater aid to understanding the processes involved in observers’ perception of shape in the stimuli, as it may be able to explain the large underestimation bias and the more veridical settings made when the cues are combined.

It is worth noting that no attempt has been made to model the binocular conditions of Experiment 1. This is because no ‘disparity-cue-only’ condition was possible, as discussed in the introduction to Chapter 6. Therefore, any attempt to model the binocular-disparity cues could only consist of allowing the parameters of a (potentially arbitrary) model to vary in order to fit the combined-cue data when disparity cues were available, an approach that cannot make any predictions and is unlikely to yield significant insight into perception.
7.1 Reliability Weighted Linear Combination Model

Averaging the perceived shape when individual cues are available in isolation provides a relatively simple approach to modelling perception when multiple cues are combined. This is a type of ‘weak-fusion’ (Clark & Yuille, 1990; Landy et al., 1995) model - no interaction between cue information is allowed (see Chapter 3 for more details). When using this approach, the relative weight given when combining estimates from each cue is usually dependent in some way on the quality of the relevant cue information. In fact, if the weightings are based on the reliability of the cue (the inverse of the variance of shape estimates made using the cue), then assuming uncorrelated and normally distributed variance, such a model produces the minimum variance estimate (Cochran, 1937) and is a Maximum Likelihood Estimator (Ernst & Banks, 2002). This ‘MLE’ model has been shown by a number of studies to predict human cue combination behaviour well in some circumstances (Ernst & Banks, 2002; Hillis et al., 2004; Lovell et al., 2012; Oruc et al., 2003), and is described below by equations 7.1 – 7.6.
\[
\mu = w_o \mu_o + w_g \mu_g \quad \text{Equation 7.1}
\]

\[
\sigma = \sqrt{\frac{\sigma_o^2 \sigma_g^2}{\sigma_o^2 + \sigma_g^2}} \quad \text{Equation 7.2}
\]

\[
r_o = \frac{1}{\sigma_o^2} \quad r_g = \frac{1}{\sigma_g^2} \quad \text{Equations 7.3 and 7.4}
\]

\[
w_c = \frac{r_o}{r_o + r_g} \quad w_g = \frac{r_g}{r_o + r_g} \quad \text{Equations 7.5 and 7.6}
\]

**Equations 7.1-7.6:**

\(\mu\) = model prediction for perceived angle in the combined cue condition. \(\mu_o, \mu_g\) = mean observer angle setting for Outline-cue-only and Gradient-cue-only conditions.

\(\sigma\) = model estimate for standard deviation in the combined cue condition. \(\sigma_o, \sigma_g\) = standard deviation of Outline-cue-only and Gradient-cue-only observations respectively.

\(r_o, r_g\) = outline cue and gradient cue reliabilities respectively. \(w_o, w_g\) = outline cue and gradient cue weightings respectively.

**7.1.1 Predictions of the Weak Fusion Model**

The ‘weighted average’ weak fusion model as described above was used, along with the observer data from the single cue conditions of Experiment 1 (Gradient-cue-only, and Outline-cue-only conditions) to predict the results of the combined cue condition. The mean data from all observers is shown together with model predictions in Figures 7.1 and 7.2. Individual observer data was also modelled and is presented later in the chapter. However,
trends in the data and predictions for the majority of observers were similar to the group data shown in this section.

The results of Experiment 1 showed that settings made by observers when the two monocular cues were combined in the outline+gradient condition were more veridical than when using either cue alone, with less underestimation of stimulus angle. It was noted in the discussion of Chapter 6 that weak fusion models that average ‘single-cue’ data to predict combined-cue perception cannot replicate this behaviour (section 6.3.4), and this is born out in the predictions of the ‘weighted average’ model shown in Figure 7.1.
7.1. While the fit of the model to the data seems reasonable (coefficient of determination, $R^2 = 0.93$), and the underestimation bias is reproduced by the model, this should be expected bearing in mind that the model estimates are based on the settings made in the Outline-cue-only and Gradient-cue-only conditions, and mean shape settings in all the monocular conditions of Experiment 1 were quite similar (see Figure 6.6). However, model predictions are always lower than the mean shape settings made by observers.

A weak fusion model with cue weightings based on cue reliability provides a statistically optimal combination of cue information, so predicted variance for combined cues is always lower than when either cue is available alone (equation 7.2). The mean variance for all observers in the Gradient+Outline condition, together with the ‘weighted average’ model predictions for variance are shown in Figure 7.2. The variance prediction of the model is not as good as the prediction of the mean angle set ($R^2=0.58$), but no systematic differences are evident between the measured and predicted variance. Although there was no significant difference between variance in the monocular Outline-cue-only and Gradient+Outline conditions of Experiment 1 (see section 6.2.2), the fact that the model predicts the Gradient+Outline condition variance reasonably well indicates that the lack of a statistically significant difference between variance in these conditions is not necessarily inconsistent with an optimal combination of cue information, because the addition of the relatively weak gradient cue to the outline cue is not expected to reduce variance very much.
Figure 7.2 Mean observer variance for all observers in the monocular Gradient+Outline condition of Experiment 1, and ‘MLE’ weighted average model (equations 7.1-7.6) predictions based on the setting made in the Outline-cue-only and Gradient-cue-only conditions. Blue dots show mean observer variance; the green dashed line indicates the prediction of the weighted average model.

7.1.2 Cue Reliability

As discussed in chapter 6, the gradient depth cue was found to be in general considerably less reliable than the outline depth cue. When the ‘weighted average’ model is used to predict Gradient+Outline behaviour relative cue weightings are found to follow a consistent pattern. The outline cue is always weighted more heavily since it is the more reliable cue, but decreases in weighting, with a corresponding increase in gradient cue weighting, with increasing stimulus angle. This pattern can be seen in Figure 7.3.
Figure 7.3 Relative cue weightings for the ‘MLE’ weighted average model when fitted to the data of all observers in the monocular conditions of Experiment 1. Stimulus angle (x) vs. cue weighting (y). Blue diamonds indicate outline weighting; the green circles indicate the gradient cue weighting.

The gradient cue might be expected to increase in reliability as the stimulus angle increases, since the gradients change more with changing stimulus angle when the stimulus angle is higher (see Figure 6.2, this is discussed in greater detail in section 9.4.1). The apparent increase in the weighting given to the gradient cue as stimulus angle increases suggests that the reliability of the gradient cue also increases relative to the outline cue. How we might expect the reliability of the outline cue to change with stimulus angle is less clear-cut, since it depends on what specific information in the outline of the object is used by observers as their cue. The angle of the ‘horizontal’ (top and bottom) edges of the card could provide the cue. This metric changes
more rapidly with changes in stimulus angle when the angle is larger; meaning cue reliability would be expected to be higher at larger stimulus angles (where the signal to noise ratio is higher). An alternative option is that observers use the relative length of the centre-fold and sides of the card. This cue changes less rapidly with changes in stimulus angle when the angle is larger, resulting in reduced reliability. The increasing variance in shape settings in the ‘Outline-cue-only’ condition of Experiment 1 as the stimulus angle increases, and the cue weighting of the ‘weighted average’ model shown above (Figure 7.3) both indicate that this second option may provide a better explanation of observer behaviour.

### 7.2 Bayesian Cue-Combination Models.

While models such as the ‘weighted average’ model used above can be successful in predicting perceived shape when information from several cues is combined, they cannot, at least explicitly, account for the influence of any factors outside the available image information. Prior knowledge may be available to the visual system about the likelihoods of certain shapes or factors affecting image cues. For example, perception of shape from shading appears to be biased towards the expectation that lighting is from above (the well known ‘light from above’ prior: Adams (2007), Adams et al. (2004), Brewster (1826), Mamassian & Goutcher (2001), Sun & Perona (1998); see section 2.1.3.2). It is often the case that such prior assumptions (or constraints) about the visible scene are required for the extraction of unambiguous information from shape cues (e.g. Adams & Mamassian (2004)). This requirement provides a powerful argument that models based
on Bayesian inference, which explicitly incorporate the effects of prior knowledge, may be the most appropriate method of modelling shape perception and cue integration. These models usually treat prior information in a similar way to image information, as if prior information was acting as an independent additional cue (Mamassian & Landy, 2001). See section 3.1.2 for more detail on the previous use of Bayesian approaches to model 3D shape perception.

Although, as noted above, it has been shown that weak fusion models performing a linear combination of individual cue estimates can often predict cue combination behaviour successfully, it should be noted that if the likelihood functions and prior distributions of a Bayesian model are Gaussian distributions, the model can in many cases be reduced to the reliability weighted linear combination (‘MLE’) used in the previous section of this chapter. This model does not explicitly encode prior information, but any priors that are present will influence depth estimates made using single cues. Thus the influence of prior information is implicitly contained in combined cue predictions made using any combination of these estimates. However, the details of any priors that are present cannot be extracted from such a model. For this reason, Bayesian models that explicitly describe the effects of prior information allow a potentially deeper analysis of underlying cue combination behaviour.

The following sections detail a range of Bayesian models fitted to the data from the monocular ‘single-cue’ conditions of Experiment 1. This enables
estimates of cue reliability and the ability to probe for possible priors or biases in perceived shape. Second, to assess how the shape cues in the stimuli are integrated by the visual system, the predictions made by cue-combination versions of these Bayesian models are compared to the Gradient+Outline condition data.

7.2.1 Bayesian Modelling of the Single Cue Data

A range of Bayesian models were constructed to model the Gradient-cue-only and Outline-cue-only data. These models each contain a function describing the likelihood of the stimulus image ($I$), given a particular stimulus angle ($\theta$), with the assumption that image information is corrupted by noise. This is known as the cue likelihood function, $p(I|\theta)$. The results of all the conditions of Experiment 1 show that observers always appear to perceive the stimuli with a lower than veridical angle (see section 6.3.3). The literature on shape perception suggests that this may be due to observers employing prior constraints that lead to a ‘flattening’ in shape, particularly for slanted surfaces (e.g. Adams & Mamassian (2004), Hillis et al. (2004), Mingolla & Todd (1986), Mitchison & Westheimer (1984), van Ee et al. (2003)). The underestimation of stimulus angle present in the Experiment 1 data can be modelled by assuming that such a prior influences observers’ settings. For this reason, the Bayesian models constricted for the ‘single-cue’ data include a prior distribution with a peak probability for a stimulus angle of zero (a flat stimulus). This ‘prior for flatness’ (or frontoparallel surface orientation) describes the probability of a particular stimulus angle, $p(\theta)$. Separate models were created for the gradient and outline cue conditions, but both included a
common prior, since it seems likely that this same ‘prior for flatness’ affects shape perception using both cues.

7.2.1.1 Posterior distribution and decision rule

For each model, the cue likelihood function and the prior for flatness are independent and combined by Bayes’ rule to give a posterior probability distribution:

\[ p(\theta | I) \propto p(\theta) p(I|\theta) \]  

Equation 7.7

The posterior distribution gives the probability of a stimulus angle ($\theta$) given the available image ($I$) and prior information. A decision rule, which selects the preferred shape from the posterior distribution, is required to determine a final model estimate of the perceived shape. The Bayesian estimator is therefore deterministic and yields only a single predicted percept for a given stimulus. However, to provide a complete model of visual perception, the model must also make predictions of the variance in estimates made by observers. How this variance arises within the visual system is not entirely clear, but it has been suggested that, within the Bayesian framework, observation variance could arise due to internal noise that causes variation in the cue likelihood function (Stocker & Simoncelli, 2006). Alternatively, others suggest that the visual system employs a ‘non-committing’ decision rule, where observers do not pick a particular angle from the posterior distribution, but instead the distribution of their settings matches the posterior distribution. This is known as probability matching (Mamassian & Landy, 2001; Mamassian et al., 2002; Wozny, Beierholm, & Shams, 2010).
The width (variance) of the posterior distribution provides a prediction of observation variance. Because the posterior distribution is assumed to match the distribution of observer settings, the mean of the posterior provides an estimate of the mean shape settings made by observers. If all the distributions of the model are Gaussian in shape this rule is also equivalent to the maximum a posteriori (MAP) rule (Yuille & Bülthoff, 1996). In the Bayesian models described below, this probability matching approach is used because it provides a way to fit the models to the ‘single cue condition’ data such that they provide a maximum likelihood estimate for the model parameters (see Section 7.2.1.2 for details).

![Figure 7.4](image)

**Figure 7.4** Example of a Bayesian model containing a single Gaussian cue likelihood (blue dot-dash line) and a single Gaussian prior distribution (red dashed line) for a 40 degree ‘single-cue’ stimulus. These functions are multiplied to give the posterior distribution (solid purple line). Typically the maximum of the posterior distribution may be taken as the model prediction for perceived shape (maximum a posteriori decision rule).
Figure 7.4 shows an example of a Bayesian model as described above, containing a Gaussian likelihood function (for single cue) centred on the stimulus angle, and a Gaussian prior distribution centred on zero degrees. Both the position and the width of the cue likelihood and prior determine the position and shape of the posterior distribution, which in turn determines the predicted perceived shape.

7.2.1.2 Fitting the Models to the Data

The Bayesian models created were fitted to observer data for both of the monocular single-cue conditions of Experiment 1 (Outline-cue-only and Gradient-cue-only) simultaneously. This was because it was assumed that the ‘prior for flatness’ discussed above would influence settings made in each condition, so a single prior must be found to best fit the data from both single-cue conditions. Fitting the models to the data was achieved via a Matlab program that used an optimisation algorithm to vary the available free parameters of the model to maximise the probability of all the observed shape settings (given the model parameters). This method requires the assumption of a ‘probability matching’ decision rule in the model (i.e. the probability of a particular shape setting being observed is equal to the posterior probability of that shape) and provides a maximum likelihood estimate of the model parameters. In this way the model is fitted to the entire data set, taking into account the distribution of the data. The model predictions for the mean shape setting and variance can therefore be compared directly to the measured mean and variance.
The optimisation algorithm was run 20 times for each set of observer data using random starting values for the model parameters and the resulting fitted parameters were averaged. This was to mitigate the possibility of finding local rather than global maxima. The models were fitted to data from each observer separately.

7.2.2 The Models

Initially three Bayesian models were constructed with differing complexity. Comparison of how well each of these models is able to describe the Experiment 1 data allows an investigation of the potential underlying processes that occur within the visual system.

7.2.2.1 Model 1: Gaussian Prior and Likelihoods

This is the simplest of the models developed and is constructed using Gaussian functions for both the cue likelihood and the prior distribution:

**Cue likelihood:**

The probability of the image, as a function of the card angle, is assumed to be described by a Gaussian function centred on the physical stimulus angle. The standard deviation of the function, $\sigma_c$, is left as a free parameter for data fitting and is allowed to vary with stimulus angle (i.e. the modelled cue reliability $(1/\sigma_c^2)$ is not constrained to be the same for all stimulus angles):
Equation 7.8 Bayesian Model 1, cue likelihood. $\theta = \text{card angle, } \theta_c = \text{physical angle of the stimulus, } \sigma_c = \text{standard deviation of likelihood function.}$

A Gaussian function as described in Equation 7.8 was used to model both the gradient and outline cues. A separate standard deviation parameter, $\sigma_c$, was fitted for each cue.

**Prior:**

The prior describes the known probability of a particular card angle and is described by a Gaussian distribution, centred on zero stimulus angle (flat stimulus). Again the standard deviation parameter, $\sigma_p$, is free to vary to obtain the best fit to the data, but unlike the cue likelihood, is the same for all stimulus angles:

$$p(\theta; \sigma_p) = \frac{1}{\sigma_p \sqrt{2\pi}} \exp \left( -\frac{\theta^2}{2\sigma_p^2} \right)$$

Equation 7.9 Bayesian model 1, prior distribution. $\theta = \text{card angle, } \sigma_p = \text{standard deviation of prior distribution}$

This model was implemented such that the prior was the same for all stimulus angles. Thus a single value for the standard deviation of the prior was found that provided a maximum likelihood estimate for the model, considering data at all stimulus angles. This approach is typical when implementing Bayesian models of perception, since prior constraints are
considered to be unaffected by the current sensory information (Beierholm, Quartz, & Shams, 2009; Mamassian et al., 2002).

This ‘Simple Gaussian’ model (Models 1; equations 7.7, 7.8 & 7.9), contains a total of 31 free parameters across the 15 stimulus angles (the standard deviation of the outline likelihood function, the standard deviation of the gradient likelihood function at each stimulus angle, and a single standard deviation for the prior distribution at all stimulus angles). Table 7.1 shows the maximum likelihood estimates of these parameters for each observer:
<table>
<thead>
<tr>
<th>Observer</th>
<th>AC</th>
<th>BM</th>
<th>LL</th>
<th>LM</th>
<th>MD</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma_p)</td>
<td>27.504</td>
<td>14.575</td>
<td>28.667</td>
<td>21.484</td>
<td>22.945</td>
<td>19.156</td>
</tr>
<tr>
<td>-70° (\sigma_o)</td>
<td>18.175</td>
<td>48.694</td>
<td>14.041</td>
<td>43.756</td>
<td>38.999</td>
<td>28.043</td>
</tr>
<tr>
<td>-60° (\sigma_o)</td>
<td>39.229</td>
<td>47.180</td>
<td>25.732</td>
<td>40.696</td>
<td>44.803</td>
<td>38.416</td>
</tr>
<tr>
<td>-50° (\sigma_o)</td>
<td>36.593</td>
<td>39.535</td>
<td>28.9</td>
<td>37.264</td>
<td>40.283</td>
<td>33.094</td>
</tr>
<tr>
<td>-40° (\sigma_o)</td>
<td>35.049</td>
<td>30.519</td>
<td>23.517</td>
<td>32.897</td>
<td>31.167</td>
<td>26.647</td>
</tr>
<tr>
<td>-30° (\sigma_o)</td>
<td>26.046</td>
<td>23.869</td>
<td>15.963</td>
<td>22.878</td>
<td>22.918</td>
<td>24.183</td>
</tr>
<tr>
<td>-20° (\sigma_o)</td>
<td>18.024</td>
<td>16.329</td>
<td>11.584</td>
<td>14.455</td>
<td>15.46</td>
<td>14.543</td>
</tr>
<tr>
<td>-10° (\sigma_o)</td>
<td>9.9991</td>
<td>10.12</td>
<td>9.4166</td>
<td>8.8132</td>
<td>7.9841</td>
<td>8.0242</td>
</tr>
<tr>
<td>0° (\sigma_o)</td>
<td>0.1</td>
<td>1.8977</td>
<td>1.7879</td>
<td>1.2645</td>
<td>1.4122</td>
<td>1.1422</td>
</tr>
<tr>
<td>10° (\sigma_o)</td>
<td>9.8196</td>
<td>10.214</td>
<td>6.524</td>
<td>7.9226</td>
<td>7.3284</td>
<td>6.2502</td>
</tr>
<tr>
<td>20° (\sigma_o)</td>
<td>18.478</td>
<td>18.498</td>
<td>15.424</td>
<td>15.396</td>
<td>14.809</td>
<td>20.817</td>
</tr>
<tr>
<td>30° (\sigma_o)</td>
<td>26.105</td>
<td>25.378</td>
<td>23.983</td>
<td>23.062</td>
<td>22.23</td>
<td>20.817</td>
</tr>
<tr>
<td>40° (\sigma_o)</td>
<td>34.635</td>
<td>31.838</td>
<td>22.074</td>
<td>32.023</td>
<td>32.129</td>
<td>24.404</td>
</tr>
<tr>
<td>50° (\sigma_o)</td>
<td>40.589</td>
<td>40.562</td>
<td>26</td>
<td>39.457</td>
<td>37.148</td>
<td>36.064</td>
</tr>
<tr>
<td>60° (\sigma_o)</td>
<td>42.473</td>
<td>44.083</td>
<td>26.699</td>
<td>43.981</td>
<td>43.749</td>
<td>37.053</td>
</tr>
<tr>
<td>70° (\sigma_o)</td>
<td>22.468</td>
<td>37.691</td>
<td>58.316</td>
<td>8.0965</td>
<td>57.523</td>
<td>34.244</td>
</tr>
<tr>
<td>-70° (\sigma_g)</td>
<td>41.285</td>
<td>61.293</td>
<td>24.687</td>
<td>42.486</td>
<td>48.758</td>
<td>53.632</td>
</tr>
<tr>
<td>-60° (\sigma_g)</td>
<td>21.836</td>
<td>35.937</td>
<td>38.42</td>
<td>33.253</td>
<td>53.619</td>
<td>40.473</td>
</tr>
<tr>
<td>-50° (\sigma_g)</td>
<td>24.177</td>
<td>35.292</td>
<td>56.529</td>
<td>23.935</td>
<td>37.198</td>
<td>29.707</td>
</tr>
<tr>
<td>-40° (\sigma_g)</td>
<td>21.736</td>
<td>24.197</td>
<td>41.948</td>
<td>17.934</td>
<td>33.541</td>
<td>23.052</td>
</tr>
<tr>
<td>-30° (\sigma_g)</td>
<td>18.107</td>
<td>17.609</td>
<td>26.934</td>
<td>20.61</td>
<td>34.298</td>
<td>12.696</td>
</tr>
<tr>
<td>-20° (\sigma_g)</td>
<td>15.699</td>
<td>7.1248</td>
<td>21.42</td>
<td>11.922</td>
<td>26.759</td>
<td>3.1585</td>
</tr>
<tr>
<td>-10° (\sigma_g)</td>
<td>10.507</td>
<td>5.3265</td>
<td>26.603</td>
<td>5.7056</td>
<td>19.965</td>
<td>10.187</td>
</tr>
<tr>
<td>0° (\sigma_g)</td>
<td>21.623</td>
<td>11.715</td>
<td>30.517</td>
<td>10.407</td>
<td>37.078</td>
<td>16.531</td>
</tr>
<tr>
<td>10° (\sigma_g)</td>
<td>25.743</td>
<td>21.533</td>
<td>42.514</td>
<td>20.697</td>
<td>35.335</td>
<td>22.431</td>
</tr>
<tr>
<td>20° (\sigma_g)</td>
<td>29.335</td>
<td>28.644</td>
<td>42.61</td>
<td>28.599</td>
<td>45.463</td>
<td>30.297</td>
</tr>
<tr>
<td>30° (\sigma_g)</td>
<td>30.895</td>
<td>28.523</td>
<td>49.159</td>
<td>37.177</td>
<td>59.509</td>
<td>28.546</td>
</tr>
<tr>
<td>40° (\sigma_g)</td>
<td>34.457</td>
<td>34.402</td>
<td>57.05</td>
<td>46.346</td>
<td>60.976</td>
<td>35.493</td>
</tr>
<tr>
<td>50° (\sigma_g)</td>
<td>23.614</td>
<td>38.682</td>
<td>31.628</td>
<td>47.073</td>
<td>83.902</td>
<td>44.564</td>
</tr>
<tr>
<td>60° (\sigma_g)</td>
<td>30.573</td>
<td>40.353</td>
<td>35.943</td>
<td>29.176</td>
<td>65.229</td>
<td>34.18</td>
</tr>
<tr>
<td>70° (\sigma_g)</td>
<td>29.176</td>
<td>35.943</td>
<td>30.573</td>
<td>40.353</td>
<td>65.229</td>
<td>34.18</td>
</tr>
</tbody>
</table>

Table 7.1 Maximum likelihood estimates of the parameters of the ‘Simple Gaussian’ Bayesian Model (Model 1). \(\sigma_p\) = prior distribution standard deviation; \(\sigma_o\) = outline function standard deviation; \(\sigma_g\) = gradient function standard deviation;

For all observers, the fitted parameters show similar trends: Outline cue likelihood functions are generally narrower for smaller stimulus angles, indicating greater cue reliability when the angle is small (the card is flatter). A similar, but far less strong trend is seen for the gradient cue likelihoods. These trends are in keeping with what might be expected from the variance.
of observer settings seen the Outline-cue-only and Gradient-cue-only conditions of experiment 1 (see Figure 6.5).

7.2.2.2 Model 2: Non-Gaussian Prior

The data of Experiment 1 shows a large underestimation bias but this reduces as stimulus angle increases. Replicating this trend in a Bayesian model requires increasing the relative influence of the cue (reducing the influence of the prior) at higher stimulus angles. This can be achieved by using a non-Gaussian shape for the prior distribution that includes a greater probability of higher stimulus angles (i.e. a ‘heavy tailed’ prior distribution). Therefore, a ‘Heavy-Tailed Prior’ model was created where the Gaussian prior distribution used in Model 1 (‘Simple Gaussian’) was replaced with a location-scale version of Student’s t-distribution (Bishop, 2006):

\[
p(\theta; \sigma_p, \nu) = \frac{1}{\sigma_p \sqrt{\nu \pi}} \left[ 1 + \frac{(\theta - \theta_0)^2}{\nu \sigma_p^2} \right]^{-\frac{\nu + 1}{2}}
\]

Equation 7.10 Bayesian Model 2 prior distribution. \( \theta = \) card angle, \( \sigma_p = \) scale of prior distribution, \( \nu = \) shape parameter. \( \Gamma \) is the Gamma function.

Like the Gaussian prior distribution used for Model 1, this distribution was centred on zero degrees (a flat object) and a scale/width parameter was varied during the fitting procedure to find the correct ‘weight’ for the prior. However, the t-distribution contains one additional free parameter that
describes how heavy tailed the distribution is ($\nu$). When this parameter is small the prior distribution is very heavy tailed, meaning that there is a higher probability of a larger angle occurring than would be found with a Gaussian distribution; when the $\nu$ parameter is large, the t-distribution approximates a Gaussian. Cue likelihoods take the same form as in Model 1, and the ‘Heavy-Tailed Prior’ model contains 32 free parameters across the 15 stimulus angles. Table 7.2 shows the maximum likelihood estimates of these parameters for each observer:
<table>
<thead>
<tr>
<th>Observer</th>
<th>AC</th>
<th>BM</th>
<th>LL</th>
<th>LM</th>
<th>MD</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_p$</td>
<td>12.122</td>
<td>13.382</td>
<td>28.37</td>
<td>10.425</td>
<td>20.883</td>
<td>17.313</td>
</tr>
<tr>
<td>$\nu_p$</td>
<td>1.4393</td>
<td>12.125</td>
<td>73.256</td>
<td>1.8988</td>
<td>10.579</td>
<td>10.584</td>
</tr>
<tr>
<td>$-70^\circ \sigma_o$</td>
<td>18.138</td>
<td>48.614</td>
<td>14.045</td>
<td>43.539</td>
<td>39.996</td>
<td>28.028</td>
</tr>
<tr>
<td>$-60^\circ \sigma_o$</td>
<td>34.957</td>
<td>47.157</td>
<td>25.718</td>
<td>40.428</td>
<td>44.848</td>
<td>38.289</td>
</tr>
<tr>
<td>$-50^\circ \sigma_o$</td>
<td>33.279</td>
<td>39.574</td>
<td>28.871</td>
<td>37.052</td>
<td>40.507</td>
<td>33.12</td>
</tr>
<tr>
<td>$-40^\circ \sigma_o$</td>
<td>32.35</td>
<td>30.513</td>
<td>23.52</td>
<td>32.853</td>
<td>31.442</td>
<td>26.553</td>
</tr>
<tr>
<td>$-30^\circ \sigma_o$</td>
<td>25.407</td>
<td>23.864</td>
<td>15.95</td>
<td>22.844</td>
<td>23.71</td>
<td>24.137</td>
</tr>
<tr>
<td>$-20^\circ \sigma_o$</td>
<td>17.813</td>
<td>16.332</td>
<td>11.592</td>
<td>14.523</td>
<td>15.119</td>
<td>14.118</td>
</tr>
<tr>
<td>$-10^\circ \sigma_o$</td>
<td>9.9568</td>
<td>10.122</td>
<td>9.3964</td>
<td>8.8201</td>
<td>8.026</td>
<td>8.026</td>
</tr>
<tr>
<td>$0^\circ \sigma_o$</td>
<td>0.1</td>
<td>1.8981</td>
<td>1.7895</td>
<td>1.2643</td>
<td>0.1</td>
<td>1.4118</td>
</tr>
<tr>
<td>$10^\circ \sigma_o$</td>
<td>9.76</td>
<td>10.22</td>
<td>6.5101</td>
<td>7.9212</td>
<td>7.5761</td>
<td>6.2957</td>
</tr>
<tr>
<td>$20^\circ \sigma_o$</td>
<td>18.516</td>
<td>18.52</td>
<td>15.404</td>
<td>15.394</td>
<td>14.404</td>
<td>13.178</td>
</tr>
<tr>
<td>$30^\circ \sigma_o$</td>
<td>25.329</td>
<td>25.397</td>
<td>23.988</td>
<td>23.023</td>
<td>22.645</td>
<td>20.805</td>
</tr>
<tr>
<td>$40^\circ \sigma_o$</td>
<td>32.219</td>
<td>31.846</td>
<td>22.046</td>
<td>32.251</td>
<td>32.659</td>
<td>24.461</td>
</tr>
<tr>
<td>$50^\circ \sigma_o$</td>
<td>36.166</td>
<td>40.537</td>
<td>25.971</td>
<td>39.584</td>
<td>37.733</td>
<td>35.982</td>
</tr>
<tr>
<td>$60^\circ \sigma_o$</td>
<td>37.403</td>
<td>44.203</td>
<td>26.683</td>
<td>44.089</td>
<td>44.711</td>
<td>36.997</td>
</tr>
<tr>
<td>$70^\circ \sigma_o$</td>
<td>21.338</td>
<td>43.351</td>
<td>16.207</td>
<td>42.334</td>
<td>42.243</td>
<td>34.019</td>
</tr>
<tr>
<td>$-70^\circ \sigma_g$</td>
<td>22.433</td>
<td>37.704</td>
<td>58.203</td>
<td>8.1245</td>
<td>57.501</td>
<td>52.501</td>
</tr>
<tr>
<td>$-60^\circ \sigma_g$</td>
<td>24.003</td>
<td>41.296</td>
<td>61.274</td>
<td>24.784</td>
<td>40.202</td>
<td>48.112</td>
</tr>
<tr>
<td>$-50^\circ \sigma_g$</td>
<td>21.727</td>
<td>35.934</td>
<td>38.333</td>
<td>33.083</td>
<td>52.832</td>
<td>40.321</td>
</tr>
<tr>
<td>$-40^\circ \sigma_g$</td>
<td>23.557</td>
<td>35.302</td>
<td>56.517</td>
<td>23.946</td>
<td>36.453</td>
<td>29.773</td>
</tr>
<tr>
<td>$-30^\circ \sigma_g$</td>
<td>17.3</td>
<td>24.193</td>
<td>41.898</td>
<td>17.999</td>
<td>34.083</td>
<td>23.04</td>
</tr>
<tr>
<td>$-20^\circ \sigma_g$</td>
<td>17.85</td>
<td>17.611</td>
<td>26.925</td>
<td>20.603</td>
<td>32.869</td>
<td>12.721</td>
</tr>
<tr>
<td>$-10^\circ \sigma_g$</td>
<td>15.484</td>
<td>7.12</td>
<td>21.429</td>
<td>11.947</td>
<td>27.132</td>
<td>3.1557</td>
</tr>
<tr>
<td>$0^\circ \sigma_g$</td>
<td>10.344</td>
<td>5.3274</td>
<td>26.55</td>
<td>5.7215</td>
<td>20.816</td>
<td>10.226</td>
</tr>
<tr>
<td>$10^\circ \sigma_g$</td>
<td>21.411</td>
<td>11.708</td>
<td>30.514</td>
<td>10.437</td>
<td>37.784</td>
<td>16.487</td>
</tr>
<tr>
<td>$20^\circ \sigma_g$</td>
<td>24.885</td>
<td>21.531</td>
<td>42.488</td>
<td>20.694</td>
<td>34.829</td>
<td>22.436</td>
</tr>
<tr>
<td>$30^\circ \sigma_g$</td>
<td>27.989</td>
<td>28.639</td>
<td>42.567</td>
<td>27.888</td>
<td>43.552</td>
<td>30.248</td>
</tr>
<tr>
<td>$40^\circ \sigma_g$</td>
<td>29.593</td>
<td>28.521</td>
<td>49.135</td>
<td>37.116</td>
<td>60.83</td>
<td>28.489</td>
</tr>
<tr>
<td>$50^\circ \sigma_g$</td>
<td>31.636</td>
<td>34.39</td>
<td>57.016</td>
<td>45.806</td>
<td>59.682</td>
<td>35.223</td>
</tr>
<tr>
<td>$60^\circ \sigma_g$</td>
<td>23.216</td>
<td>38.683</td>
<td>31.579</td>
<td>46.487</td>
<td>74.017</td>
<td>44.278</td>
</tr>
<tr>
<td>$70^\circ \sigma_g$</td>
<td>29.354</td>
<td>40.325</td>
<td>35.947</td>
<td>29.1</td>
<td>62.794</td>
<td>34.181</td>
</tr>
</tbody>
</table>

**Table 7.2** Maximum likelihood estimates of the parameters of the ‘Heavy Tailed Prior’ Bayesian Model (Model 2). $\sigma_p$ = prior distribution scale; $\nu_p$ = prior distribution tail weight; $\sigma_o$ = outline function standard deviation; $\sigma_g$ = gradient function standard deviation.

The trends seen in the cue likelihood parameters of Model 2 are similar to those seen for Model 1. For two of the six observers (AC and LM), the prior is very heavy tailed. Only one observer has a fitted prior that is almost Gaussian in shape without significant heavy tails (LL).
7.2.2.3 Model 3: Non-Gaussian Likelihoods

While it is possible that the priors used by the observers who took part in Experiment 1 are heavy tailed, it could be argued that it is more reasonable to expect that the likelihood functions of the model could be non-Gaussian in shape. For example, if cue reliability changes with stimulus angle (which is likely due to changes in the signal to noise ratio - see Hogervorst & Eagle (1998)), then we would expect likelihood functions to be skewed. In fact we can calculate how the likelihood functions should be skewed for the outline cue, based on the geometry of the 2D projected image. To do this we must choose a particular property of the stimulus outline that serves as the true cue in the image. Several options exist – one is the angle of the ‘horizontal’ edges at the top and bottom of the stimulus. However, this cue would be most informative at higher stimulus angles (since the larger the stimulus angle, the greater the change in this cue with changing angle), meaning observer variance would be expected to decrease with increasing stimulus angle – the opposite trend to that seen in the ‘Outline-cue-only’ data. A potentially better option for the true cue within the outline stimulus is the relative length of the central fold and the left and right edges of the stimulus. The amount this factor changes in response to changing stimulus angle is large when the stimulus angle is small, decreasing as the angle increases. This should result in decreased cue reliability at higher stimulus angles, and therefore greater observation variance – exactly what is seen in the ‘Outline-cue-only’ data. Therefore, a ‘Relative-Side-Length’ model was created where the outline cue likelihood function was represented by a Gaussian function, based not on
stimulus angle (as in Model 1) but the relative length of the central fold and left/right edges in the stimulus:

\[ p(I|\theta; \sigma_o) = \frac{1}{\sigma_o \sqrt{2\pi}} \exp \left[ -\frac{(R-R_0)^2}{2\sigma_o^2} \right] \]

**Equation 7.11** Bayesian Model 3, outline cue likelihood. \( R \) = relative length of central fold and left/right edge, \( R_0 \) = relative length of central fold and left/right edge of the physical stimulus, \( \sigma_o \) = standard deviation of likelihood function.

All other aspects of this ‘Relative-Side-Length’ model were the same as ‘Simple Gaussian’ Model 1. The prior and gradient cue likelihoods were represented by Gaussian functions based on the stimulus angle (Equations 7.8 and 7.9). Again, likelihood functions were allowed to vary in width for each stimulus angle so the model contained 31 free parameters across the 15 stimulus angles. Table 7.3 shows the maximum likelihood estimates of these parameters for each observer:
### Table 7.3

<table>
<thead>
<tr>
<th>Observer</th>
<th>σ_p</th>
<th>BM</th>
<th>LL</th>
<th>LM</th>
<th>MD</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>-70° σ_p</td>
<td>0.13514</td>
<td>0.052006</td>
<td>0.010246</td>
<td>0.044493</td>
<td>0.039068</td>
<td>0.024344</td>
</tr>
<tr>
<td>-60° σ_o</td>
<td>0.045898</td>
<td>0.059667</td>
<td>0.028368</td>
<td>0.047941</td>
<td>0.055343</td>
<td>0.044838</td>
</tr>
<tr>
<td>-50° σ_o</td>
<td>0.048789</td>
<td>0.053414</td>
<td>0.036954</td>
<td>0.049717</td>
<td>0.055328</td>
<td>0.045006</td>
</tr>
<tr>
<td>-40° σ_o</td>
<td>0.051867</td>
<td>0.044219</td>
<td>0.03346</td>
<td>0.048215</td>
<td>0.055328</td>
<td>0.03796</td>
</tr>
<tr>
<td>-30° σ_o</td>
<td>0.040789</td>
<td>0.037089</td>
<td>0.024119</td>
<td>0.035463</td>
<td>0.037054</td>
<td>0.037771</td>
</tr>
<tr>
<td>-20° σ_o</td>
<td>0.029637</td>
<td>0.026746</td>
<td>0.018955</td>
<td>0.023555</td>
<td>0.024934</td>
<td>0.023764</td>
</tr>
<tr>
<td>-10° σ_o</td>
<td>0.017068</td>
<td>0.017289</td>
<td>0.016028</td>
<td>0.015051</td>
<td>0.013286</td>
<td>0.013652</td>
</tr>
<tr>
<td>0° σ_o</td>
<td>0.001001</td>
<td>0.003308</td>
<td>0.0031415</td>
<td>0.0022214</td>
<td>0.001001</td>
<td>0.0024512</td>
</tr>
<tr>
<td>10° σ_o</td>
<td>0.017353</td>
<td>0.01805</td>
<td>0.011522</td>
<td>0.01403</td>
<td>0.013539</td>
<td>0.011093</td>
</tr>
<tr>
<td>20° σ_o</td>
<td>0.032749</td>
<td>0.032736</td>
<td>0.027309</td>
<td>0.02734</td>
<td>0.025693</td>
<td>0.036402</td>
</tr>
<tr>
<td>30° σ_o</td>
<td>0.045786</td>
<td>0.044468</td>
<td>0.042037</td>
<td>0.040393</td>
<td>0.040513</td>
<td>0.037927</td>
</tr>
<tr>
<td>40° σ_o</td>
<td>0.059255</td>
<td>0.054303</td>
<td>0.037048</td>
<td>0.054988</td>
<td>0.055822</td>
<td>0.041341</td>
</tr>
<tr>
<td>50° σ_o</td>
<td>0.066357</td>
<td>0.066273</td>
<td>0.064755</td>
<td>0.062305</td>
<td>0.058291</td>
<td>0.058291</td>
</tr>
<tr>
<td>60° σ_o</td>
<td>0.063381</td>
<td>0.066877</td>
<td>0.06708</td>
<td>0.06854</td>
<td>0.054534</td>
<td>0.04136</td>
</tr>
<tr>
<td>70° σ_o</td>
<td>0.025277</td>
<td>0.056891</td>
<td>0.017961</td>
<td>0.055181</td>
<td>0.057394</td>
<td>0.043136</td>
</tr>
<tr>
<td>-70° σ_o</td>
<td>22.471</td>
<td>37.674</td>
<td>58.236</td>
<td>8.1239</td>
<td>59.11</td>
<td>53.748</td>
</tr>
<tr>
<td>-60° σ_o</td>
<td>24.701</td>
<td>41.307</td>
<td>61.289</td>
<td>24.728</td>
<td>40.443</td>
<td>48.817</td>
</tr>
<tr>
<td>-50° σ_o</td>
<td>21.836</td>
<td>35.934</td>
<td>38.377</td>
<td>33.23</td>
<td>54.08</td>
<td>40.539</td>
</tr>
<tr>
<td>-40° σ_o</td>
<td>24.182</td>
<td>35.272</td>
<td>56.533</td>
<td>23.956</td>
<td>37.082</td>
<td>29.77</td>
</tr>
<tr>
<td>-30° σ_o</td>
<td>17.355</td>
<td>24.185</td>
<td>41.901</td>
<td>17.974</td>
<td>34.56</td>
<td>23.056</td>
</tr>
<tr>
<td>-20° σ_o</td>
<td>18.116</td>
<td>17.61</td>
<td>26.905</td>
<td>20.628</td>
<td>34.203</td>
<td>12.695</td>
</tr>
<tr>
<td>-10° σ_o</td>
<td>15.696</td>
<td>7.1278</td>
<td>21.438</td>
<td>11.95</td>
<td>27.167</td>
<td>3.1626</td>
</tr>
<tr>
<td>0° σ_o</td>
<td>10.506</td>
<td>5.3299</td>
<td>26.571</td>
<td>5.7259</td>
<td>20.987</td>
<td>10.217</td>
</tr>
<tr>
<td>10° σ_o</td>
<td>21.612</td>
<td>11.716</td>
<td>30.503</td>
<td>10.471</td>
<td>37.81</td>
<td>16.494</td>
</tr>
<tr>
<td>20° σ_o</td>
<td>25.73</td>
<td>21.523</td>
<td>42.499</td>
<td>20.717</td>
<td>34.735</td>
<td>22.45</td>
</tr>
<tr>
<td>30° σ_o</td>
<td>29.333</td>
<td>28.644</td>
<td>42.598</td>
<td>28.726</td>
<td>45.375</td>
<td>30.328</td>
</tr>
<tr>
<td>40° σ_o</td>
<td>30.896</td>
<td>28.518</td>
<td>49.15</td>
<td>37.168</td>
<td>62.571</td>
<td>28.572</td>
</tr>
<tr>
<td>50° σ_o</td>
<td>34.411</td>
<td>34.388</td>
<td>57.035</td>
<td>46.212</td>
<td>62.865</td>
<td>35.519</td>
</tr>
<tr>
<td>60° σ_o</td>
<td>23.613</td>
<td>38.674</td>
<td>31.623</td>
<td>47.136</td>
<td>89.008</td>
<td>44.573</td>
</tr>
<tr>
<td>70° σ_o</td>
<td>30.574</td>
<td>40.307</td>
<td>35.955</td>
<td>29.23</td>
<td>67.001</td>
<td>34.192</td>
</tr>
</tbody>
</table>

The numerical value of $\sigma_o$ is much smaller when using equation 7.11 for the outline cue likelihood rather than a Gaussian function, so the numbers cannot be directly compared. However, the trend for smaller values at smaller angles is typically not as strong as in Models 1 and 2 (with the exception of zero degrees where the outline cue seems to be unusually reliable). This is
expected because the different form of the ‘relative-side-length’ cue likelihood function makes an increase in cue reliability for smaller angles inherent without any change in $\sigma_o$.

Figure 7.5 shows the outline cue likelihood functions (Equation 7.11) for an example observer (BM) when constructed with the maximum likelihood estimates of $\sigma_o$ obtained using the single-cue condition data of Experiment 1, as described above in section 7.2.1.2. These fitted likelihood functions exhibit large skews for larger stimulus angles. The direction of this skew, together with a widening of the distributions at higher angles, is consistent with a decrease in cue reliability for higher angles. Note that the slight asymmetry is because the central fold of the card was fixed at 1m distance, so the projected relative length of the sides and central fold is slightly different for convex (sides further away form observer) and concave (sides closer to observer) shapes.

The likelihood function for an angle of zero degrees (flat stimulus) is conspicuously narrow compared to those for -10 and +10 degrees for most observers (see Table 7.3). This suggests the completely flat stimulus is unusually reliable. It is possible that this is an artefact of the stimuli used, specifically the lack of any aliasing on the horizontal lines of the zero degree Outline-cue-only stimulus – a property unique to this angle of stimulus.
Figure 7.5. Outline cue likelihood functions of the ‘Relative-Side-Length’ Bayesian model, for an example observer (BM) when constructed with the maximum likelihood estimates of $\sigma_o$ obtained using the single-cue condition data of Experiment 1.

7.2.2.4 Model Likelihood

The parameters shown in Tables 7.1, 7.2 and 7.3 are maximum likelihood estimates for each of the three Bayesian models previously described. Therefore, a comparison of the quality of these models can be made by comparing the likelihood of each model with these parameters. The model likelihood is equal to the probability of the observed data under the model posterior distribution. By summing the negative log probability of each data point the total (negative log) likelihood of the model can be calculated. In this case, a lower number means a better model. Figure 7.6 shows the Negative Log Likelihood of each of the three Bayesian models described above:
The heavy tailed prior that is used in Model 2 only provides only a small improvement in likelihood for some observers. This improvement is largest for observers AC and LM. The fitted prior distributions of Model 2 for these two observers are also those with the heaviest tails. The ‘relative-side-length’ outline cue likelihood function of Model 3 has a much larger effect on model likelihood, resulting in a large reduction in negative log likelihood for all observers. This suggests that this model of the outline cue (Equation 7.11) provides a much better description of the visual process than the Gaussian function used in Models 1 and 2 (equation 7.8).
7.2.2.5 Model 4: ‘Best Guess’ Simplified Model

The models described above all contain a large number of free parameters since the cue likelihood functions are allowed to vary in width for each stimulus angle separately. A simplified model with fewer parameters is therefore desirable. In order to create such a model some analysis of how the previously described models fit the data is useful:

First, it is apparent that in general model parameters are fairly symmetrical about zero degrees stimulus angle, a result that is not surprising since the data for both 'single-cue' conditions of Experiment 1 are also largely symmetrical. Therefore a simplified model could use the same likelihood functions for both convex and concave shapes with little loss of power.

Secondly, the model likelihood values shown in Figure 7.6 indicate that using a heavy tailed prior distribution, and particularly an outline cue likelihood function based on the relative length of the central fold and stimulus edges, increases the likelihood of the model. This suggests that both these changes provide a better model than using Gaussian functions. Although the effect of the heavy tailed prior is small when comparing Models 1 and 2, priors with heavier tails than those of Model 2 are likely to provide an improved fit when the 'relative-side-length' outline cue likelihood is included in the model:

Because this particular likelihood function reduces the reliability of the outline cue at larger stimulus angles, but observer shape settings become more veridical (Figure 6.3), so the effect of the prior must be further reduced at higher stimulus angles if the model is to fit the data well.

In addition, the fitted parameters of Model 3 (Table 7.3) suggest that the change in shape of the outline cue likelihood function with stimulus angle can
account for most of the change in observation variance with changing angle: the width parameter of both the outline cue and gradient cue likelihood functions are similar for most angles (except for close to zero angle / flat). For these reasons, a ‘best guess’ simplified Bayesian model was created that consisted of:

- A single heavy tailed prior distribution (see equation 7.10) that was allowed to vary in width and tail weight to best fit the data.
- A single Gaussian cue likelihood function (see equation 7.8) to describe the Gradient cue at all stimulus angles, which was allowed to vary in width to best fit the data.
- A single cue likelihood function based on the relative length of the central fold and the left/right sides of the stimulus (see equation 7.11) to describe the Outline cue at all stimulus angles, which was allowed to vary in width to best fit the data.

The resulting model therefore contained only 4 free parameters across all 15 stimulus angles and both Gradient and Outline cue conditions. Table 7.4 shows the maximum likelihood estimates of these parameters for each observer:
### Table 7.4
Maximum likelihood estimates of the parameters of the ‘Best Guess’ Simplified Bayesian Model (Model 4).

<table>
<thead>
<tr>
<th></th>
<th>AC</th>
<th>BM</th>
<th>LL</th>
<th>LM</th>
<th>MD</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_p$</td>
<td>12.11</td>
<td>13.34</td>
<td>28.495</td>
<td>10.468</td>
<td>20.857</td>
<td>17.382</td>
</tr>
<tr>
<td>$\nu_p$</td>
<td>1.4369</td>
<td>11.611</td>
<td>127.38</td>
<td>1.9072</td>
<td>10.319</td>
<td>10.91</td>
</tr>
<tr>
<td>$\sigma_o$</td>
<td>0.041871</td>
<td>0.046115</td>
<td>0.02828</td>
<td>0.043517</td>
<td>0.044274</td>
<td>0.036569</td>
</tr>
<tr>
<td>$\sigma_g$</td>
<td>24.192</td>
<td>29.643</td>
<td>43.181</td>
<td>27.363</td>
<td>50.293</td>
<td>32.186</td>
</tr>
</tbody>
</table>

As with Model 2, fitted priors are very heavy tailed for observers AC and LM, and not appreciably heavy tailed for observer LL. While numerical values for $\sigma_o$ and $\sigma_g$ are not easily compared, outline and gradient cue functions are typically not very different in width (if fact this is the case for all models except for a stimulus angle of zero in Models 1, 2 and 3, where the outline cue appears to be more reliable). This is an interesting aspect of the models, since variance in shape settings was much lower for the Outline-cue-only condition of Experiment 1 than in the Gradient-cue-only condition (Figure 6.5). This could indicate that the use of a single, identical, prior for both conditions causes the models to underestimate the reliability of the outline cue and overestimate the reliability of the gradient cue.

Figure 7.7a shows the Negative Log Likelihood of Model 4, alongside that of Models 1, 2 and 3 (also shown in Figure 7.6). For all observers, Model 4 has a higher likelihood than Models 2 and 3, but is slightly less likely than model 3 (the ‘relative-side-length’ model).
Figure 7.7 (a) Negative Log Likelihood for all four Bayesian models describing the data from the ‘single-cue’ conditions of Experiment 1. (b) Akaike Information Criterion for the same models. Likelihoods/AIC for each model and each observer are shown separately.
However, as previously noted, Model 4 has far fewer parameters than the other models. The Akaike Information Criterion (AIC) allows comparison of model likelihood, taking into account the number of free parameters (Akaike, 1974). Figure 7.7b shows AIC values for the four models. Model 4 has a much lower AIC value than any of the other models (for all observers) and therefore is the preferred model, despite the slightly lower likelihood than Model 3.

The Bayesian modelling described above provides an insight into the possible mechanisms involved in the perception of 3D shape for the stimuli of Experiment 1, something the MLE model described in section 7.1 is unable to offer. The modelling shows that for this type of model to fit the data well, the influence of the prior must be reduced at higher stimulus angles, and that this can be achieved by the use of either a non-Gaussian prior (Model 2), or non-Gaussian likelihood functions (note: the particular likelihood functions used for the outline cue in the ‘Relative-Side-Length’ Model 3 actually have the opposite effect – the effect of the prior is increased at higher angles, but the model provides an better fit overall because it is able to replicate the change in variance seen in the data). The following section addresses the question of whether these Bayesian models can also provide a good account of perceived 3D shape when stimuli contain combined outline and gradient cues.
7.2.3 Predicting the Combined-Cue Data

Combined-cue versions of the four Bayesian models detailed in section 7.2.2 were constructed to predict the outcome of the monocular combined-cue ‘Gradient+Outline’ condition of Experiment 1. These combined cue versions used all the components of both the ‘Outline-cue-only’ and ‘Gradient-cue-only’ models, combined by Bayes’ rule. Equation 7.12 shows the form of the combined-cue model, built from the components described above in section 7.2.2:

\[
p(\theta|I) \propto p(\theta)p(g|\theta)p(o|\theta) \quad \text{Equation 7.12}
\]

In equation 7.12, only a single term for the prior is present since it the same prior is used in both gradient and outline ‘single-cue’ models. The combined-cue models use the parameters already obtained from fitting the single-cue models to the single-cue condition data in the previous section, and provide a prediction of observers’ perceived shape in the combined-cue ‘Gradient+Outline’ condition of Experiment 1. Again a ‘non-committing’ (probability matching) decision rule is assumed (Mamassian & Landy, 2001; Mamassian et al., 2002; Wozny et al., 2010). The mean and variance of the posterior distribution therefore provide the predictions for observer settings.

7.2.3.1 Model Predictions

Figure 7.8 plots the mean observer angle settings for the monocular ‘Gradient+Outline’ condition of Experiment 1 (black line) and shows the predicted angle settings from the weighted average weak fusion model, and
the Bayesian models. Figure 7.9 plots the observer variance from the ‘Gradient+Outline’ condition with the various models’ predicted variances.

Figure 7.8 Mean shape settings made by individual observers in the Gradient+Outline condition of Experiment 1 (solid black line), together with ‘Weighted Average’ Model predictions (dash/dot red line) and Bayesian Model predictions (dashed lines).
As a means of quantifying how well each model predicts the combined-cue data, Figure 7.10 shows the coefficient of determination ($R^2$) for the shape setting predictions of each model. The use of $R^2$ as a measure of goodness-of-fit has limitations. For example, with non-linear data, it cannot explicitly
distinguish between models that provide the correct pattern in their prediction and those that do not; nor can it provide any information about the relative levels of systematic and random error in the fit – $R^2$ does not distinguish between errors that are both below and above the true value and consistent error in one direction, e.g. a systematic underestimation. However, it does provide a simple, easy to understand metric for initial comparison.

![Figure 7.10](image)

Figure 7.10 Coefficient of Determination ($R^2$) values for the model predictions of mean perceived angle in the Gradient+Outline condition data of Experiment 1.

For four of the six observers, of the Bayesian models, the ‘best guess’ simplified model provides the best predictions of perceived shape (Figure 7.10). This model also provides equal or better performance than the
‘weighted average’ weak fusion model for 4 of the six observers. However, for observers that showed very non-linear shape perception (e.g. AC) none of the Bayesian models replicate the shape of the curve very well (Figure 7.8), and the ‘weighted average’ is considerably better in this respect. Examination of the data from the monocular single-cue conditions of Experiment 1 (Figures 6.2 and 6.4) reveals that for both cues, the underestimation bias reduces as the stimulus angle increases, while the observation variance stays similar in the case of the gradient cue, and increases in the case of the outline cue. This pattern explains why the ‘Simple Gaussian’ model (Model 1) cannot fit the data well: To reduce underestimation as the stimulus angle increases, the influence of the cue must increase relative to the prior (i.e. the likelihood becomes narrower). This causes a reduction in the variance predicted by the model - a pattern not seen in the data. Therefore, if the model is to contain one cue likelihood and one prior (and there is no obvious reason to expect additional components), it must have non-Gaussian components to enable a better fit to these data. For this reason, the Bayesian models typically fail to predict the pattern that is seen in the observer data (black line in Figure 7.8) of reduced underestimation (reduced bias) as the stimulus angle increases, and in fact predict the opposite pattern to some extent. The ‘best guess’ simple model, largely due to its heavy tailed prior, is better able to replicate this pattern in the data.

The ‘weighted average’ weak fusion model produces a variance estimate that for two of the six observers is too low (red dot/dash line in Figures 7.9 -
observers BM and LM). The Bayesian models should be expected to provide higher estimates of variance since, compared to the ‘weighted average’ model the effects of any prior are reduced when cues are combined, since the prior is included only once in the Bayesian models (but implicitly included twice in the ‘weighted average’ model). This effect is seen in the variance predictions of the Bayesian models (Figure 7.9) but all these models noticeably overestimate combined-cue variance for all observers. It is not clear why this is the case, but it could be that the models are missing a component – typically the reliability of the outline cue is underestimated (see Section 7.2.2.5) and therefore the model’s predicted variance for the Outline-cue-only condition will be overestimated. This suggests that some other prior may be a factor when making shapes estimates using the outline-cue that is not applicable in when using the gradient cue, and the lack of this component in the models could lead to the overestimation of variance for the combined-cue condition.

7.2.4 Discussion of Bayesian Modelling

Experiment 1 was designed to explore how humans use 3D shape information from complex illumination gradient cues, both in isolation and when in combination with other cues to shape. Because shape settings made by observers showed a large bias towards underestimating the angle of the stimuli, a range of Bayesian models were developed to try and examine how this bias might occur, and to explore the suitability of this type of model in describing the visual processes involved in shape perception. In this discussion section, I first consider how well the Bayesian models can
replicate the biases seen in the results of the monocular conditions of Experiment 1, and how this may improve our understanding of how the biases in the data arise. Secondly, I discuss how useful this type of Bayesian modelling is for helping us understand shape perception.

### 7.2.4.1 Sources of Bias

A significant aspect of the data from all the conditions of Experiment 1 is the tendency of observers to underestimate the stimulus angle. Section 6.3.3 discussed how this bias might arise and how it seems likely that the underestimation of stimulus angle may be because observers integrate prior knowledge about object shapes, or surface orientations into their estimations, using a prior distribution that includes a greater frequency of objects or surfaces with low slant than high slant. As previously noted, the authors of a number of studies have also suggested the existence of priors for flatness, or fronto-parallel surface perception. For these reasons, a prior distribution for ‘flatness’ was included in the Bayesian models described above, and in some implementations is successful in replicating the pattern of perceived shape seen in the data. The models that use non-Gaussian components (Models 2, 3 and 4), particularly the ‘best guess’ model (Model 4), are often able to more accurately fit the shape settings made by observers in the monocular ‘single-cue’ conditions of Experiment 1 (Figure 7.10) than Model 1, which contains only Gaussian components.

An important aspect of the ‘Relative-Side-Length’ model (Model 3) is that it provides an explanation of the increasing observer variance as the stimulus
angle increases in the Outline-cue-only condition of Experiment 1 (see Figure 6.5). Higher observation variance at high stimulus angles suggests that the cue is less reliable at higher angles. This feature is built into the model by basing the cue likelihood function on the relative length of vertical card fold and edges in the projected 2D image. The ability to explain patterns in the data in this way increases confidence that an approach based on Bayesian inference may be able to provide a model that accurately represents the processes involved in 3D shape perception.

7.2.4.2 Model Performance

When considering how a Bayesian model might account for the single cue data, it is clear that because the underestimation bias in observers’ perceived angle reduces as stimulus angle increases (see Figure 6.3), then any priors that might cause lower than veridical angle estimation must have reduced influence at higher stimulus angles. However, since Bayesian priors must, in a proper Bayesian framework, be independent of the stimulus (and indeed Beierholm et al. (2009) verified that this is seems to be the case for 3D shape perception), the reliability of the prior cannot be lower for higher stimulus angles. An increase in cue reliability as an alternative means of reducing bias also seems to be ruled out by the lack of any reduction in variance at higher stimulus angles, and this is why Model 1 is the least likely of the Bayesian models to account for the data from the ‘single cue’ conditions of Experiment 1 (Figure 7.7). The use of non-Gaussian components in the model (Models 2, 3 and 4) provides a means of overcoming these issues. One possibility is a model
containing a prior distribution with heavy tails (Model 2). It was found that this ‘Heavy-Tailed Prior’ model delivered a slightly better fit to the ‘Gradient-cue-only’ and ‘Outline-cue-only’ condition data for some observers, and better predictions of combined-cue behaviour for 4 out of 6 observers (Figure 7.10), than were possible than when using Gaussian priors. Similar heavy tailed prior distributions have been shown to explain biases in motion perception (Stocker & Simoncelli, 2006).

Another, although not mutually exclusive, possibility is that the cue likelihoods are non-Gaussian. If cue reliability changes with stimulus angle, then the likelihood functions should be skewed. The ‘Relative-Side-Length’ model creates cue likelihood functions for the outline cue that are skewed and heavy tailed for large stimulus angles (see Figure 7.5). Heavy tailed likelihood distributions (Girshick & Banks, 2009) and priors (Knill, 2007b) have been suggested as a means of explaining robust cue integration when large cue conflicts are present (see also Lange, Little, & Taylor (1989)). Heavy tailed prior distributions (Models 2 and 4) result in a reduced bias in the perceived shape (but increased variance) when large conflicts occur between the prior and cue information. Heavy tailed likelihood functions for the outline cue (Models 3 and 4) result in increased variance at higher stimulus angles when the outline cue is present. In both cases, this matches the patterns seen in the measured shape settings. Model 4 incorporates both a heavy tailed prior and skewed likelihood functions for the outline cue. As a result is the best performing of the Bayesian models described here, providing a better fit to the single-cue condition data (when the number of
free parameters are taken into account – see Section 7.2.2.5 and Figure 7.5), and better predictions of combined-cue perception (Section 7.2.3.1, Figures 7.8 and 7.10).

The biases that the Bayesian models assume to be introduced by priors are quite large in all the conditions of Experiment 1. This results in what may be surprisingly narrow prior distributions in the models (often narrower than the cue likelihoods – see Table 7.1 for examples). Although this could be taken as an indication that other sources of bias may have influenced the results, similar narrow priors have been found in some previous studies (e.g. Lages (2006) and Welchman, Lam, & Bülthoff (2008)) and the introduction of large biases and possible 'less optimal' behaviour when strong priors are in use has also been noted (Vilares & Kording, 2011).

If observation variance is assumed to originate from variation (due to internal noise) in position of the likelihood function (Stocker & Simoncelli, 2006), rather than the use of a probability matching decision rule, as assumed by the models above, then the form of prior distribution affects how this variation influences the predictions of the model. For example a prior that is relatively weaker at higher stimulus angles (i.e. a heavy tailed prior) causes the position of peak of the posterior distribution to vary more at higher stimulus angles (for the same level of internal noise). This could also account for greater observation variance without the need for reduced cue reliability at higher stimulus angles.
Finally, the Bayesian models described here predict the more veridical angle settings that occur in the combined-cue ‘Gradient+Outline’ condition for the majority of observers (see Figures 6.6 and 7.8). This is something that the ‘weighted average’ weak fusion model cannot do. The reason this is possible for the Bayesian models is that a single prior constraint was used to model all the conditions. Therefore, in the combined-cue condition, the influence of the prior is reduced compared to the ‘single-cue’ conditions because there are two cues rather than one (thus greater cue reliability) combined with the same prior information. This results in a smaller bias due to the prior in the predictions of the Bayesian models for the combined cue condition, similar to that seen in the experimental data.

### 7.2.4.3 Model Validity

#### 7.2.4.3.1 Why Use a Bayesian model?

While Bayesian inference has been suggested by many as a suitable model for the cognitive processes involved in shape cue integration, weak fusion models that do not explicitly account for prior information, such as reliability weighted average (MLE) models (e.g. Ernst & Banks (2002), Hillis et al. (2004), Lovell et al. (2012), Oruc et al. (2003)) have also been shown to work well. In fact, these simpler models can often produce good results even when prior constraints or biases are apparent. However, it could be argued that these models only work well because the effects of any priors are contained in the individual cue data before combination. In other words the biases are implicitly part of the model. In fact the MLE model described in section
7.1 is equivalent to a Bayesian model, built using Gaussian likelihoods and prior (similar to Bayesian Model 1 described in section 7.2), but where the prior is allowed to vary in width dependent on the stimulus angle. However, as noted in section 7.1, the use of such models cannot provide any insight into the precise details of biases in observation data. For example, findings such as the possibility that observers could use priors distributions that are not Gaussian, and that the outline cue may be based on the apparent relative lengths of the card fold and vertical edges, could not have been arrived at without using Bayesian modelling.

7.2.4.3.2 How Much Can Bayesian Model Really Tell Us?

Inferring the precise way in which prior knowledge affects depth perception is a difficult task. Direct measurement of prior distributions that may be encoded within the visual system is not possible, since we cannot probe such information without confounding effects. In addition, because any priors are always present, even responses to ‘single-cue’ stimuli (in a shape estimation task) do not measure cue likelihood functions, but estimates made using both the single cue and any priors. Thus the calculation of the likelihoods and priors of a Bayesian model must be made by working ‘backwards’ from the settings made by observers, or calculated from physical attributes of the stimuli. Unfortunately, even with only one cue likelihood and one prior, there are mathematically an infinite number of possible combinations that could produce a given predicted perceived shape. To surmount this problem, and constrain the problem sufficiently to deduce the processes that may be occurring in the visual system, some assumptions about the form of the
model must be made. First, assumptions must be made about how the posterior distribution of the model translates into the predicted average perceived shape and variance (in the models described in this chapter a ‘non-committing’ decision rule is used). Second, suitable assumptions must be made about the location and form of the likelihood functions and prior distributions in the model (here a prior centred on zero angle, and likelihood functions centred on the stimulus angle are assumed, and different distribution and function shapes are explored). In this way it is possible to explore the reliability of the cues and the shape of the prior distribution for each of the experimental conditions.

The models could also be further constrained by additional experimental data. If the reliability of shape settings at each stimulus angle were measured using a two-alternative forced choice paradigm, the results of which would not be affected by biases due to any priors, then the widths of the likelihood functions of the model could be fixed. However, this method still leaves unknown the precise shape of the likelihood functions: we must still make assumptions about the distribution of the data when fitting psychometric functions to find the cue reliability, for example by using a cumulative Gaussian function. This is important because likelihood function shape can be critical in how well the model can fit observation data (Figure 7.5).

As a result of the number of assumptions that must be made when creating a Bayesian model, and the number of different ways in which such models can be made to fit the data, there is clearly a limit to how much Bayesian modelling can tell us. Certainly it would be difficult to argue that any of the
specific models presented in this chapter represents a true picture of the precise processes occurring within the visual system; a criticism that has also been levelled against Bayesian modelling of perception by others (e.g. Bowers & Davis, 2012). Fit to ‘single-cue’ data (Figure 7.7a) is best for the ‘Relative-Side-Length’ model, but this model has a large number of free parameters. On the other hand, the failure of the ‘Simple Gaussian’ model to fit and predict the combined-cue data, does suggest that features such as heavy tailed priors and likelihoods (that facilitate more robust behaviour) are likely if the visual system does process shape information in a way that is equivalent to Bayesian inference. Additionally, the ‘Best-Guess’ Simple Model that includes these features was found to be the most likely to account for the measured ‘single-cue’ data, once the number of free parameters is taken into account (Figure 7.7b), and it makes good predictions of the combined cue behaviour (Figures 7.8, 7.9 and 7.10).

7.3 Modelling Summary

In this chapter a number of approaches to modelling the perception of 3D shape in the monocular conditions of Experiment 1 are undertaken. A ‘weighted average’ weak fusion model is shown to be capable of making reasonable predictions for perceived shape and observer variance when the gradient and outline cues are combined, but this type of model is unable to explain biases in the ‘single-cue’ data, or the more veridical settings made when cue are combined. Because biases in the results of Experiment 1 suggest that observers seem to use prior information about the statistical likelihood of stimulus shapes when making estimates using either cue alone,
or a combination of cues, Bayesian models of perception and cue-combination seem to provide a more suitable approach. Several Bayesian models are therefore constructed to investigate the perceptual biases demonstrated by observers in Experiment 1. Non-Gaussian components (either the prior or cue likelihood) are found to be required in a Bayesian model to fit the data well. The model that is most likely to account for the observed data (Model 4) contains both a heavy tailed prior and non-Gaussian likelihoods for the outline cue that are based on the apparent relative length of the central fold in the stimulus and its vertical edges.
Chapter 8- Experiment 2: Learning to use the Gradient Cue

In the Gradient-cue-only condition of Experiment 1, four of the six observers were able to reliably distinguish concave and convex versions of the stimulus (see Chapter 6, Figure 6.4). In this condition only a monocular gradient cue to shape was available. Illumination gradients (shading) are an inherently ambiguous cue to shape (Belhumeur, Kriegman, & Yuille, 1999; Horn, 1975; Pentland, 1984, 1988), so in order to make unambiguous settings observers must be using prior knowledge, or making some assumptions about the stimuli. The outline cue is, like monocular cues in general, also ambiguous, so the same applies in this condition too. However, the assumptions required to use the outline cue are quite simple. Specifically that the surfaces are flat, rectangular and rigid. As noted in section 6.3.2, there is evidence that the human visual system does make these assumptions and it is therefore not surprising that settings for the task with outline cue only are consistent and reliable.

Section 6.3.2 also discusses the fact that the use of the gradient cue requires assumptions that are much more complicated because of the complex interactions between object shape, material and illumination that give rise to the specific gradients present on the stimuli. Using such variations in luminance and chromaticity to infer object shape or surface colour is therefore a difficult problem for the visual system. If humans gain shape from
shading via a similar process to many shape from shading computer
algorithms (e.g. Horn (1975)) then assumptions about the light source form
and position, along with assumptions about the reflectance properties of the
surfaces of the stimulus (albedo, colour, glossiness) would be needed. As
well as the results of Experiment 1, some previous work has suggested that
the human visual system is sensitive to both the luminance and chromatic
information contained in complex colour illumination gradients and that
features such as inter-reflections (mutual illumination) may be useful for
unambiguous shape perception (Ruppertsberg et al., 2008). However, it is
unclear what disambiguating information is learnt from the scene, or
assumed, and how observers go about this disambiguation process.
This question regarding what is required by the visual system to make use of
complex illumination gradients as shape cues, leads to the first question that
Experiment 2 is intended to answer: is the visual system capable of utilising
illumination gradients as a cue to shape without any other cues, and without
prior exposure to the stimulus? Because the gradient cue is ambiguous, to do
this the visual system would have to use ‘built-in’ prior knowledge /
assumptions about the ‘global’ properties of shading to make use of the cue
information. To answer this question perception of object shape was tested,
using gradient information only, in the same way as the Gradient-cue-only
condition of Experiment 1. This condition (‘No Training’) of Experiment 2 was
performed by a new group of naïve observers who, unlike those in
Experiment 1, had no prior experience of the stimuli before making their
shape settings. Observers in Experiment 1 were shown a movie before the
experiment that contained smaller, low-resolution versions of both the top-
down view setting task and the experimental stimulus. Observers in this ‘No Training’ group of Experiment 2 were not shown this video.

If visual priors alone are not sufficient to make use of gradients, the results of the ‘Gradient-cue-only’ condition of Experiment 1 must mean that the information required to disambiguate the gradient cue can be easily learnt. Observers in Experiment 1 had an opportunity to learn some properties of the lighting and stimulus from the video they were shown before the experiment started. In the video, unlike in the Gradient-cue-only condition of the experiment itself, the stimulus contained both outline and gradient cues (Figure 5.6 right column). The movie went through two cycles of stimulus angles, from the steepest convex to the steepest concave angle and then vice versa, over a time course of around 30 seconds. While observers were asked to view the movie simply as an aid to understanding the task, and not as training to learn any specific aspects of the stimulus, it seems that this short movie may be sufficient to provide enough information for observers to disambiguate the gradient cue effectively.

The second question Experiment 2 was intended to answer is that if observers do need to learn from the training video in order to use the gradient cue, do they learn an arbitrary mapping between shape and gradient (i.e. use memory of the video), or are they able to use the video as a means of gathering information about the physical connection between shape and gradient, and make the required assumptions to use the gradients as a true visual cue to shape? To investigate this question, three further groups of
naïve observers were tested. Two of these groups were shown training videos containing simpler, pared down, versions of the stimuli: one group saw a video where the average luminance and chromaticity of each side of the card was applied uniformly over the relevant side (preserving the effects of shape on the cards’ average colour/luminance, but removing gradients); the other group saw a video containing only a wire frame version of the stimulus, with no gradients or colour information. A third group were exposed to video containing stimuli with incongruent gradients, corresponding to the inverse (concave vs. convex) stimulus shape. This last condition was designed to reveal if observers used memory of the video to such an extent that they would learn unrealistic mappings between gradient and outline, and therefore make corresponding inverse shape settings.

8.1 Method and Stimuli

The general method and stimuli used in Experiment 2 were identical to those in the Gradient-cue-only condition of Experiment 1 (see Chapter 6, section 6.1). Stimuli contained only the gradient cue to shape, and other cues were removed or minimised by the use of a viewing aperture (see section 6.1). Experiment 2 differed from the gradient-only condition of Experiment 1 only in the training video seen by observers before they made their shape settings, as outlined above.
8.1.1 Training Videos

Four conditions were tested (the Gradient-cue-only condition of Experiment 1 can be thought of as a fifth condition, which is henceforth referred to as the ‘Full Training’ condition):

‘No Training’ - observers did not see a training video before the experiment.

‘Outline Training’ - the training video used a version of the stimulus that contained only the outline cue.

‘Uniform Colour Training’ - the training video used a version of the stimulus that contained an outline cue plus uniform coloured stimulus surfaces (one colour for each side of the card) that matched the mean luminance and colour of the gradient cue. Note that the mean luminance and colour was different for each stimulus angle.

‘Incongruent Training’ - the training video used a version of the stimulus that contained an outline cue plus a conflicting gradient cue from the reverse shape. For example, a +40 degree (concave) outline was paired with -40 degree (convex) shading.

In addition to the training video, all observers received a verbal explanation of the task. Examples of these different training video stimuli are show in Figure 8.1.
8.1.2 Observers

Because each condition of Experiment 2 required observers to be trained in a different way, separate groups of observers were required for each condition. 16 naive observers (mean age 38) took part in Experiment 2, 13 females and 3 males. These were split into groups of four observers per condition. All subjects had normal colour vision and stereo acuity, and normal or corrected to normal acuity. Normal colour vision was verified using the Farnsworth-Munsell 100 hue test (score below 100). Normal stereo acuity was verified with the TNO stereo test (120 seconds of arc or less).
8.2 Results and Analysis

In the analysis of Experiment 1, I considered if observers were able to correctly and consistently set the shape to be either a corner or roof when presented with only the gradient cue. This is an important consideration, because this will only be possible if observers are able to resolve the complexity of the gradient information and use it as a consistent shape cue.

Figure 8.2a shows mean shape settings for four naïve observers that only received a verbal explanation of the task before starting the experiment, and had not seen any training video (‘No Training’ group). Observer angle setting as a function of physical stimulus angle for convex (negative) and concave (positive) angles is plotted. There was some variation of setting with stimulus angle, with large angles being set as larger, but there were no negative average responses (corresponding to perception of a convex ‘roof’ shape), despite half the stimuli specifying a convex shape. All observers in the ‘No Training’ group were unable to distinguish concave and convex shapes (Fisher’s exact test, per observer; SA: p=0.8, AC: p=0.5, CM: p=0.2, WH: p=0.2).
Figure 8.2 Observer shape settings for the four training conditions of Experiment 2. Mean angle settings (all observers) as a function of physical angle for roof (negative angle, red squares) and corner (positive angle, blue diamonds). Error bars show standard error of the mean. ‘N’ indicated the number of observers. The dashed line on (a) indicates veridical performance for all figures. (a) ‘No Training’ (verbal instruction only). (b) ‘Full Training’ (short training video with a stimulus containing both gradient and outline cues and correct setting lines for the displayed stimulus angle). (c) ‘Outline Training’ (video with a stimulus containing the outline cue and setting lines only). (d) ‘Uniform Colour Training’ (video without detailed gradient cue, but correct mean colour and luminance on each card surface, along with the outline cue and setting lines).

In contrast, four of the six observers in the ‘Full Training’ group (i.e. the Gradient-cue-only condition of Experiment 1) made settings that were clearly...
of opposite sign for roof and corner, and that increased with physical angle (see Chapter 6, section 6.2.2). These four observers were able to assign roofs and corners to the correct category (Fisher’s exact test; AC: p= 0.01, BM: p=0.01, LM: p=0.002, SC, p=0.01), whilst the remaining two were not (LL: p=0.3, MD: p=0.54). Average settings for this group are re-plotted in Figure 8.2b for comparison.

The results so far suggest that observers in the ‘No Training’ group could not make consistent settings because they did not have sufficient visual information to disambiguate the gradient cue. Alternatively, these observers may have failed to understand the task correctly without viewing the training video. This second hypothesis was tested by asking a further two groups of four observers to make settings after watching training videos containing either a wire frame version of the stimulus (‘Outline Training’) or a stimulus where the luminance and chromaticity of the card sides was averaged (for each stimulus angle independently) such that each side of the card was a uniform colour, containing a spatially coarse representation of the gradient information (‘Uniform Colour Training’). Figure 8.2c shows the mean settings of the ‘Outline Training’ group, who showed similar behaviour to that in of the ‘No Training’ group (Figure 8.2a). Three of the four observers in this group were unable to distinguish concave and convex shapes (Fisher’s exact test; CH: p=0.1, YR: p=0.5, HH: p=0.7). Results for one observer did show a significant ability to assign shapes to the correct category (Fisher’s exact test; NI: p=0.05). Mean observer settings for the ‘Uniform Colour Training’ group are shown in Figure 8.2d. This training video provided enough information for
the observers to make similar shape settings to those made with ‘Full Training’ and with the more reliable outline cue (Figure 8.3), and 2-way between subjects ANOVA revealed no significant difference between the ‘Full Training’ and ‘Uniform Colour Training’ groups. Three of the four observers in the ‘Uniform Colour Training’ group were able to assign roofs and corners to the correct category (Fisher’s exact test; AC: p=0.05, MA: p=0.04, PC: p<0.001, CS: p=0.1).

Figure 8.3 Comparison of mean observer settings for the ‘Full Training’ and ‘Uniform Colour Training’ conditions, and the Outline-Cue-Only condition of Experiment 1. Vertical axis: observer setting (degrees); Horizontal axis: stimulus angle (degrees). Red circles: Outline-Cue-Only condition of Experiment 1; Green squares: ‘Full Training’ (Gradient-Cue-Only condition of Experiment 1); Purple triangles: ‘Uniform Colour Training’ condition.

The settings made by the ‘Outline Training’ and ‘Uniform Colour Training’ groups suggest that observers must first learn some information about the
characteristics of the scene and illumination in order to use the gradient cue (training with an outline-only defined stimulus is not sufficient), but exposure to detailed gradients are not needed to learn the required information. Although the results of the conditions of Experiment 2 described so far suggest that observers need to learn something about the object / scene properties in order to use the gradient cue, they do not need to learn the gradients themselves. However, there remains the possibility that observers could be learning a simple mapping of gradient (or average colour) to shape, by relying on their memory of the training video. The final condition (‘Incongruent Training’) was designed to investigate if this was the case. In this condition observers were trained using a video containing a version of the stimulus with incongruent shape cues, with the outline and gradient representing opposite shapes. For example a +40 degree (concave) stimulus in the training video had a +40 stimulus outline and setting line position, but the shading from the -40 degree stimulus (convex) was applied. If observers make shape matches based on an shape-gradient mapping they learned during the training video, it might be expected that this group would make inverted settings, incorrectly assigning concave gradients to convex shapes and vice versa. In fact, inverted settings were made by only 1 of the 4 observers in the ‘Incongruent Training’ group. Figure 8.4 plots the results for this group, showing the mean settings for each of the four observers separately. The three observers who did not make inverted settings could assign concave and convex shapes to the correct category (Fisher’s exact test; SM: p=0.03, LP: p=0.03, HK: p=0.05) and performed similarly to those in the ‘Full Training’ group (Figure 8.2b). The fourth observer made inverted
shape settings as if relying on memory of the training video and did not assign stimulus shapes to the correct category in general (Fisher’s exact test; GD: p=0.2). This result provides evidence that, for the majority of observers in the ‘Incongruent Training’ group (3/4), learning from the training video affords an opportunity to establish assumptions about the scene that allow them to correctly use the gradient in the stimuli as a shape cue, rather than making their settings based directly on the gradient-shape correspondence seen during training (i.e. using memory). It seems possible that assumptions about illuminant position and surface reflectance are established during the learning, in order to disambiguate the gradient cue.
Figure 8.4 Observer shape settings for the ‘Incongruent Training’ group. Mean angle settings as a function of physical angle are shown for each of 4 observers separately. Error bars show standard error of the mean. ‘Incongruent Training’ consisted of a short video with the gradient cue indicating the opposite shape to the outline cue. For example: -40 degree shading displayed with +40 degree outline. Only one of the four observers (GD) made the ‘reversed’ pattern of settings that would be expected from reliance on memory of the training video.

8.3 Discussion

8.3.1 Can Priors Disambiguate the Gradient Cue?

Luminance and colour gradients are inherently ambiguous cues to shape, dependent on object shape, material, and the lighting environment. The literature on the use of shading as a depth cue is not conclusive on what prior information or assumptions might be required by the visual system to effectively use these cues. While classical computer shape-from-shading
algorithms (for example Horn (1975)) typically require knowledge of light position and surface reflectance properties in order to calculate local surface orientation, it is not clear from the literature if this is also the case for the human visual system. Some work has shown that the visual system may not use these assumptions and instead rely on a process dependent on global properties of shading on 3D shapes (Erens et al., 1993; Koenderink & van Doorn, 1980; Mingolla & Todd, 1986; Nefs et al., 2006). However, it has also been shown that humans are able to judge, and are sensitive to changes in, light position in 3D scenes (for example Gerhard & Maloney (2010), Maloney (2002), Ruppertsberg et al. (2008)), and may estimate illuminant position (Bloj et al., 2004; Boyaci et al., 2003).

The first question outlined in the introduction of this chapter was whether or not prior knowledge / assumptions used by the visual system (such as the ‘light from above prior’) are sufficient to disambiguate the gradient cue. Such priors would enable observers to make consistent, unambiguous settings without learning anything about the specific stimuli in the experiment before making their settings. Observers in the ‘No Training’ group were not given any training before making shape settings. Results from this group (Figure 8.2a) show that without prior exposure to the stimulus, observers are unable to make reliable shape settings, indicating that visual priors, or shape extraction mechanisms using only global shading and 3D shape properties are not sufficient to disambiguate the gradient cue. Observers in the ‘Full Training’ group (i.e. the Gradient-cue-only condition of Experiment 1) had an opportunity to learn assumptions about scene properties, such as illuminant
position and surface reflectance before the experiment started. The short learning phase, with a time course of just 30 seconds, appears to provide enough information to later interpret complex changes in gradient (see section 6.3.2, Figure 6.15) across the scene as a shape cue. This was possible without the addition of any other shape cues (except during the training video). For the majority of the observers who received full training, there was no ambiguity in responses over whether gradient depicted a convex roof, or a concave corner.

8.3.2 What do Observers Need to Learn in Order to Use the Gradient Cue?

The second question set out in the introduction asks, if visual priors alone cannot disambiguate the gradient cue (and the results of the ‘No Training’ condition show this to be the case), what is the information observers learn from the training that enables them to make unambiguous settings in the ‘Full Training’ condition? The results of the ‘Full Training’ condition alone cannot tell us if observers learn an arbitrary mapping between shape and gradient from the video training, or if they are really able to make and use suitable assumptions, perhaps about lighting and object properties to understand the connection between shading gradients and shape.

To answer this question, three further groups of observers made shape settings. Each of these groups was provided with different information about the gradients during the training video. In the ‘Outline Training’ group, observers were trained using a video describing the task and showing a
representation of the stimulus without any shading (wireframe). This provided more visual information about the task than that received by the ‘No Training’ group, but also did not provide any information about the shading on the stimulus. Observers in the ‘Outline Training’ group, like those who received no training, performed poorly when making settings using the gradient cue alone (Figure 8.2c). This suggests that in the ‘Full Training’ case, the training video provides not only an aid to understanding the task, but also visual information about the scene that is needed to make accurate shape judgements.

The ‘Uniform Colour Training’ video contained a version of the stimulus with a spatially coarse representation of the shading, such that no gradients were present across the card (each side was of uniform colour and luminance), but the mean luminance and chromaticity of each side was the same, for each card angle, as in the ‘Full Training’ video. When trained using this video, observers performed as well as those exposed to the ‘Full Training’ video. This suggests that observers do not need to see the specific gradients in the training video in order to learn enough to use the gradient cue in the stimuli to successfully establish shape.

The final group of observers were shown the ‘Incongruent Training’ video, which was specifically designed to test if the training could result in a learnt mapping between any gradients and shape, rather than providing scene information that can be used to disambiguate the gradient cue. In this video, the stimulus contained gradients for the inverse of the shape given by the
object outline (the setting lines in the video were correct for the outline cue, not the gradient cue). Observers who made ‘memory matches’ of gradient to shape would therefore be expected to make reversed shape settings. Three of the four observers did not make reversed settings, but instead performed similarly to those who were trained using congruent stimuli. This behaviour demonstrates that these observers were using the shading gradients in the stimuli as a visual cue to shape when making their settings, and not relying on their memory of the training video. The ‘Incongruent Training’ video did not provide observers with a consistent mapping of shading to shape, but did yield some information about the scene, including the typical colour of the stimuli, lack of shadows and highlights, lighting position and importantly, constraints on the shape of the stimulus suggested by the strong object outline cue (i.e. that the card sides were square and made of a rigid material that was only folded at the central spine). This information about the properties of the scene appears to be enough to allow observers to establish sufficient and suitable assumptions to solve the shape from shading problem when presented with only the shading gradients in the experiment.
8.3.3 Gradients or Mean Luminance and Colour?

The experiments described above show that when observers are trained using a 'uniform colour' version of the stimuli they make similar shape judgements to observers trained using stimuli that contain gradients. While this indicates that prior knowledge of the specific gradients in the stimuli is not required to perceive accurate 3D shape, it does not tell us if observers actually require the gradients at all. It is possible that they use the average luminance and colour of the stimuli as the shape cue and the gradients are not required. The mean luminance and CIE 1976 (L*, u*, v*) chroma (related to perceived ‘colourfulness’) for left and right sides of the Gradient-cue-only stimuli, at each stimulus angle, are shown in Figure 8.5. It is difficult to make direct comparison between these properties and the total cone contrast of the gradients in the stimuli, since these are rather different qualities, but it is clear that both mean luminance and chroma change significantly over the range of card angles. Mean values for each side of the card change over the range of card angles by one or two orders of magnitude – a similar rate of change to the total cone contrast of the gradients (see Figure 6.2). Both mean luminance and mean colour of the Gradient-cue-only stimuli could therefore provide a useful shape cue if observers are able to relate these properties to the 3D shape of the object.
To investigate if observers were able to use the mean luminance and colour of the stimuli as a cue to the 3D shape, in the absence of any shading gradients, a ‘uniform colour stimulus’ control experiment was carried out: in this experiment an additional naïve set of observers were trained using the ‘full training’ video (which contained stimuli with gradients; see section 6.1), but made shape matches to a version of the stimuli that had uniformly coloured stimulus surfaces (one colour for each side of the card). The colour of each side matched the mean luminance and chromaticity of the Gradient-cue-only stimuli (similar to the ‘uniform training’ video described in section 8.1.1 and shown in Figure 8.1c). The luminance and chromaticity of each
side of the card was therefore different for each stimulus angle (see Figure 8.5) and provided a potential shape cue, but no shading gradients where present. The experiment was carried out using exactly the same method as used previously in Experiments 1 and 2 (see section 6.1). Seven naïve observers took part in this control experiment (5 female). All subjects had normal colour vision and stereo acuity, and normal or corrected to normal acuity. Normal colour vision was verified using the Farnsworth-Munsell 100 hue test (score below 100). Normal stereo acuity was verified with the TNO stereo test (120 seconds of arc or less).

As noted above, observers were trained identically in the Gradient-cue-only condition of Experiment 1 and the ‘uniform colour stimulus’ control. The only difference between the two experiments being the removal of the gradients from the stimuli used in the latter.

Shape settings made by each of the observers in this ‘uniform colour stimulus’ control experiment are shown in Figure 8.6. It can be seen from this figure that the majority of observers we unable to distinguish concave and convex shapes, or distinguish between stimuli at different angles. The results of only two of the seven observers (AN and SA) provide an obvious indication that the observer perceived any 3D shape. Statistically, only two of the seven observers could tell convex from concave shapes (Fisher’s exact test; AM: \( p=0.10 \); AN: \( p=0.02 \); CA: \( p=0.26 \); CU: \( p=0.01 \); SA: \( p=0.07 \); No statistics for observers MA and MR – for both observers mean shape settings were always concave).
Figure 8.6 Results of the ‘Uniform Colour Stimulus’ control experiment. Data are shown for each observer separately. For each observer, data are shown for each trial. Vertical axis: observer setting (degrees); Horizontal axis: stimulus angle (degrees). The red line represents expected mean settings for veridical perception.
The observers in this control made clearly worse shape settings than those in the Gradient-cue-only condition of Experiment 1 where gradients were present in the stimuli (see section 6.2.2 and Figure 6.4 for comparison). The combined mean shape settings of all observers show that observers cannot in general gather much useful 3D shape information from the ‘uniform colour’ stimuli. These data are shown in Figure 8.7, together, for comparison, with the mean settings made in the ‘Gradient-cue-only’ condition of Experiment 1.

**Figure 8.7** Mean shape settings made by all observers in the ‘Uniform Colour Stimulus’ control experiment (red diamonds). Also plotted for comparison, the mean shape settings made in the ‘Gradient-Cue-Only’ condition of Experiment 1 (blue squares). The red line indicates veridical performance.
Mean shape settings are smaller (flatter shape) than when the gradients are present in the stimuli. A two-way between subjects ANOVA shows that this is a significant effect for the settings made by all observers (F(1,14)=31.4; p<0.001).

The results of the ‘uniform colour stimulus’ control indicate that the luminance and/or chromatic gradients in the ‘Gradient-cue-only’ stimuli do provide a shape cue and without the gradients observers are typically not able to make useful judgements of 3D shape for stimuli that lack any outline shape cue. The average luminance and colour, while in theory providing shape information, do not appear to provide a useable shape cue in isolation, and gradients appear to be critical in using the colour shading cue. Experiment 3 (Chapter 9) extends the investigation further with the aim of determining if it is the luminance component or the chromatic component of the gradients (or both) that provide the shape cue.

**8.4 Experiment 2 Summary**

The results of Experiment 2 demonstrate that visual prior assumptions do not appear to be sufficient to enable observers to effectively use the ambiguous gradient cue, when this is the only shape cue available. However, a small amount of training seems to provide enough opportunity for observers to gain sufficient information about the scene to disambiguate the gradient cue and make consistent shape settings. Not only does the visual system appear to make suitable assumptions, after training, about the reflectance properties of the card surfaces and about the light position in order to use the gradient cue
unambiguously, but settings made using the gradient cue are also very
similar to those made using the outline cue alone (see Experiment 1 analysis,
section 6.2.2.1). If the assumptions made about the scene are very different
from the true scene parameters, then perception will be inaccurate. However
observers in both the ‘Full Training’ and ‘Uniform Colour Training’ groups
made angle settings that were not significantly different and were similar to
those made with the outline cue (see Figure 8.3). Thus, it seems reasonable
to conclude that the visual systems of observers in these groups must have
learned a reasonable set of assumptions about the scene, and this is
possible with limited training that does not need to have the same detail level
as the stimuli. Additionally, the results of the ‘Incongruent Training’ condition
show that the majority of observers do not make inverted settings when
trained using incongruent gradients (specifying the inverse shape), providing
evidence that observers really can learn assumptions about the light
environment and object properties, rather than learning a simple mapping
between outline and gradient. Finally, a control experiment in which
observers made shape settings for stimuli that contained only a ‘uniform
colour’ shading cue indicated that the luminance and/or chromatic gradients
in the shading are important and provide more useable shape information
than just the mean luminance and chromaticity values.
Chapter 9 - Experiment 3: Shape from Chromatic Gradients

The results of Experiment 1 demonstrated that observers were capable of using the complex colour illumination gradients within the stimuli, which contained both luminance and chromatic variation, to determine object shape consistent with that indicated by other shape cues. This was possible without the addition of any other cues to shape in the stimuli that might help disambiguate the gradient information. Experiment 2 showed however, that a small amount of experience of the 3D scene is required to enable observers to make suitable assumptions about the object and lighting properties that are needed to use the gradient cue. One of the interesting aspects of this result is that the visual system is able to use the gradient cue despite the complex nature of the chromatic and luminance components (see Figure 6.15) that make up the realistic gradients in the stimuli, even when additional chromatic variation is present due to mutual illumination. As noted in Chapter 6, the majority of the work that has been undertaken to investigate human shape perception from shading has been limited to the study of achromatic shading. In other words, the gradients resulting from the 3D shape of objects varied only in luminance. Very little work has considered the specific contribution to shape perception of chromatic gradients.

Some work has investigated the interactions of shape-from-shading and colour (Kingdom, 2003; Kingdom et al., 2005), suggesting that colour and 3D
shape-from-shading may be linked by means of natural colour-luminance relationships. Additionally, in the real world, mutual illumination between coloured surfaces creates chromatic as well as luminance gradients (see Figure 2.1). These chromatic gradients are potentially an additional source of information about the 3D structure of the scene (e.g. Funt et al. (1991)). Considering these previous studies, it seems possible that the visual system is capable of using chromatic gradients resulting from mutual illuminations to elicit a sense of 3D shape, without the addition of other cues (such as luminance gradients, or outline contour). If this is the case, we might also expect that information from this chromatic gradient cue would be combined with information from other shape cues in a similar way to that seen with other cues to 3D shape - see Chapters 3 (3D Shape Cue Integration) and 7 (Modelling the Integration of the Gradient and Outline Cues). Therefore, Experiment 3 was designed to investigate in more detail the effectiveness of chromatic gradients as a shape cue, and how perception is affected when chromatic and luminance gradient cues are combined.

9.1 Methods

The experiment consisted of three conditions that used different stimuli (Figure 9.1). Two ‘single-cue’ conditions tested how well observers could discriminate between stimuli defined by luminance gradients alone (‘luminance-gradient-only’ condition; Figure 9.1a), or chromatic gradients alone (‘chromatic-gradient-only’ condition; Figure 9.1c). This was done to measure the reliability of each cue in isolation. These ‘single-cue’ conditions were performed using a 2IFC discrimination experiment (see section 4.1). In
the ‘luminance-gradient-only’ condition stimuli were isochromatic and contained only luminance gradients. In the ‘chromatic-gradient-only’ condition stimuli were isoluminant and contained only chromatic gradients. By measuring discrimination between stimuli when only one cue is available, it is possible to compare how effectively each cue is used by the visual system to determine 3D shape. The results of the ‘single-cue’ condition can also provide an input to a MLE cue combination model as described in sections 3.1.1 and 7.2. Using the variance of the ‘single-cue’ conditions as the input, the model provides a prediction of variance and perceived shape when the cues are combined. This is achieved as follows: The predicted overall variance in perceived shape when both cues are available ($\sigma^2$) is dependent on the variance when each cue is available independently. In this experiment, the two cues under investigation are chromatic gradient and luminance gradients, with variances given by $\sigma_c^2$ and $\sigma_l^2$ respectively:

$$\sigma^2 = \frac{\sigma_c^2 \sigma_l^2}{\sigma_c^2 + \sigma_l^2}$$  \textbf{Equation 9.1}

The weightings ($w_c$ and $w_l$) of each cue are also related to the relative variance of the perceived shape when individual cues are used:

$$w_c = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_l^2}$$  \textbf{Equation 9.2}  \quad w_l = \frac{\sigma_l^2}{\sigma_c^2 + \sigma_l^2}  \textbf{Equation 9.3}

The mean predicted shape ($\mu$) is then given by the weighted combination of the shapes indicated by each cue separately, $\mu_c$ and $\mu_l$:  

213
\[ \mu = w_c \mu_c + w_l \mu_l \]  
Equation 9.4

In order to investigate the relative weightings of chromatic and luminance gradient cues to 3D shape when the cues are combined, a third ‘combined-cue’ condition was tested. This condition used an experimental method known as ‘cue perturbation’ (Hillis et al., 2004; Landy et al., 1995; Young et al., 1993). This method uses a two-interval forced choice (2IFC) paradigm, similar to that used for the ‘single-cue’ conditions, but where one interval contains a stimulus where there is a small conflict between the shapes indicated by each of the two cues under investigation (i.e. one cue is perturbed by a small amount). The other interval contains a stimulus with no conflict between the cues. By comparing the perceived shapes of the cue-conflict and no-conflict stimuli, the relative influence of the two cues can be measured.

Finally, the predictions made using the MLE model, based on the data from the ‘single-cue’ conditions, can be compared with the results of the ‘cue-conflict’ experiment, providing a measure of how well the model predicts perception when the cues are combined.
9.1.1 Apparatus

The stimulus viewing apparatus used for Experiment 3 was identical to that used in Experiment 2 and the monocular conditions of Experiment 1 (see Figure 5.8), with the exception that the separate monitor used in the previous experiments to display the matching lines of the observer task was not present. This monitor was not required due to the 2IFC discrimination task used in Experiment 3.

9.1.2 Stimuli

Because the aim of Experiment 3 was to examine the efficacy of luminance and chromatic gradients as cues to shape separately, we decomposed the hyperspectrally rendered chromatic Mach card images used for the Gradient-cue-only stimuli of Experiment 1 (see section 5.4) into separate luminance gradient and chromatic components. This allowed the creation of two ‘single-cue’ (‘luminance-gradient-only’ and ‘chromatic-gradient-only’) sets of stimuli. In addition, because the chromatic and luminance components of the rendered images were separated, it was possible to manipulate of each cue independently and subsequently re-combined the cues to create a third set of stimuli containing both cues with varying levels of conflict between angles indicated by each cue (‘cue-conflict’ stimuli). Stimulus angles ranged from 50 to 70 degrees, in steps of 2 degrees (where 0 degrees corresponds to a flat, frontoparallel card). The angle of the fold in the stimuli was chosen so that there was sufficient mutual illumination between the cards for the chromatic gradients to be clearly visible (for details of the visibility of chromatic gradients, see Ruppertsberg et al. (2008)).
To create the stimuli, the rendered images were converted, using calibration measurements for the CRT monitor used in the experiment, from RGB to CIE xyY colour space (CIE, 1931) and split, pixel by pixel, into separate luminance (Y) and chromatic (x,y) components. These components were then recombined in different ways to create the three stimulus sets: -

**Combined Cue-conflict stimuli:** Luminance (Y) and chromatic (xy) components from images of cards with slightly different angles (+/- 5°) were used to create a new image in xyY space that was subsequently converted to RGB colour space for display. Thus the resulting images contained conflicts between the card angles indicated by the chromatic and luminance information (Figure 9.1a)

**Chromatic-gradient-only stimuli:** The chromatic component (xy) of the original image was combined with a uniform luminance (Y=10cd/m²) and subsequently converted to RGB colour space for display. These new images were consequently isoluminant and contained only chromatic gradients (Figure 9.1b).

**Luminance-gradient-only stimuli:** The luminance component (Y) of the original image was combined with a uniform chromaticity (D65 white point, x=0.3127, y=0.3290) and subsequently converted to RGB colour space for display. These new images were consequently isochromatic and contained only luminance gradients (Figure 9.1c).
Figure 9.1. Example ‘folded card’ stimuli. All stimuli are folded in a concave ‘corner’ shape with an angle of 50° between the card surfaces and a horizontal axis perpendicular to the viewing direction. (a) Stimulus with both chromatic and luminance gradients as used in the ‘cue-conflict’ condition. (b) ‘Chromatic-gradient-only’ stimulus (isoluminant). (c) ‘Luminance-gradient-only’ stimulus (isochromatic). Note: due to the use of a viewing aperture, observers were unable to see the edges of the stimuli.

9.1.3 Observers

Five participants took part in the experiment, mean age 35, two male. Three were naïve to the purpose of the study, one a lab member (JB) and another, the author (GH). All observers had normal colour and stereoscopic vision, as well as normal or corrected to normal acuity. Colour vision was verified using the Farnsworth-Munsell 100 hue test (score below 100). Stereo acuity was verified with the TNO stereo test (120 seconds of arc or less). All observers had some prior experience of combined-cue folded card stimuli (without any cue-conflict) from the previous experiments, but other than the author (GH), they had not seen the chromatic-gradient-only or luminance-gradient-only stimuli before.
9.1.4 Single-Cue Conditions Procedure

The aim of the ‘single-cue’ conditions of Experiment 3 was to provide a measurement of observers’ ability to use each cue individually to discriminate between stimuli of different angles. These data were also used as an input for a MLE cue combination model. Two conditions were tested: ‘chromatic-gradient-only’ and ‘luminance-gradient-only’. In the chromatic-gradient-only condition stimuli were isoluminant (Figure 9.1b), and in the luminance-gradient-only condition stimuli were isochromatic (Figure 9.1c).

A two-interval forced choice discrimination paradigm was used together with the method of constant stimuli (see section 4.1 for details). One interval contained a fixed (standard) stimulus at either 55°, 60° or 65°, and the other interval a test (or comparison) stimulus chosen at random from a range of angles from 50 to 70 degrees, in steps of 2 degrees. The order of presentation of the standard and test stimuli was randomised for each trial. Stimulus intervals were one second long, with an inter-stimulus time of 0.5 seconds. After an initial 60 second period of adaptation to a 10cd/m² grey background, the experiment began and observers were asked to indicate, by pressing one of two response buttons, which interval contained the ‘deeper’ stimulus (see Figure 9.2). This method allows psychometric functions to be fitted to the data (by plotting the proportion of ‘deeper’ responses as a function of test stimulus angle). Observers responded to 40 trials for each pair of stimuli, and trials for each of the three tested stimulus angles were interleaved randomly, resulting in a total of 1440 trials, split over 8 sessions.
Figure 9.2 Time line of the two-interval, forced choice experiment. The 1st and 2nd intervals contain either the standard or test stimulus (random order). Observers must choose which interval contains the ‘deepest’ stimulus. Stimuli were either isoluminant (chromatic-gradient-only condition) or isochromatic (luminance-gradient-only condition).

9.1.5 Cue-Conflict Condition Procedure

The experimental procedure followed the ‘cue-perturbation’ technique described above in section 9.1. This procedure is very similar to the 2IFC discrimination technique used in the single-cue experiment (see Figure 9.2), but stimuli contained both cues combined, and the angles indicated by each cue in the standard stimulus could conflict, while those in the test stimulus were congruent.
Observers were tested using standard stimuli with one cue indicating an angle of 60°, and the other (perturbed) cue indicating either 55°, 60° (no cue-conflict), or 65°. The test stimulus did not have any conflict between the angles indicated by each cue, and was varied randomly between trials over the same range of card angles as before. Figure 9.3 describes the set of stimuli used in the experiment. In total, 5 different standard stimuli were used, 4 of which contained cue-conflict. These stimuli are represented in Figure 9.3 by the red/grey card symbols and the angles indicated by each cue in the stimulus are depicted by the position of the symbols in the figure. For example, the uppermost symbol in the figure describes a stimulus that has a luminance cue indicating 60° and a (perturbed) chromatic cue indicating 65°. The range of angles used for the test stimulus (no cue-conflict) is indicated by the dashed blue line.

Trials for each combination of cues in the standard stimulus were interleaved randomly, resulting in a total of 1440 trials, split over 8 sessions. To minimise any effect of learning between the three conditions, the cue-conflict and single-cue sessions were interleaved randomly for each participant.
Figure 9.3. Diagram of the stimuli used in the cue-conflict experiment. The red/grey folded card symbols show the stimulus angles indicated by the chromatic and luminance cues in the cue-conflict stimuli. The blue line shows the range of angles used for the no-conflict stimuli.

9.2 Results and Analysis: Single-Cue Conditions

As noted previously, the single-cue experiment provided a measurement of the reliability of each cue separately, giving an input to the MLE model. Cumulative Gaussians were fitted to the data to deliver psychometric functions to obtain an estimate of single cue variances ($\sigma_l^2$, $\sigma_c^2$) for the model (Equations 9.1 to 9.4). When examining results based on fitted psychometric functions, it is useful to be able to quantify the uncertainty within the data. For this purpose, a bootstrap technique was employed, using 1000 bootstrapped samples, to estimate 95% confidence intervals for the resulting variance parameters (see Wichmann & Hill (2001)) of the fitted cumulative Gaussians.
Figure 9.4. Example psychometric functions fitted to the data (circles) for a single observer (JB) in the luminance-gradient-only condition of the single cue experiment. Functions are shown for each angle tested. Shaded areas indicate 95% confidence intervals for the fitted functions. The point where the curves cross the dashed line is the point of subjective equality (PSE).

Figure 9.4 shows example fitted psychometric functions, with 95% confidence intervals, for a single observer (JB) in one of the single-cue conditions. The point of subjective equality (PSE) between the standard and test stimuli is where 50% of responses indicate the test stimulus as being ‘deeper’. It can be seen that the PSEs are close to expected value (either 55, 60 or 65 degrees, depending on the standard stimulus). For this observer and condition, variance reduces as the stimulus angle increases, indicating
increasing cue reliability. The variance of each of these psychometric functions is shown in Figure 9.5.

Figure 9.5 Variance, including 95% confidence intervals, of the example psychometric functions depicted in Figure 9.4.

Figure 9.6 shows the variance for all observers in both ‘single-cue’ conditions. Variance was typically lower for the two experienced observers (GH and JB). The trend of decreasing variance with increasing stimulus angle is evident for all observers and both luminance and chromatic gradients. For each observer and card angle, in the majority of cases, the 95% confidence intervals for chromatic and luminance gradient variances overlap, suggesting that the cues were of similar reliability, and importantly, that the chromatic cue alone can be used to determine object shape. It seems probable, therefore, that both cues will influence shape perception when combined in the cue-conflict experiment.
Figure 9.6 Fitted psychometric function variance for all observers in both conditions of the single cue experiment. Variance is shown for each observer separately, for 'luminance gradient only' (blue bars) and 'chromatic gradient only' (red bars) stimuli. Left: 55 degree stimuli; Centre: 60 degree stimuli; Right: 65 degree stimuli. Lower variance indicates a lower discrimination threshold. Error bars show 95% confidence intervals.

9.3 Results and Analysis: Cue-conflict Condition

The ‘cue-conflict’ condition measured the Point of Subjective Equality (PSE) between stimuli with and without cue-conflicts. This provides a means of determining how strongly each cue influences the perceived shape of the stimulus. As with the ‘single-cue’ conditions, cumulative Gaussians were fitted to the data to obtain psychometric functions, using 1000 bootstrapped samples to estimate 95% confidence intervals (Wichmann & Hill, 2001). Observers made comparisons between test stimuli and standard stimuli with either a fixed luminance cue and a perturbed chromatic cue, or a fixed chromatic cue and a perturbed luminance cue. The angle at which the test stimulus is perceived as deeper than the standard stimulus in 50% of trials
(0.5 point of the fitted cumulative Gaussian) determines the point of subjective equality (PSE) between the cue-conflict and no-conflict stimuli (see Figure 9.7). How far the PSE is shifted towards the angle indicated by the perturbed cue is dependent on the relative influence of each cue on the perceived shape. If the relative weights of the two cues are assumed to be constant over the small range of the cue perturbation, then using Equation 9.4 (and the fact that the combined weight must equal 1), cue weightings can be calculated from the location of the PSE (see Hillis et al. (2004)).

For example, if the luminance cue is consistent with an angle of 60° and the chromatic cue is perturbed by -5° to be consistent with an angle of 55°, then a PSE of 57° would indicate that the luminance and chromatic cues are weighted at 40% and 60% respectively. This example is shown in Figure 9.7.
Figure 9.7 Example psychometric function with fitted cumulative Gaussian, for a cue-conflict stimulus where the chromatic cue indicates 55° and the luminance cue 60°. The Point of Subjective Equality (PSE) will typically fall between the angles indicated by two cues in the cue-conflict stimulus.

Figure 9.8 shows the measured PSEs for each observer individually (blue circles) together with error bars indicating the estimated 95% confidence intervals. PSEs are seen to vary as the angle of the perturbed cue changes, suggesting that, as expected from the results of the ‘single-cue’ conditions, both cues have an influence on shape perception. The position of the PSE relative to the two dashed lines on the graphs indicates the relative weightings of the two cues. If the PSE is closer to the red dashed line, this suggests that the chromatic cue is more heavily weighted by the visual system when estimating shape. If it is closer to the black dashed line, the luminance cue is weighted more heavily. For the majority of observers, the PSE lay between the angles indicated by the two cues, suggesting that
luminance and chromatic information influence perceived shape to a similar extent.

Figure 9.8 also shows the PSEs predicted by the MLE model described in section 9.1 (green stars). These predictions are calculated using the variances of the psychometric functions measured in the ‘single-cue’ conditions. Error bars indicate 95% confidence intervals. In the majority of cases, the model made good predictions for the PSE that were not significantly different from the measured values (95% confidence intervals overlap).
Figure 9.8 Point of Subjective Equality (PSE) for each of the standard stimuli used in the cue-conflict experiment. Each column shows data from an individual observer. Top row: PSE for perturbed luminance cue (chromatic cue fixed at 60 degrees). Bottom row: PSE for perturbed chromatic cue (luminance cue fixed at 60 degrees). Blue circles show measured PSEs from the cue-conflict experimental condition. Green stars show PSEs predicted by the MLE model, based on the variances measured in the ‘single-cue’ conditions. Error bars for both measured and predicted PSEs show 95% confidence intervals calculated from 1000 bootstraps. The dashed red line indicates the expected PSE if the observer relied solely on the chromatic cue. The dashed black line indicates the expected PSE if the observer relied solely on the luminance cue.

Figure 9.9 shows measured variance in the ‘cue-conflict’ condition for each individual observer (blue circles). Error bars indicate 95% confidence intervals. As in the ‘single-cue’ conditions, variance was generally lower for
experienced observers GH and JB than for the three naïve observers. The green stars in Figure 9.9 show the calculated variance of the MLE model, based on the single-cue experiment measurements (Figure 9.6, Equation 9.1). This predicted variance was always lower than variance when either cue was available alone. The model provided a very good prediction of variance for observer JB. Observer GH showed higher variance for some stimuli than the model predicted. This may be because, as the author of the study, GH is likely to have sufficient familiarity with the single-cue stimuli to enable improved discrimination (and indeed GH had low variance in the single-cue condition compared to the other observers – see Figure 6.10). This leads to a low predicted variance (green stars in Figure 9.9) for the cue-conflict condition. An improvement in discrimination in the cue-conflict condition due to knowledge of the stimuli, and a corresponding reduction in the measured variance (blue circles in Figure 9.9), seems less likely since the more complex cue-conflict stimuli are probably more difficult to learn.

The three naïve observers (KP, RS, and LH) typically showed lower variance in the cue-conflict experiment than predicted by the MLE model. This is a surprising result because the model provides an optimal, minimum variance, estimate. However, the difference is not statistically significant in the majority of cases (95% confidence intervals overlap). Since the model is statistically optimal, it predicts estimates that are more reliable when both cues are available than for either cue separately. Because observers are as reliable as the model prediction, and sometimes more so, this result suggests that the chromatic and luminance cues are combined by a process that makes best
use of the available information to improve the reliability of estimates when both cues are present.

Figure 9.9 Combined-cue variance. Each column shows data from an individual observer. Top row: variance for perturbed luminance cue (chromatic cue fixed at 60 degrees). Bottom row: variance for perturbed chromatic cue (luminance cue fixed at 60 degrees). Blue circles show measured variance from the cue conflict experimental condition. Green stars show variance predicted by the MLE model, based on the variances measured in the ‘single-cue’ conditions. Error bars for both measured and predicted variance show 95% confidence intervals calculated from 1000 bootstraps.

As noted above, because ‘optimal’ cue integration makes best use of available information to reduce uncertainty, the MLE model predicted variance when cues are combined is always lower than when either cue is available individually. If the same occurs for the measured data, this would further support the idea the luminance and chromatic gradients are combined.
in an optimal way. Figure 9.10 show the variance in all conditions of the experiment for each observer.

**Figure 9.10** Measured variance for all conditions of Experiment 3. Each column shows data from an individual observer. Blue circles: luminance cue only. Green stars: chromatic cue only. Red diamonds: combined cues without conflict. Yellow crosses: combined cues, luminance cue perturbed (to either 55 or 65 degrees, chromatic cue fixed at 60 degrees). Magenta squares: combined cues, chromatic cue perturbed (to either 55 or 65 degrees, luminance cue fixed at 60 degrees). Error bars show 95% confidence intervals calculated from 1000 bootstraps.

Figure 9.10 shows that while variance in the combined cue conditions is often not statistically different from variance in the single cue conditions (95% confidence intervals overlap), the combined cue variances are lower in the majority of cases, suggesting that some form of optimal cue integration may occur.
9.4 Discussion

9.4.1 Luminance and Chromatic Gradients as Separate Cues to Shape

The results of the ‘single-cue’ conditions of Experiment 3 indicate that observers can discriminate, on the basis of shape, between stimuli containing only luminance gradients. This is not surprising – as previously discussed, many earlier studies have shown that humans can use achromatic shading as a cue to 3D shape. A more interesting, and new, result is the fact that observers could also discriminate between shapes defined by only chromatic gradients.

Figure 9.10 shows luminance and chromaticity profiles for the three types of stimuli used in Experiment 3 (‘luminance-gradient-only’, ‘chromatic-gradient-only’, and the combined-cue stimuli used in the ‘cue-conflict’ condition). The top row shows luminance profiles for a horizontal sample across the centre of each type of stimulus image for a selection of card angles (65, 60 and 55 degrees). These profiles show that for the stimuli that are not isoluminant (‘luminance-gradient-only’ and ‘combined-cue’) the luminance gradient has a higher slope at higher card angles (card less flat). On the centre and bottom rows, profiles are shown of the x and y chromaticity coordinates (CIE 1931), again for a horizontal sample across the centre of each type of stimulus image. For the stimuli that are not iso-chromatic, chromatic gradients along both the x and y chromaticity axis also have a larger magnitude (the chromaticity slope is higher) when the card is less flat.
Figure 9.10 Top row: luminance profiles. Centre row: CIE chromaticity x-coordinate. Bottom row: CIE 1931 chromaticity y-coordinate. Profiles are for a horizontal sample across the centre of the stimuli images, and shown for stimuli of 65 degrees (solid red line), 60 degrees (dashed green line) and 55 degrees (dotted blue line).


Because the physical gradients (both luminance and chromatic) in the stimuli have a higher slope at higher card angles, we might expect that the shape cues will be more reliable and discrimination variance lower for larger card angles. This is indeed the pattern seen in the results of the ‘single-cue’ conditions (see section 9.2 and Figure 9.6).
Although this change in variance with stimulus angle is in line with what would be expected if observers used the chromatic gradients as true visual cues to 3D shape, it remains possible that observers instead use some other more general colour properties of the ‘chromatic-gradient-only’ stimuli to discriminate between them. For example, observers may use the overall colour of the ‘white’ side of the card as a way to estimate the card shape: at higher angles the amount of mutual illumination is higher and the physical colour of the light reflected from the ‘white’ side of the card becomes redder. However, the gradients within the stimuli used should be visible to observers: at the base angle used in the experiment (60 degrees), the total cone contrast (Brainard, 1996; Chaparro et al., 1993) across the ‘white’ side of the combined-cue card is 39.8%, a contrast that considering previous work we would expect to be easily visible (Chaparro et al., 1993; Ruppertsberg et al., 2008). Additionally, the ‘uniform colour stimulus’ control of Experiment 2 (section 8.3.3) indicates that, at least in the case of combined luminance and colour stimuli, the gradients themselves, rather than average colour and/or luminance, provide the critical information in the ‘gradient’ cue. The results of main part of Experiment 2 also suggest that observers do not typically rely on learnt or memorised properties (such as mean colour) of this type of stimulus when estimating shape. For these reasons, it seems possible that observers do in fact use the chromatic gradients as shape cue, accounting for how such gradients arise due to mutual illumination and shape.
9.4.2 Luminance and Chromatic Gradient Cue Combination

It is clear from the PSEs shown in Figure 8 that observers do not solely rely on the luminance shading of the ‘cue-conflict’ stimuli to estimate shape. In nearly all cases where there is a cue-conflict, the PSE lies between the shape indicated by the luminance information and that indicated by the chromatic information. This result adds considerable weight to the conclusions made from the ‘single-cue’ data - that the chromatic gradient is used by the visual system as a shape cue.

The perception of shape when the luminance and chromatic gradients are combined was analysed using a ‘weighted average’ (MLE) weak fusion model. This model contains some assumptions that may not hold in this specific case:

First, the model assumes that the cues that are to be combined are processed completely independently. As noted section 2.1.4.1.2, previous work has indicated that under some circumstances spatial correlations between luminance and chromatic gradients appear to be important to how 3D shape and depth are perceived (Kingdom, 2003). Such interactions could reduce the validity of a model that does not allow any cue interaction. In Kingdom’s work, it is suggested that colour variations in the stimuli are perceived as due to surface reflectance changes, and luminance variations due to shape changes. However, in the case of the stimuli used here, the chromatic gradients are not due to any change in surface reflectance, but are linked to 3D shape just as the luminance gradients are. In this case (and
many others), the visual system would not best utilise the available information by assuming that all chromatic variation is due to surface reflectance changes. In fact, recent work has shown that Kingdom’s ‘colour shading effect’ does not occur for all observers (Clery, Bloj & Harris, 2013) even when colour variations are not linked to shape and are purely surface reflectance patterns, suggesting interactions between chromatic and luminance shading may not always occur. It therefore seems reasonable that when chromatic and luminance information do provide separate and informative shape cues (as is the case in the stimuli used here) that the visual system could use both cues and combine them in similar way to more disparate types of shape information, without significant interactions.

Secondly, if priors are present that have an influence on perceived shape from both cues, these priors should have a reduced influence on the perceived shape (relative to the cue information) when both cues are available, leading to less biased shape perception. The ‘weighted average’ type of model may implicitly contain the effects of priors (in the ‘single cue’ data), but these priors are necessarily also independent in the model. For this reason, the model cannot predict a reduction in the influence of prior information in the combined cue case. Because the earlier experiments described in this thesis suggest that priors do influence the perception of 3D shape in the folded card stimuli used, the MLE model may not be able to account for all the factors that influence shape perception in when the cues are combined. However, any shift in perceived shape due to an increase in total cue reliability when both cues are present affects both the standard and
test stimuli almost equally, because the cue conflicts are small. Therefore any effect on the results of the experiment should also be small.

As discussed in section 9.4.1, previous studies suggest that observers should to be capable of using the chromatic gradient information, but this, together with the fact that they can discriminate between ‘chromatic-gradient-only’ stimuli does not prove that they do so in this case. However, variance in the cue-conflict experiment, where both cues are combined, is in line with the statistically optimal MLE model (see Figure 9.9); variance is reduced when more cues are available, indicating that an optimal integration of shape information from luminance and chromatic gradients takes place within the visual system. This type of cue integration behaviour has been shown to occur for the combination of a number of other depth/3D shape cues (e.g. Hillis et al. (2004), Lovell et al. (2012), Oruc et al. (2003)). Therefore, it seems a reasonable conclusion that chromatic gradients form a visual cue to shape, rather than providing higher level information that can be used to discriminate the stimuli.
Figure 9.11 Cue weights based on the measured point of subjective equality from the cue-conflict experiment. Weights for each observer are shown in separate columns. Top row: luminance cue perturbed (chromatic cue fixed at 60 degrees). Bottom row: chromatic cue perturbed (luminance cue fixed at 60 degrees). Red bars: Chromatic cue weight. Blue bars: Luminance cue weight.

Cue weightings based on the measured PSEs from the ‘cue-conflict’ condition (Equation 9.4) are shown in Figure 9.11, and show that in many cases the chromatic cue was in fact weighted more heavily than the luminance cue. It should be noted however that, because the stimuli were deliberately chosen with angles that created significant mutual illumination, and therefore large chromatic gradients, such high weighting would not always occur. Nevertheless, this result does provide a strong case for chromatic gradients being a useful shape cue where mutual illuminations are present.
9.5 Experiment 3 Summary

The aim of Experiment 3 was, first, to investigate whether the human visual system is able to use the 3D shape information that is contained within chromatic gradients, particularly those that arise from mutual illuminations. Second, if chromatic gradients are an effective cue to shape, how are they combined with other shape cues?

Previous findings relating colour and 3D shape perception have suggested that chromatic gradients should provide a useful cue to shape, and the results of Experiment 3 confirm that this is the case. Observers could discriminate between stimuli of different shapes that were defined by only chromatic gradients, and shape perception was influenced by both cues when chromatic and luminance information conflicted. The results of many previous shape cue combination studies suggest that if chromatic gradients do provide a shape cue, then their integration with other cue information should occur in a way that is close to statistically optimal. The agreement between the predictions of the MLE cue-combination model and the results of the ‘cue-conflict’ condition indicate such a process does occur.

The results of Experiment 3 also show that not only do chromatic gradients appear to provide useful shape information, in some circumstances (such as when high levels of mutual illumination are present) this information may be as, or more, important than that obtained from the corresponding luminance gradient.
Chapter 10 - General Discussion

The experiments described in this thesis were designed to investigate the role of realistic shading, containing both luminance and chromatic gradients, in three-dimensional shape perception. Such gradients arise due to the complex interactions of lighting, object reflectance properties and three-dimensional shape. A large body of literature exists that details studies of this ‘shape from shading’ phenomena, but little of this work has considered the more complex shading effects that occur due to object inter-reflections, and in particular the chromatic gradients that can arise from these ‘mutual illuminations’. It is these colour shading gradients that are the focus of the work in this thesis. In the following sections I will summarise the findings from the experimental and modelling work contained in this thesis and discuss the relevance of the findings in relation to previous studies and potential future work.

10.1 Summary of Experimental Findings

10.1.1 Shape Perception from Realistic Complex Colour Gradients

The first experiment described here (Experiment 1) was designed to investigate how well the human visual system is able to use complex colour shading gradients as a cue to 3D object shape. In this experiment, shape perception was tested for stimuli containing realistic gradients with both luminance and chromatic components, either in isolation or in combination with outline (perspective) or outline and binocular disparity shape cues. The
aim of Experiment 1 was twofold: first to determine if visual system is able to account for the very different luminance and chromatic gradients present on convex and concave versions of the ‘chromatic Mach card’ stimuli used and accurately determine shape in each case, and second, to understand how gradient shape cues might be integrated with other types of shape information.

The results of Experiment 1 showed that gradients were a less reliable cue to shape than object outline. This finding is consistent with other studies that have compared shading (although in these previous studies shading typically consisted of luminance gradients only) with other 3D shape cues (e.g. Bültzoff & Mallot (1988), Todd, Norman, Koenderink, & Kappers (1997), Vuong et al. (2006)). Experiment 1 also confirmed the findings of some previous studies of 3D shape and surface slant perception: observers typically display a large bias, setting angles that are in general considerably lower than the stimulus angle. This flattening of perceived shape is similar to that found by a number of other authors (Adams & Mamassian, 2004; Bültzoff & Mallot, 1988; Mitchison & Westheimer, 1984; Todd et al., 2005; van Ee et al., 2003).

More uniquely, Experiment 1 showed that observer’s perception of shape when given only the shading gradients as cues to shape was very similar for concave and convex versions of the stimulus. This is despite mutual illumination creating significantly altered luminance gradients and additional chromatic gradients for concave shapes only. In addition, variance was lower
for concave stimuli with high angles (where mutual illumination is most significant) than it was for convex stimuli at high angles, suggesting that the visual system makes use of the additional 3D shape information available through the presence of mutual illumination in the concave stimuli – something that computational studies have suggested might be possible (Forsyth & Zisserman, 1990; Funt et al., 1991; Nayar et al., 1991). These are interesting results because it indicates that the visual system is capable of solving the difficult problem of accounting for the complex shading effects that occur when inter-reflections of light are present, and making use of all the shape information present in the complex shading. These findings could only be arrived at because of the use of realistic stimuli with accurate colour shading properties.

If the visual system combines the gradient cue with other shape information in a similar way to that seen with many other combinations of shape cues, then observers’ perception might be expected be more reliable when additional cues are also available (e.g. Hillis et al. (2004), Lovell et al. (2012), Oruc et al. (2003)). However, while observation variance was lowered when the gradient cue was combined with either an outline (perspective) cue, or outline and binocular disparity cues, observers’ variance did not reduce below that seen when the outline cue was used in isolation. It is possible that the gradient cue is integrated with other shape cues in a way that is close to the statistically optimal predictions of many cue-combination models, but increased reliability is difficult to see because the gradient cue is weak in comparison to the other cues used in Experiment 1. However, this could also
be evidence of less optimal cue-combination strategies, such as the cue vetoing suggested by some other studies (Bülthoff & Mallot, 1988; Dosher, Sperling, & Wurst, 1986; Mather & Smith, 2000).

Comparing the settings made by observers in Experiment 1 when differing numbers of cues were available yields one further interesting result - one that indicates behaviour is not consistent with the popular ‘weighted average’ weak fusion cue-combination models: when more cues are available, observers make more veridical settings and the ‘flattening’ bias is reduced. This cumulative effect when combining 3D shape or depth cues has been noted in a few previous studies (Bruno & Cutting, 1988; Bülthoff, 1991; Bülthoff & Mallot, 1988; Hiroyasu, Katusunori, & Shin’ya, 2001).

Interestingly, while a ‘strong fusion’ model of cue integration, where interactions occur between systems processing information from each cue, could explain this behaviour, such an approach is not necessary. This behaviour can also be explained by a statistically optimal model of cue integration based on Bayesian inference that does not allow cue interaction, but takes into account the use of prior knowledge within the visual system. This is because the biasing effect of priors may be reduced as more visual cue information becomes available.

10.1.2 Disambiguating Complex Colour Gradient Cues - Are ‘Priors’ Enough?

Before observers took part in Experiment 1 they were shown a short, 30-second, training video showing the stimulus (with both gradient and outline
cues) moving through the full range of angles that were to be tested. This video was initially intended simply to train the observers in the magnitude estimation task that was used in Experiment 1. However, because the video allowed the observers to gain some information about the link between the gradient cue, outline cue and object shape, the results of Experiment 1 could not be fully examined without considering further the effects the training video might have. For this reason Experiment 2 was designed specifically to investigate what prior information, if any, observers require in order to make consistent shape settings.

Because many previous studies have shown that humans are able to use the type of perspective cue to surface slant (which is what the outline cue consisted of in the stimuli used here) without any training (e.g. Ames (1951), Clark et al. (1955), Olson (1974), Stevens (1981), Stevens & Brookes (1987), Zimmerman et al. (1995)), the stimuli of Experiment 2 did not contain any outline cue. Similarly, the binocular disparity cue available in some conditions of Experiment 1 was disposed of for Experiment 2. Therefore the stimuli of Experiment 2 contained only the gradient cue to 3D shape, and the conditions of this experiment differed only in the type of training observers received before making shape settings. In this way an investigation of the prior information needed to use specifically the gradient cue was possible.

The results of Experiment 2 showed that if the human visual system contains prior knowledge or assumptions about lighting or surface reflectance properties that might help in the use of shading gradients as cues to 3D
shape, then they are not sufficient to enable observers to set consistent shape settings for the chromatic Mach card stimuli when only the gradient cue is available. Some form of prior information or suitable assumptions are required to make use of the gradient cue, since it is inherently ambiguous. Experiment 2 showed that observers must specifically learn this information, but that only a small amount of training is enough to enable observers to gain sufficient information about the scene to disambiguate the gradient cue and make consistent shape settings. A control experiment also confirmed that observers required the gradient information in the shading, rather than only the spatially averaged luminance and chromaticity, to make these consistent shape estimates. The similarity between settings made using the very different gradient and outline cues alone, suggests that the visual system is capable of quickly making reasonable assumptions about the scene parameters, such as light direction and object surface reflectance in order to solve the gradient-shape relationship (in a similar way to many 'shape-from-shading' algorithms, e.g. Horn (1975)), rather than simply learning an arbitrary mapping between the gradients in the training videos and object shape. This conclusion is supported by the fact that majority of observers in Experiment 2 who were deliberately trained with an incorrect mapping between shape and shading gradients still made similar shape settings to those trained with the correct gradients.

The finding that observers seem to make assumptions about scene in order to solve the complex illumination problem that relates shading and shape is in contrast to the work of a number of other authors who have suggested that
the visual system might instead rely on a process dependent on the global properties of shading on 3D shapes (e.g. Erens et al. (1993), Koenderink & van Doorn (1980), Mingolla & Todd (1986), Nefs et al. (2006)). However, considering that other previous studies have shown that humans are sensitive to changes in light position in 3D scenes (Gerhard & Maloney, 2010; Maloney, 2002) and the resulting changes to complex shading gradients (Ruppertsberg et al., 2008), and may estimate illuminant position (Bloj et al., 2004; Boyaci et al., 2003), the suggestion that observers may make use of such information or assumptions about the wider scene in order to use the gradient cue should perhaps not be surprising.

10.1.3 Chromatic Gradients as an Independent Shape Cue

The results of Experiments 1 and 2 indicated that the human visual system is capable of accounting for complex interactions of 3D shape and illumination, such as mutual illuminations. In the concave versions of the stimuli used, mutual illumination created chromatic gradients, as well as altered luminance gradients. Because very little of the work that has investigated human shape-from-shading capabilities has included realistically coloured stimuli, whether these chromatic gradients are useful to the visual system as shape cues has remained unknown, although theoretical work has suggested that colour information in mutual illuminations might be useful for separating the confounding effects of illumination, reflectance and shape; e.g. (Funt et al., 1991).
Experiment 3 was therefore designed to investigate whether the human visual system is able to use the 3D shape information that is contained specifically within chromatic gradients from mutual illumination, even when luminance cues are removed. By splitting the Gradient-cue-only stimuli of Experiments 1 and 2 into chromatic-gradient-only and luminance-gradient-only stimuli, and using a two-interval forced choice paradigm for Experiment 3, it was possible to measure how well observers could discriminate between stimuli at different angles using only chromatic or luminance gradients. This yielded a unique result – observers were capable of differentiating 3D shape based on chromatic gradients only.

The second part of Experiment 3 used a cue-perturbation procedure (Young et al., 1993) that is well known, but never before applied to measure the weights assigned to the chromatic and luminance components of colour shading gradients. This second part of the experiment confirmed that chromatic gradients are indeed useful as shape cues – shape perception was influenced by both the chromatic and luminance cues. Further to this, in some cases chromatic gradients were found to be given an equal, or even greater, weight than the luminance component. This is an important result considering the lack of attention previously paid to chromatic gradients as 3D shape cues.

One of the few previous studies to investigate colour gradients as a 3D shape cue concluded that isoluminant colour gradients that changed in saturation, but not hue, could be effective at conveying surface slant
(Troscianko et al., 1991). These colour saturation gradients are similar to the chromatic gradients present in the chromatic-gradient-only stimuli of Experiment 3, and typical of the type of gradients created by mutual illuminations. It seems likely that the reason the visual system is able to use gradients in saturation as 3D shape cues is because of the link with shape through mutual illumination.

In the analysis of the data from Experiment 3, I also considered if chromatic gradients are combined with other shape cues in a similar way to that seen in previous cue-combination studies. By using a ‘reliability weighted average’ (MLE) model (Ernst & Banks, 2002) and the discrimination data for chromatic-gradient-only and luminance-gradient-only stimuli, it was possible to predict perception when the cues were combined in the second part of the experiment. The model provided quite a good account of both the perceived shape and variance when the cues were combined, suggesting that chromatic and luminance gradients are processed separately and the information from each cue combined linearly and in a way that is near to statistically optimal. This result is in broad agreement with many previous studies that have investigated shape cue combination for other cue types (e.g. Ernst & Banks (2002), Hillis et al. (2004), Knill & Saunders (2003), Landy et al. (1995), Lovell et al. (2012), Oruc et al. (2003), Young et al. (1993)). Experiment 3 as a whole suggests that both chromatic and luminance gradients can be used by the visual system, individually and in combination with other cues, much like other sources of 3D shape information.
10.2 Findings from the Modelling

The modelling of 3D shape perception, particularly when multiple cues to shape are available, has been the focus of many studies over recent years, with the hope that such theoretical and computational work can provide a useful description of the processes that occur within the visual system. If mathematical models can predict the results of experiments that probe shape perception, then perhaps the mechanisms by which the visual system generates our three-dimensional perception of the world can be deduced. The relatively simple ‘weighted average’ (MLE) weak fusion model (Ernst & Banks, 2002) was used to try and predict the results of cue-combination in both Experiments 1 and 3. As discussed above, in the case of Experiment 3, the MLE model provides a good description of observer behaviour, indicating that the visual system uses a cue-combination mechanism that processes information from each available cue separately, before combining shape estimates in a way that makes optimal use of the information available to reduce uncertainty. While this result is useful, and in agreement with several other studies (see discussion in section 10.1.3 above), there are some significant limitations to this type of perceptual model when we try to explain data from magnitude estimation, rather than discrimination, experiments (e.g. Experiment 1): The MLE model is unable to provide any explanation of biases in the ‘single-cue’ data of Experiment 1 (specifically the underestimation of stimulus angle and increased variance at higher stimulus angles), or predict the more veridical settings made when cues are
combined. Such a model can only provide at best a simplified description of the processes involved.

Indeed the MLE model is mathematically equivalent to a specific variation of the more complex class of Bayesian models (that is: normally distributed likelihoods and priors, and a *maximum a posteriori* decision rule, although importantly with a prior that is not restrained to be the same for all shapes). These Bayesian models can explicitly account for biases in perception due to the factoring of prior information into shape estimates, and the work in Chapter 7 shows how a prior distribution with a peak probability for flat shape, or frontoparallel surfaces, can explain the underestimation bias seen to the results of Experiment 1. This type of bias has also been seen in some other 3D shape perception studies (e.g. Adams & Mamassian (2004), Hillis et al. (2004), Mingolla & Todd (1986), Mitchison & Westheimer (1984), van Ee et al. (2003)).

However, the situation is not quite as simple as it first appears. A Bayesian model is not capable of accurately describing the ‘single cue’ data using either the outline or gradient cues, unless the model has either a non-Gaussian prior, or non-Gaussian likelihood functions (or both). While non-Gaussian priors have been demonstrated before (Stocker & Simoncelli, 2006), we should not actually expect likelihood functions to be Gaussian in form, if cue reliability is not fixed (Hogervorst & Eagle, 1998).
A Bayesian model with non-Gaussian likelihood functions for the outline cue, based on the relative lengths of the vertical edges and the central fold of the visible 2D projection of the 3D shape, was shown to provide the best fit to the single-cue data of Experiment 1. This ‘Relative-Side-Length’ method of generating the outline cue likelihood functions allows the model to replicate the increase in variance seen in the Outline-cue-only condition of Experiment 1 when the stimulus angle is larger, suggesting that the relative length of the vertical card sides and central fold may be the true cue in the card outline that is used by observers to perceive the 3D shape. In addition, a separate model that implemented a heavy-tailed distribution for the prior was shown to be able to reproduce the more veridical shape settings that observers made at higher stimulus angles, indicating that observers’ priors for flattened shape (or fronto-parallel surface orientation) may not be normally distributed.

The findings described above were made with models that were fitted to the data while allowing different likelihood function widths at each stimulus angle and were therefore quite complex. For this reason, a simplified model using a single ‘Relative-Side-Length’ outline cue likelihood for all stimulus angles, a single Gaussian likelihood function for the gradient cue at all angles, and a heavy-tailed prior distribution was created that had only four free parameters for all 15 stimuli angles and both cues. The ability of this simplified model to fit the ‘single cue’ condition data of Experiment 1 was only slightly reduced by the large reduction in free parameters and the simplified model is much preferred when assessing its performance using the Akaike Information Criterion, which takes into account the number of free parameters. The
simplified model was able to provide a better prediction of observers’ shape settings made for stimuli where outline and gradient cues were combined, for the majority of observers, than the other Bayesian models tested and a ‘weighted average’ weak fusion model.

One final point of interest regarding Bayesian modelling is that this type of model is capable of predicting the more veridical angle settings (reduced underestimation bias) made by observers in Experiment 1 when more cue information is available – something that is seen when gradient and outline cues are combined, binocular disparity is added and when cue-rich real card stimuli are used. If a single prior influences shape estimates made using different cues when they are available in isolation, then when these cues are combined the overall influence of the prior, compared to the cue information, will be lower. Therefore Bayesian models predict more veridical, less biased, perception when more cues are available. This prediction cannot be replicated by the popular ‘weighted average’ type of weak fusion model, indicating that such an approach is unsuitable when priors have significant influence on perceived shape, and that this particular type of model is unlikely to represent the true cue combination process.
10.3 Concluding Remarks and Suggestions for Future Research

The experimental work described in this thesis demonstrates that the human visual system is capable of accounting for the complex interactions of illumination, surface reflectance and 3D object shape, and to some extent can disentangle these factors in real three-dimensional scenes. Humans can accurately estimate 3D object shape when complex colour shading is the only information available, provided they have some opportunity to assess the scene and conclude a reasonable set of assumptions about object and lighting properties. We do not perceive erroneous 3D shape when presented with shading that contains mutual illuminations, and further to this, can use the additional information available when mutual illuminations are present to make more reliable estimates of shape. The visual system also appears to treat the luminance and chromatic gradients in colour shading as separate cues to shape, and can use either in isolation, or combine them optimally. Because mutual illumination is a feature of more complex three-dimensional scenes, where light is reflected between the surfaces of different objects, the ability to use mutual illuminations and the resulting chromatic gradients as effective cues to 3D shape allows the visual system to more effectively build an accurate representation of the relative positions of object and surfaces within such (real world typical) complex scenes than would otherwise be possible. Gilchrist & Jacobsen (1984) showed how achromatic inter-reflections of light in complex 3D scenes can be crucial to correctly identifying major aspects of the scene, and the work presented in this thesis suggests that the chromatic content of inter-reflected light is also important.
The modelling of the visual system undertaken in an attempt to explain the experimental data confirms the conclusions of a number of previous studies - that weak fusion models, where cues are processed separately, may be adequate in many cases to predict perception when multiple 3D shape cues are combined. However, more sophisticated models based on Bayesian inference have the potential to provide a much more thorough explanation of the underlying processes involved in shape perception in general. Although Bayesian modelling of shape perception is a technique with apparently significant explanatory power, we must be careful in how we use this type of model, because of the very large number of ways that the models can be constructed and used to explain differing data.

The findings from the experiments and modelling described within this thesis go some way towards answering the question of whether the human visual system can make use of the complicated luminance and chromatic gradients that occur as part of object shading in the real world to estimate three-dimensional shape. However, there still remain some significant gaps in our knowledge: Further experiments that examine the combination of chromatic gradients with other unrelated shape cues (for example texture or binocular disparity, rather than the luminance gradients use in Experiment 3) might be useful to confirm that the chromatic components of mutual illuminations are truly treated in the same way as other shape cues. In addition, new experiments will be needed to understand more precisely what information the visual system must learn in order to use isolated gradient cues (and can
a process like the ‘cue-promotion’ of Landy et al. (1995) replace learning when other cues are available). Finally, as the need to learn how to use the gradient cue reminds us, object shape perception is not merely passive, but a process that makes use of learnt information about the three dimensional structure of the world, acquired over both short and long time scales. For shading gradient cues, the assumptions, or acquired information, needed to make use of the cue information are often complex in comparison to other monocular shape cues (e.g. perspective). It is perhaps for this reason that brief experience of a scene, to gather information such as lighting and surface reflectance properties, seems to be required to make use of gradient cues in some situations. Entirely naïve observers sometimes cannot make use of shading gradients in the same way as observers who have had this opportunity to learn, and further research is needed to fully understand how the learning process occurs, and how the information is integrated with the current sensory input.
References


Garcia-Suarez, L., & Ruppertsberg, A. I. (2010). Why higher resolution graphics cards are needed in colour vision research. *Color Research & Application* (Accepted for publication).


