UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL AND HUMAN SCIENCES

Psychology

Vision as Informed by the Statistics of the Natural Environment

Jennifer Anne Elizabeth Josephs

Thesis for the degree of Doctor of Philosophy

December 2013

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL AND HUMAN SCIENCES

Psychology

Thesis for the degree of Doctor of Philosophy

VISION AS INFORMED BY THE STATISTICS OF THE NATURAL ENVIRONMENT

Jennifer Anne Elizabeth Josephs

The natural statistics approach is based on the theory that there is a strong relationship between perceptual systems and the statistical properties of the environment they evolved in. The experiments contained within this thesis were varied in focus and methodology, but were all centred on characterising this relationship. Psychophysical observations suggest that human observers can exploit regularities and differences in amplitude spectra to perform "high level" tasks such as scene categorisation. In two carefully controlled priming experiments, I found no evidence to support this idea. In another experiment, I investigated whether observers could use the sign or magnitude of 2D contour curvature to influence estimates of metric depth near the boundary. Evidence demonstrated that participants were significantly influenced by the magnitude of contour curvature, suggesting that observers can use the learned relationship between 2D and 3D surface properties to recover depth when other cues are unreliable. Lastly, I developed a novel stimulus and response set-up that allowed recording of an observer's perception of slant about multiple tilt axes. Using this novel technique, it was demonstrated that observers demonstrate a frontoparallel prior. This work suggests that measurement of the statistical regularities in natural scenes has the potential to explain many aspects of visual performance. For future work, a complete characterisation of these relationships will require systematic measurements of the properties of the natural environment and combining research knowledge across carefully controlled, novel psychophysical tasks.

Contents

ABSTRACT ..... **I**

CONTENTS ..... **III**

LIST OF TABLES ..... **IX**

LIST OF FIGURES ..... **XI**

DECLARATION OF AUTHORSHIP ..... **XXIII**

ACKNOWLEDGEMENTS ..... **XXV**

CHAPTER 1 ..... **1**

Vision as informed by the statistics of the natural environment **1**

1.1 Vision as problem solving 1

1.1.1. Theoretical approaches to understanding visual perception 1

1.1.2. The ecological approach to understanding visual perception 4

1.2 Outline of methods in natural scenes research 6

1.2.2. Encoding efficiency and image statistics 6

1.2.3. Scene statistics and perceptual biases 7

1.3 Perceptual systems shaped by the environment 7

1.4 Image statistics 11

1.4.1 Regularity and Redundancy 11

1.4.2. Internalised statistics 13

1.5 Efficient coding 13

1.6 Scene statistics 14

1.7 Priors 17

1.7.1. The origin of priors 20

1.8 Correlations between images and physical scenes 21

1.9. Behavioural studies in natural statistics research 24

1.10 Future research 25

CHAPTER 2 ..... **27**

Suppression, but not Facilitation, of amplitude statistics in natural image categorisation **27**

2.1. Introduction 27

2.1.1. Methods of studying the use of amplitude in scene categorisation 29

2.2 Experiment 1 35

2.3. Methods 35
2.3.1. Participants 35
2.3.2. Apparatus 35
2.3.3. Stimuli 35
2.3.4. Procedure 38
2.4. Results 38
2.5. Experiment 2 41
2.6. Method 42
2.6.1. Procedure 42
2.7. Results 42
2.8. Experiment 3 45
2.9. Method 45
2.9.1. Procedure 45
2.10. Results 45
2.11. Experiment 4 47
2.12. Method 48
2.12.1. Procedure 48
2.13. Results 48
2.13.1. Threshold analysis 48
2.13.2. Response Time Analysis 48
2.14. Discussion 51
2.14.1. Amplitude Suppression 51
2.14.2. Object Recognition 52
2.14.3. Amplitude-Phase Interaction 53
2.14.4. Amplitude and Response Time 54
2.14.5. Conclusion 54
CHAPTER 3 56
Can prior exposure to characteristic scene amplitudes tip the balance in binocular rivalry? 56
3.1. Introduction 56
3.1.1. Binocular Rivalry 58
3.1.2. Motivation Behind Binocular Rivalry Approach 59
3.1.3. Study Outline 60

3.1.4. Conditions and Predictions 61
3.2. Experiment 1 63
3.3. Method 63
3.3.1. Participants 63
3.3.2. Apparatus 64
3.3.3. Stimuli 64
3.3.4. Procedure 64
3.4. Results 66
3.5. Discussion 68
3.6. Experiment 2 70
3.7. Method 70
3.7.1. Procedure 70
3.8. Results 71
3.9. Discussion 73
3.8. General Discussion 74
3.8.1. Overview of findings 75
3.8.2. The indirect role of unlocalised amplitude in scene perception 76
3.8.3. Why is the visual system sensitive to amplitude statistics? 78
3.8.4. Conclusion 79
CHAPTER 4 81
Inferring 3D surface shape from 2D contour curvature 81
4.1. Introduction 81
4.2. Method 86
4.2.1 Participants 86
4.2.2.Apparatus 86
4.2.3. Stimuli 87
4.2.4. Procedure 89
4.3 Results 90
4.4 Discussion 92
CHAPTER 5 97
Measuring biases in slant and tilt perception 97
5.1. Introduction 97
5.2. Methodological Considerations 105

5.2.1. Stimulus presentation 105
5.2.2. Recording of attitude responses 106
5.2.3. High- and low reliability stimuli 108
5.3. Outline of Current Experiment 109
5.4 Possible Outcomes 111
5.5. Method 113
5.5.1. Participants 113
5.5.2. Apparatus 113
5.5.3. Procedure 118
5.6. Results 120
5.6.1. Data Handling 120
5.6.2. Calculation of Perceived Attitude from Set Attitude 122
5.6.3. Haptic response bias in Part 1: Full-cue condition 122
5.6.4. Perceived attitude in Part 2: Limited cue condition 125
5.6.5. Data Analysis 129
5.7. Discussion 138
5.7.1. Do observers have a prior for slant? 138
5.7.2. Haptic response biases 139
5.7.3. Benefits and Limitations of the Method 140
CHAPTER 6 143
General Discussion 143
6.1. Findings, Implications and Limitations 144
6.1.1. Chapters 2 and 3 144
6.1.2. Chapter 4 145
6.1.3. Chapter 5 146
6.2. Future Directions 147
6.2.1. Perceptual biases 147
6.2.2. Image statistics 148
6.2.3. Scene statistics 148
6.3. Applications and Collaborations 149
6.4. Conclusions 150
REFERENCES 151
APPENDIX 163

Slant-Tilt Project Manual...163

List of Tables

Table 1 132

Table 2 133

List of Figures

Figure 1.1 Left: The Necker Cube (see text). Right: The Ponzo illusion. The two red lines are the same size on the retina, yet the converging lines made by the tiles in the scene are a cue to depth that makes the right line look longer. 2

Figure 1.2 Demonstration of high and low spatial frequency in images with sinusoidally changing luminance variations. 6

Figure 1.3 Limits of colour perception in humans and honeybees. Figure adapted from Goldstein, 2007..... 8

Figure 1.4 Contrast sensitivity function (CSF) for an adult observer. Human ability to perceive high spatial frequencies is finite, ability drops sharply after about 5 cycles per degree of visual angle. Adapted from Banks and Salapatek (1978)..... 9

Figure 1.5 The amplitude spectra for 6 natural images (averaged across all orientations). The spectra are plotted on log coordinates to show the 1/f slope found in frequency analysis of natural scenes. Adapted from Field (1987). 9

Figure 1.6 A demonstration of elevation and azimuth. The laser range scanner can rotate about the azimuth and elevation to obtain a 3D map of distances from objects in the scene.....15

Figure 1.7 (a) The specific distance tendency describes how participants generally assume objects are about 2-4m away, this means they both over and underestimate physical distances. The equidistance theory (b) describes how observers tend to assume two neighbouring patches in an image are at the same distance away from them, yet in reality their physical distances can vary. Adapted from Yang and Purves (2003b). 16

Figure 1.8 Demonstration of perception as Bayesian inference (see text for example). 18

Figure 1.9 Demonstration of 'light from above' prior using shape from shading. Rotating each disc by 180° will reverse the sign of the shape, though the global image properties remain constant. 19

Figure 1.10 The bottom left image shows the range information for the corresponding full colour image above. The images on the right show how straight line templates were used to identify different angles within the image. Adapted from Howe & Purves (2002 and 2005). 23

Figure 2.1 A square wave (F) can be constructed by adding several different components (A to E) that have the correct phase and amplitude. If one were to alter the phase of some of the sinusoidal gratings, then the contours in F would lose their definition. Note that this example only uses a single orientation, whereas images of natural scenes are much more complex and can be deconstructed into gratings of any orientation. Figure adapted from Sekuler & Blake (2002). 28

Figure 2.2 Figure adapted from Torralba & Oliva (2003), summarising average power spectra across multiple images from natural/manmade and basic scene categories. A) and C) show averaged power spectra for natural and manmade scenes respectively. Contrast energy is plotted along the z-axis, with larger peaks indicating higher amplitude; spatial frequency increases from the centre of the plot (units in cycles per degree of visual angle). B) and D) are contour plots representing this same information, with orientation plotted about 360° (though plots are rotated 90° so that contrast energy at the horizontal component is displayed vertically). Contours represent 50% (inner contour) and 80% (outer contour) of cumulative contrast energy; contours close to the central point indicate high contrast energy at low spatial frequencies. E) Spectral signatures representing average power spectra for several basic scene categories. Contours represent 60, 80 and 90% of contrast energy. 30

Figure 2.3 Top: original natural and manmade scene. Middle: images adjusted to contain the average luminance and global contrast across the entire 1152 image set. Bottom: Fourier analysis was used to calculate amplitude spectra of each image and the average across the entire set; ELA stimuli (equalized amplitude spectrum) were created by pairing image phase with the average amplitude and then applying an

inverse Fourier transform Adapted from Joubert, Rousselet, Fabre-Thorpe & Fize (2009).....32

Figure 2.4 A city scene manipulated to contain different levels of phase noise from 0 (original image) to 1 (phase entirely randomised). Adapted from Loschky & Larson (2008).....33

Figure 2.5 Example manmade (left) and natural (middle) image with equalised luminance and global contrast and average amplitude across entire image set. Example test image (right), showing a hybrid of the two individual images, with 50% natural, 50% manmade phase. 37

Figure 2.6 Schematic diagram of priming/adaptation trial from Experiment 1. An example caricature stimulus is shown for each prime category. The test image shown contains 50/50 phase blend and amplitude information averaged across the entire image set.....37

Figure 2.7 Psychometric fits across averaged data (*N* = 8) in Experiment 1 for baseline and priming conditions (6 durations) for (a) natural and (b) manmade primes; the x-axis represents the proportion of manmade phase in test images, from 25% manmade, 75% natural, to 75% manmade, 25% natural; circles indicate 50% categorisation threshold. (c) Bars represent the difference in threshold between each prime duration and baseline. Error bars indicate the standard error of the mean (±1 *SEM*). 39

Figure 2.8 Psychometric fits across averaged data (*N* =11) from Experiment 2 for baseline and priming conditions for (a) natural primes and (b) manmade primes. (c) Bars represent the difference in threshold between each inversion condition and its baseline, with data from the shortest prime condition (35ms) from Experiment 1 (*N* = 8) shown for comparison. Priming condition is plotted on the x-axis. Error bars indicate ±1 *SEM.* 43

Figure 2.9 A and B: pair of intact images, followed by example test images from Experiments 3a and b, with phase blend from 75% natural, 25% manmade, to 25% natural, 75% manmade, paired with natural, averaged or manmade amplitude.

Amplitudes in 4a are category averages, while amplitudes in 4b originate from either the natural or manmade intact image, or an average of the two. C and D: Psychometric fits across averaged data from Experiments 3a (*N* = 7) and 3b (*N* = 5) respectively. Fits are shown for average, natural and manmade amplitude categories. 47

Figure 2.10 Psychometric fits across averaged data for Experiment 4 (*N* = 11). Fits for baseline and priming conditions for (a) natural primes and (b) manmade primes. (c) Bars represent the difference in threshold performance in the reaction time task for each experimental condition relative to the baseline. Error bars indicate ±1 *SEM.* 49

Figure 2.11 Response time data from Experiment 4, averaged across participants (*N* = 11) for (a) 1-frame prime condition and (b) 3-frame condition. Bars represent correct responses in baseline conditions (no prime), and prime conditions in which the prime was either consistent (CS) or inconsistent (IC) with the dominant phase of the test. Bars are labelled by correct response (natural/manmade) and by prime/test consistency. Error bars indicate ±1 *SEM.* 50

Figure 3.1 When the patchwork stimuli shown below are presented to different eyes during binocular rivalry, instead of seeing just one eye's image as would be expected in eye rivalry, sometimes participants perceived only red or only green spots, indicating that rivalry was resolved by pattern coherence. Adapted from Kovács et al., 1996. 59

Figure 3.2 Schematic diagram of procedure. Here, primes (all natural) are depicted at 7%, 20% and 50% contrast for match, mismatch and amplitude component primes respectively. Bottom prime stimulus shows white noise (baseline) at 90% contrast. 66

Figure 3.3 A to C: City scene dominance plotted as a function of RMS contrast for forest and city scene primes for each of three conditions (A: Match, B: Mismatch and C: Amplitude Component) and baseline, averaged across all participants (*N*=8). D: Difference in city scene dominance between city prime and forest prime for each condition. Error bars represent ±1 S.E.M 67

Figure 3.4 The figures show the proportion of city scene dominance as a function of RMS contrast for a) phase-match and baseline, and b) identical-amplitude conditions.

Figure 3.4c shows the difference in city scene dominance between city prime and forest prime for each condition. Error bars represent ±1 S.E.M, *N* = 3. 72

Figure 4.1 Familiar objects can be recognised from their silhouette (left) or outline (right). Here, the bird (top) and bunch of grapes (bottom) are easily identified when shown at a non-accidental viewpoint. Adapted from Wagemans et al., 2008. 81

Figure 4.2 Points A and B above can easily be ordered in depth, since the object is familiar and its 3D shape is known. Likewise, D appears closer to the observer than C within the recognisable apple silhouette. The rightmost figure is unfamiliar, yet it is still possible to infer 3D shape and order the probe points, perhaps due to internalised knowledge of depth regularities within real objects (rightmost figure adapted from Norman & Raines, 2002). 82

Figure 4.3 Demonstration of how slant, and thus depth, increases towards the rim of an object. 83

Figure 4.4 Left: Participants marked salient positions along the bounding contour; the lengths of radial bars indicate the cumulative frequency of points chosen for that location. Right: Attneave's cat, created by drawing lines between the dots that mark curvature maxima and minima (adapted from Attneave, 1954). 84

Figure 4.5 The two images above can be cross-fused to gain a sense of 3D depth. Note how the objects are positioned slightly differently in the left versus right eye's image. Adapted from Held, Cooper & Banks, 2012 86

Figure 4.6 The magnitude of a smooth contour curvature is calculated by fitting a circle with radius (r) that closely fits the curve (C). In the current study segments of perfect circles were used as contours. Figure adapted from Norman et al., 2006. 88

Figure 4.7 Demonstration of curvature magnitudes (black) used in the study. Left: maximum concavity, centre: flat contour, right: maximum convexity. Note that stimuli have been rotated for viewing purposes and would normally be presented with the contour to the left or right. 88

Figure 4.8 The top (concave contour) and bottom (convex contour) are examples of a single trial from the experiment; left and right images can be cross-fused for a 3D percept. Participants were asked to move the red dot (shown) until it appeared to rest on the surface of the potato. 89

Figure 4.9. Plot showing set depth of probe points, for each of 3 probe distance conditions and seven curvature magnitudes. Data was averaged across participants (*N*=7); bars represent ±1 SEM. 91

Figure 4.10. Plot showing set depth of probe points, for each of 3 probe distance conditions and 3 curvature signs. Data was averaged across participants (*N*=7), bars represent ± 1 SEM 92

Figure 5.1 In this demonstration of slant, the centre image has a slant of zero and is thus frontoparallel. The images either side have been rotated about a vertical axis by various angles, so that the surface is receding in depth. 98

Figure 5.2 In this demonstration of slant and tilt, slant increases across the radial axis, while the angular location represents tilt (the axis of rotation). Note that tilt cannot be defined when there is zero slant. Adapted from Norman, Todd, Norman, Clayton & McBride (2006). 99

Figure 5.3 Demonstration of slant underestimation as predicted by a frontoparallel bias. The Pan-Tilt unit is a device that can mechanically change the slant and tilt of a planar stimulus 99

Figure 5.4 The diagram shows a cyclopean eye viewing the same surface at different degrees of rotation about a vertical axis. This demonstrates the relationship between surface slant and its projected area on the retina. When the surface is not slanted, it projects to the largest space in the retinal image. As the slant increases the projected area becomes smaller. The surface is not visible at 90˚. 100

Figure 5.5 The tessellated sphere represents an approximation of all possible slant and tilt combinations in the world. Only the centre patch is frontoparallel. Slant prevalence increases as patches become farther from the centre 101

Figure 5.6 Projected surface area has a maximum value when slant is zero. Slant prevalence increases as values move away from frontoparallel. Here, the prior for slant peaks at ±45˚. 102

Figure 5.7.Figure showing the independent probability distribution of a) slant and b) tilt of planar surfaces in natural scenes (where zero tilt corresponds with slants about a horizontal axis) c) Shows the joint distribution of slant and tilt; contours are coloured according to the colour bar, with red representing high probability and blue representing low probability. Note the clear asymmetries in slant and tilt in the environment. Adapted from Yang & Purves (2003)..... 104

Figure 5.8 This image shows a participant setting a palm board to match the slant of the stimulus surface (right; wooden block). An inclinometer is attached to the palm board (positioned roughly at waist height, and rotated about the horizontal axis) to digitally record set slant. A head mount was worn to hide the hand from view. Adapted from Taylor-Corvill & Eves, 2013. 107

Figure 5.9 Photograph of haptic paddle in use. An unseen axis at the back of the paddle allows it to rotate left to right; the metal arm (right) rotates to allow paddle movements towards the ground or ceiling plane..... 110

Figure 5.10 Layout of experiment room for Part 2 (limited cue condition). Participants could not see the haptic paddle as it was occluded by the mirror. In Part 1 the occluder was removed and a longer mirror was used for binocular viewing. Not to scale; see Appendix for dimensions. 110

Figure 5.11 This is a Gauss map, showing slant and tilt with respect to the observer. The slant is the same at *cd* and *ab*, while tilt is equal for *ac* and *bd*, yet tilt is more difficult to discriminate at *cd* because the arc length is shorter. Adapted from Norman, Todd, Norman, Clayton & McBride (2006)..... 113

Figure 5.12 Outline and textural cues to surface attitude. a) The outline of the surface will become compressed when slanted; its aspect ratio also provides a cue to tilt, b) size gradient and c) density gradient can be a cue to surface attitude if elements have a uniform size, and d) compression gradient (or aspect ratio) can be informative if

texture elements have a regular shape. Adapted from Saunders and Backus (2006). 114

Figure 5.13 Textures used in the current study. Texture 1, left: reliable cues to slant (high contrast, small circles of the same radius). Texture 2, right: less reliable cues to slant (randomly oriented ellipses of varying size). Note: texture contrast decreased substantially when printed on Perspex. 115

Figure 5.14 The two images represent example stimulus attitudes in the full cue condition (Part 1; Texture 1). Here, images show a slant of 60 degrees from frontoparallel (towards left wall and right wall slant respectively) about a vertical axis. Note that the compression of the stimulus outline (change in aspect ratio) and relative density cues to slant sign. The sense of slant is compelling in this demo, despite it being shown in 2D. The textures are not shown to scale. 116

Figure 5.15. Here, images show a slant of 60 degrees from frontoparallel (towards left wall and right wall slant respectively) about a vertical axis, thus demonstrating how the stimuli might look in Part 2 (limited cue condition, Texture 2). The aperture occludes the edges of the stimulus and disparity cues are lost in monocular viewing. 116

Figure 5.16. Top: The polar plot illustrates the 36 test attitudes used in the study. Tilt is defined along the circumference, slant across the radius. Bottom: Demonstration of the coordinate system for slant and tilt. Blue cylinders are added for visualisation of surface normal. 119

Figure 5.17 Left image shows a 70° right wall slant as it might appear through the circular aperture. Note that although it is clearly very slanted relative to frontoparallel, the sign of tilt is ambiguous. The right image shows the same texture, yet has the added cue of converging lines that aid interpretation of tilt. 121

Figure 5.18. Coordinate system for UV space, as used in analysis of perceived surface attitude 122

Figure 5.19 Figure showing data from Part 1 (full-cue condition) for participant CM. Set attitude is shown as a function of stimulus U or V (left), and as a function of both

stimulus U and V (right). The 2 IV cubic polynomial regression fits are shown in red on left hand plots; the fits are represented by the 3D surface in right hand plots (changes in surface colour from light yellow to dark red represent changes in perceived attitude from positive to negative set U or set V). 123

Figure 5.20. Part 1: Full Cue Condition. Percentage of trials in which tilt was reversed. Each data point represents the average for each stimulus attitude, across 10 repetitions and 9 participants. Data are plotted as a function of tilt (top row) and slant (bottom row). Error bars represent ±1 S.E.M. 125

Figure 5.21. Part 2, Texture 1: Perceived surface attitudes and 2IV cubic polynomial fit for Participant CM. Left: perceived U as a function of stimulus U and perceived V as a function of stimulus V. Right: perceived U and V as a function of U and V with fits represented by the 3D surface plots (changes in surface colour from light yellow to dark red represent changes in perceived attitude from positive to negative perceived U or V)..... 126

Figure 5.22. Part 2, Texture 2: Perceived surface attitudes and 2IV cubic polynomial fit for Participant CM. Left: Left: perceived U as a function of stimulus U and perceived V as a function of stimulus V. Right: perceived attitude as a function of U and V with fits represented by the 3D surface plots (changes in surface colour from light to dark represent changes in perceived attitude from positive to negative perceived U or V). 127

Figure 5.23. Part 2, Texture 1: perceived surface attitudes and 2IV cubic polynomial fit, plotted as an average across all participants (N = 9) and all reps (n=10). Left: perceived U as a function of stimulus U and perceived V as a function of stimulus V. Error bars represent ±1 SEM. Right: perceived attitude as a function of U and V with fits represented by the 3D surface plots (changes in surface colour from light yellow to dark red represent changes in perceived attitude from positive to negative perceived U or V). 128

Figure 5.24. Part 2, Texture 2: perceived surface attitudes and 2IV cubic polynomial fit, plotted as an average across all participants (N = 9) and all reps (n=10). Left: perceived U as a function of stimulus U and perceived V as a function of stimulus V.

Error bars represent ±1 SEM. Right: perceived attitude as a function of U and V with fits represented by the 3D surface plots (changes in surface colour from light yellow to dark red represent changes in perceived attitude from positive to negative perceived U or V). 129

Figure 5.25. Part 2, Texture 1 (left), Texture 2 (right): bias in perceived attitude as a function of U and V with fits represented by the 3D surface plots, plotted as an average across all participants (*N* = 9) and all reps (*n*=10). Error bars represent ±1 SEM. Changes in surface colour green to blue represent changes in bias from positive to negative. 131

Figure 5.26. Average bias in perceived slant for Part 2 (limited cue condition; *N* = 9). Left: Texture 1, right: Texture 2. Quadratic fit to averaged data shown in red. Error bars represent ±1 SEM. 136

Figure 5.27. Percentage of tilt reversals for Part 2: Limited Cue Condition; Texture 1 (left) and Texture 2 (right). Each data point represents the average for each stimulus attitude, across 10 repetitions and 9 participants. Data are plotted as a function of tilt (top row) and slant (bottom row). Error bars represent ±1 S.E.M. 137

Figure 0.1**.** Layout of experiment room with detailed measurements 165

Figure 0.2 Diagram showing coordinate system for haptic paddle. 166

Figure 0.3. Rotary encoders attached to the haptic paddle code rotations about x (horizontal axis) and y (vertical axis). The PT Unit rotates about y (slant) and z (tilt). 167

Figure 0.4. PTU mounted to wall (shown rotated from home by 90). Mounting bracket allows entire unit to be shifted left/right or up/down if desired. Stimulus surface (shown from behind on the left) is 40cm from the wall at frontoparallel. Cable boxes are stuck to the wall to avoid collision with the surface as it rotates. 170

Figure 0.5**.** Example of haptic paddle in use. Left: steep right wall slant; centre: slight right wall slant, close to frontoparallel, right: steep ground-plane/left wall slant. 171

Figure 0.6. View of haptic paddle and armrest. Paddle is positioned in line with the mirror, central to the eyes. 171

Figure 0.7. Full stimulus (Texture 1) in darkened room. 172

Figure 0.8. A:C show mirror set up for Part 1; D shows mirror set up for Part 2. A) Front diagram of mirror and chin rest, B) top diagram of mirror and chin rest, C) photograph of binocular mirror set up, D) photograph of monocular mirror set up, showing 45° mirror and button operated shutter (left). Mirror 1 length 20cm, mirror 2 length 10cm (designed to accommodate binocular and monocular viewing respectively). 173

Declaration of Authorship

I, *Jennifer Anne Elizabeth Josephs* declare that the thesis entitled *Vision as Informed by the Statistics of the Natural Environment* and the work presented in it are my own and has been generated by me as the result of my own original research. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- Either none of this work has been published before submission, or parts of this work have been published as:

Signed: ……...

Date:……..

Acknowledgements

I would like to thank my supervisors for their support and guidance and I am really grateful to have had the opportunity to do PhD research in such an interesting area. I am also thankful to the department for funding all of my equipment needs, without which I would have a much less interesting thesis.

I really appreciate the support of my advisor, Dr Matt Garner, for being available to talk to and Dr Ed Redhead for student support. James Elder was incredibly helpful for talking through coordinate systems and mathematical concepts. I also thank Pete Dargie, Richard Bainbridge, Paul Reynolds, Matt Jones, Darius Beben for their help and patience with engineering the haptic paddle and Pan Tilt Unit. I am really grateful to my participants and friends who took part in many of my (often very long) studies, and to my friends who have stood by me, particularly at the end! I am massively indebted to Dr Natalie Mestry for help with formatting, and to Dr Katie Gray for bringing me food when I was working late. Also, thanks to Nick for making me laugh and putting up with my strange behaviour.

The research in this thesis was funded by the University of Southampton.

Chapter 1

Vision as informed by the statistics of the natural environment

1.1 Vision as problem solving

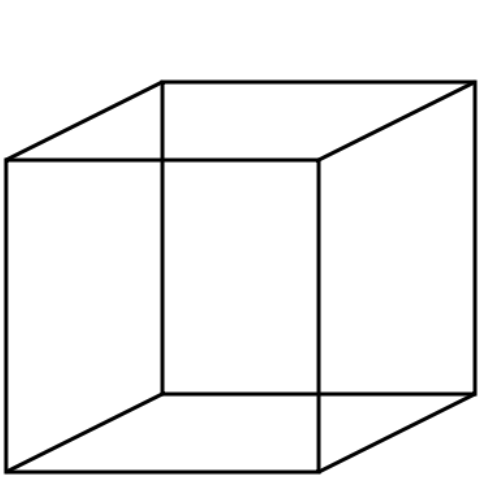
The survival of a species depends on adaptations that allow it to provide itself with food and avoid danger, and this means we must accurately judge properties of the natural environment. However, despite the apparent ease of seeing the world in three dimensions, in reality visual perception is a complex task, involving bottom-up processing and the many inferences that the brain must make about two-dimensional retinal images. For instance, since the retinal image contains no metric depth information the visual system must use rules to segment the image into objects, surfaces and their relative depths (Martin, Fowlkes, Tal & Malik, 2001).

A simple demonstration of perceptual problem solving is the Necker Cube (Figure 1.1), where the perceived shape alternates between one of two interchangeable perspectives. The brain rejects an infinite number of alternative solutions but then switches between two that are equally probable. It is not always obvious how such outcomes are reached so researchers strive to understand how and why those interpretations are made. One explanation is that the brain can utilize prior knowledge and experience to aid image interpretation, as demonstrated in some visual illusions. For example, when observers view the Ponzo illusion in Figure 1.1 they often perceive the two equal lines to be different in length. One possible explanation is that observers have internalised the knowledge that as lines converge they become more distant. This explanation would be consistent with an ecological approach to understanding visual perception and this will be discussed in detail further on, though first I will introduce key questions in visual perception and outline several theories that have sought to address these.

1.1.1. Theoretical approaches to understanding visual perception

Even since the time of Aristotle, scientists and philosophers have asked how it is that the visual system can make sense of the visual world. Often posed questions include 'is perception innate or learned?', 'is perception direct, or must the observer make inferences about the possible stimulus?' and 'do we see things the way they are

because of the way the brain and its systems are designed, or because of our interactions with the environment we live in?' (Palmer, 2002).



Picture removed for copyright reasons

Figure 1.1 Left: The Necker Cube (see text). Right: The Ponzo illusion. The two red lines are the same size on the retina, yet the converging lines made by the tiles in the scene are a cue to depth that makes the right line look longer.

The experiments contained within this thesis are designed in adherence to an empiricist view– that humans perceive the world in a certain way because we have acquired knowledge about the environment. The alternative, nativist view is that perception is hardwired at birth. As with many theories, these views do not have to be mutually exclusive, as it is also possible that humans (and other organisms) have evolved certain hardwired adaptations due to interaction with the environment. This concept will be further discussed in section 1.3. Similarly, the studies within this thesis were created with the view that the environment plays a strong role in perception, or that vision is shaped by the natural environment. However, organic systems are necessarily limited in their capabilities, thus one could argue that the structure of the visual system is important in how we perceive the world. Below I explain several other theories and how their views complement or contradict those adopted throughout this thesis:

- Holism vs. atomism: Gestaltism is the view that the whole is more than the sum of its parts. Advocates of this movement argue that one cannot analyse minute details of sensory information alone and that humans perceive things in a certain way because we interpret information as a whole as opposed to separate parts.

For example, given a sensory input of 5 dots (. . . . .) , a structuralist (focussing on primitive sensory data) would argue that an observer would simply perceive 5 dots. In contrast, Gestalt rules of perceptual grouping explain how one would instead perceive the dots arranged as a straight-line contour. Gestaltism is revisited in Chapter 2, in which I argue that context and structure are important for perception of natural scenes.

- Inference vs. direct perception: Helmholtz's (1867) principle of unconscious inference assumes that following the input of sensory information there must be some processing stage at which the input must be interpreted. Marr (1982) uses the inverse optics problem to support the argument for unconscious inference; any one retinal image can have been generated by an infinite possibility of stimuli in the real world, and thus this ambiguity must at some point be resolved. The Necker Cube is a good example that provides evidence against direct perception, as sensory information remains the same, yet perception switches intermittently.
- A similar concept to unconscious inference is that of internalised heuristics. This view assumes that perception makes inferences based on learned information about the natural environment, and can employ these rules when sensory information is ambiguous. A good example is the Ponzo illusion, where the visual system appears to use the rule that things appear smaller on the retina if they are more distant. Interestingly, some support for the idea of heuristics in perception was provided by Gregory and Wallace (1963), when they discovered that a blind adult who then had vision restored could did not experience the same heuristics as adults who had lived with normal vision (e.g. the Necker cube appeared to be flat, and the participant had poor distance scaling). This also provides support for a more empiricist rather than nativist view of visual perception.
- Ecological optics: Gibson (1950) advocated an ecological approach, arguing that in order to understand vision one must first understand the distribution of properties within the natural environment. Supporters of this view would also recognise the importance of paying attention to the tasks that the organism must carry out within its environment, as these surely impact on visual perception. Interestingly, Gibson rejected the unconscious inference approach in favour of direct perception, believing that the wealth of sensory visual information is so great (particularly when the organism is in motion) that there is no need to make

inferences. Though the ecological approach is appealing and is perhaps the closest theory upon which the experiments within this thesis are based, my view is that perception also makes use of heuristics or prior knowledge, as is discussed below with regard to Baye's Theorem.

- The Bayesian approach: Baye suggested that perception is the outcome of both sensory input (the likelihood) and internalised knowledge about the distribution of properties within the natural environment (the prior). When sensory information is ambiguous, the observer can make a weighted combination of prior and sensory information depending on how reliable the incoming information is determined to be (Geisler & Kerstsen, 2002; see section 1.7 for more thorough explanation). Indeed, the literature covered within this thesis has a strong focus on investigating how prior knowledge affects perception.
- Computational theories: Marr (1982) takes a slightly different approach to the other theories discussed, arguing that computers allow us to model the problems faced by the visual system, and investigate potential solutions that might result in what is actually perceived. It is also of note that computer technology allows much more detailed analysis of natural images, from which it is possible to extract geometrical properties (see Chapter 4) and examine the relationships between image statistics and visual perception (Chapters 2 and 3).

1.1.2. The ecological approach to understanding visual perception

Researchers in the last century realised that it is important to study vision in the real environment if they were to understand perceptual phenomena like the Necker cube and Ponzo illusion. Indeed, Geisler and Ringach (2009) argue that to better understand the human visual system we must first consider the tasks that it faces within the environment. Such visual tasks might include judging shape, size, and distance (useful for reaching and grasping objects) or convexity (useful for placing the feet when walking).

Rather than just creating ecologically valid stimuli, Brunswik and Kamiya (1953) proposed that we should study the natural statistics of the environment itself, arguing that there are strong links between perceptual systems and statistical properties of the environment they evolved in. Technological advances and a recent increase in this approach has meant that there are now a number of studies dedicated

to increasing our knowledge about the links between vision and the environment. Indeed, the overarching framework of the natural statistics approach allows us to draw comparisons between several divisions of research that are explained below:

1. Image statistics: Researchers analyse the information contained within photographs of real world scenes as they represent the retinal image (Field, 1989). There is a great deal of regularity in natural images; they contain edges that join to create structure, have smooth patches containing the same or similar luminance and adjacent pixels or sections that may be of the same colour. If the visual system has an understanding of this statistical regularity then it can increase efficiency by exploiting redundancy, having cells that are selectively sensitive to particular image features. Natural image studies also aim to discover what information observers find most useful for different tasks, such as whether contours are used for grouping (Geisler, Perry, Super & Gallogly, 2001) or whether colour is necessary for rapid scene categorisation (Yao & Einhauser, 2008).
2. Scene statistics: The physical properties of real-world scenes can be measured to obtain frequency distributions that describe different features. For instance, laser range scanners have been used to measure depth, surface slant, the size of objects and distance between points within a scene (e.g. Potetz & Lee, 2003; Yang & Purves, 2003a). This information can then be used to explain results from behavioural studies, such as why people misperceive distance or the length of lines (e.g. Howe & Purves, 2002). Scene statistics can also be used to ecologically validate human biases in perception.
3. Perceptual priors: Behavioural research has shown systematic biases in human perception for tasks such as judging object shape (e.g. Langer & Bulthoff, 2001; Adams, 2007), line orientation (Girshick, Landy & Simonceli, 2011) or the speed of moving objects (Weiss, Simoncelli & Adelson, 2002). As perceptual information becomes more ambiguous, then these biases become more evident. Researchers call these biases 'perceptual priors' as perception is shifted in the direction of known statistics of the environment. In addition to behavioural research on priors, we also look to natural scene and image statistics to explain where the biases originate.

Thus, the goals of the natural statistics approach are to improve our understanding of how the visual system works by better understanding both the environment and how it projects to our retinas. There are several methods that can

be used to reach these goals, including behavioural studies, image and scene analysis, and studies of neuronal encoding. The following review will address current progress and future direction in natural statistics research, beginning with an outline of several methods and an explanation of how they can be used in conjunction with each other.

1.2 Outline of methods in natural scenes research

1.2.1. Image statistics and behaviour

It is possible to analyse images, or groups of images by obtaining statistical summaries of information such as contrast, luminance and the size or orientation of features. One example is spatial frequency, which describes the pattern of luminance throughout an image. The left hand image in Figure 1.2 has more luminance changes per degree of visual angle than the right hand image, thus it has a higher spatial frequency. Researchers can manipulate image information in order to see how it affects human behaviour. For instance, we can selectively remove high or low spatial frequencies leaving either coarse or fine details respectively and examine how this affects scene categorisation (e.g. Schyns & Oliva, 1994).

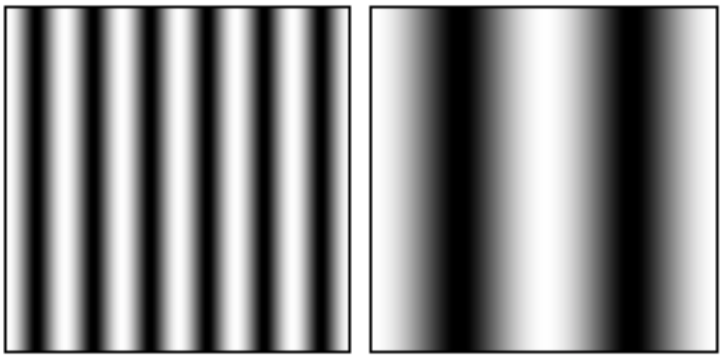


Figure 1.2 Demonstration of high and low spatial frequency in images with sinusoidally changing luminance variations.

1.2.2. Encoding efficiency and image statistics

Another approach looks at how efficiently the visual system can encode information from the natural environment. An efficient system should be sparsely coded so that each group of neurons can encode as much relevant information as possible by reducing or exploiting redundancies in information (Simoncelli & Olshausen, 2001). Techniques such as functional magnetic resonance imaging (fMRI)

can be used to measure the brain's responses to stimuli and compare neuronal activity across different conditions. For example, orientation sensitivity was measured by showing participants sinusoidal gratings at different orientations and increasingly reduced contrast (Furmanski & Engel, 2000). Three methods were used: first, by measuring V1 responses to different oriented stimuli, second, by measuring thresholds for contrast detection, and finally by evaluating observers' discrimination sensitivity. The authors demonstrated an oblique effect in all tasks; sensitivity was greater at cardinal axes (0° and 90°) than at oblique angles (45° and 135°). Importantly, it would seem that the visual system is efficiently coded with respect to orientation in natural scenes as studies show there is greater prevalence of surfaces in natural images at cardinal axes relative to oblique angles (Coppola, Purves, McCoy and Purves, 1998; Girshick, Landy & Simoncelli, 2011).

1.2.3. Scene statistics and perceptual biases

Recent technological advances have made it possible to accurately measure physical properties of natural scenes themselves. Data from laser range scanners, for example, can be used to measure depth and distance (Yang & Purves, 2003b), physical angles (Howe & Purves, 2005) or shape (Yang & Purves, 2003a). Data collected can be used to describe distributions of visual features in the real world. If humans are sensitive to regularities in the environment this may be revealed in behavioural studies that test for perceptual biases.

1.3 Perceptual systems shaped by the environment

The idea that our environment and the tasks that we engage in have shaped our perceptual systems can be evaluated across species as well. Humans possess three colour cones that reach peak sensitivity at short (blue), medium (green) and long (red) wavelengths. Honeybees also have three colour cone pigments but they peak at different wavelengths to humans (Figure 1.3). Their perceivable wavelengths include ultraviolet light that is useful for their survival, but is not necessary for ours and thus is unperceivable to humans (Goldstein, 2007). The overlap in human cone responses for middle and long wavelengths is argued to be a product of evolution that allows discrimination between red and green and thus may aid tasks such as finding red fruit amongst foliage (Parraga et al, 2002). Indeed, Maloney (1986) conducted a

principal components analysis on the reflectance spectra of natural objects and found most of the variance was accounted for by a restricted number of components. Thus, a substantial distribution of wavelengths can be coded with three cone receptors, making the joint responses of our three cones a good approximation of object reflectance spectra. This is evidence that human colour vision is designed in a way that allows high efficiency with respect to the natural environment and the tasks we carry out.

An additional link between the environment and the human visual system is that humans are more sensitive to some spatial frequencies than others, with peaks at around 2 -10 cycles per degree of visual angle (Figure 1.4; Banks & Salapatek, 1978; Campbell & Green, 1965; Campbell & Robson, 1967). One can examine the distribution of spatial frequencies and orientations within natural images using a technique called Fourier analysis. Indeed, after averaging across a large number of natural images, the Fourier amplitude spectrum falls at roughly $1/f$ (where f is spatial frequency, Figure 1.5; Field, 1987; Oliva & Torralba, 2001; Torralba & Oliva, 2003). Indeed, research has shown that images with amplitude spectra of $1/f$ dominate over those with different slopes (Baker & Graf, 2009), thus demonstrating that human perceptual sensitivity is tuned to the statistics of natural images.

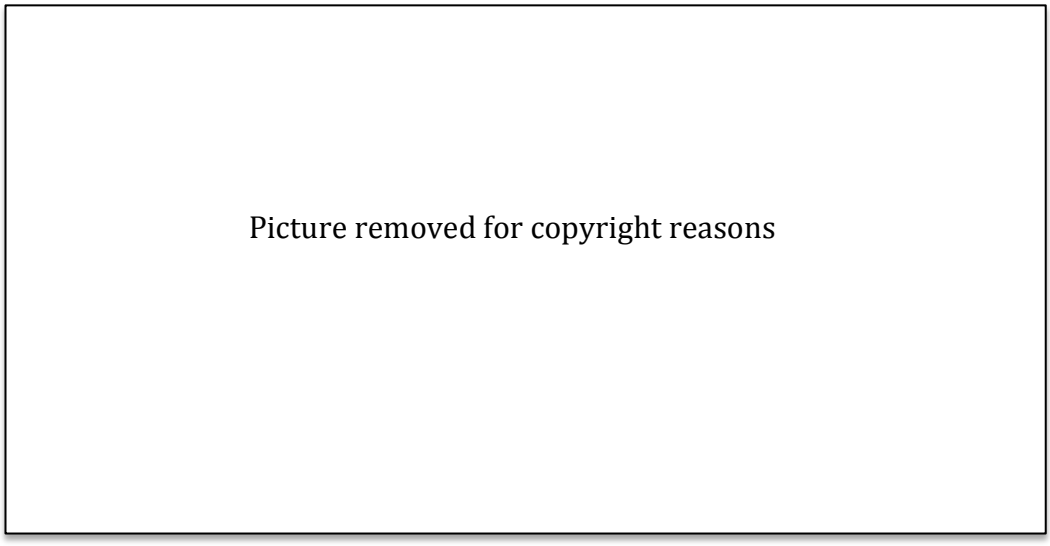


Figure 1.3 Limits of colour perception in humans and honeybees. Figure adapted from Goldstein, 2007.

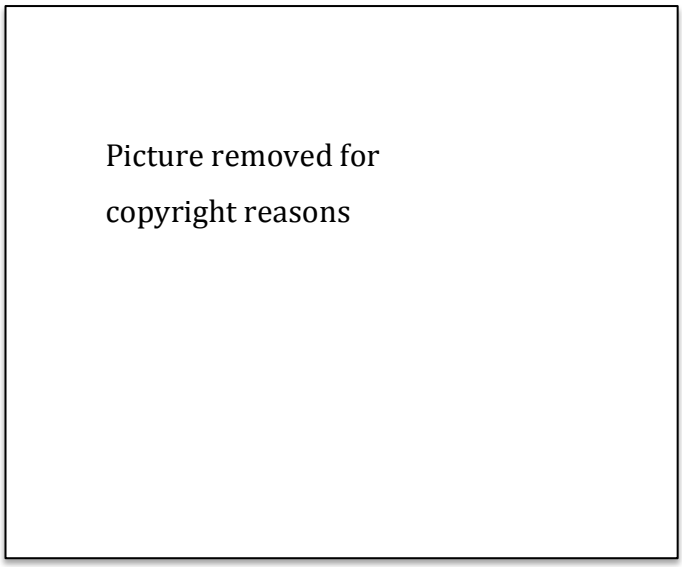


Figure 1.4 Contrast sensitivity function (CSF) for an adult observer. Human ability to perceive high spatial frequencies is finite, ability drops sharply after about 5 cycles per degree of visual angle. Adapted from Banks and Salapatek (1978).

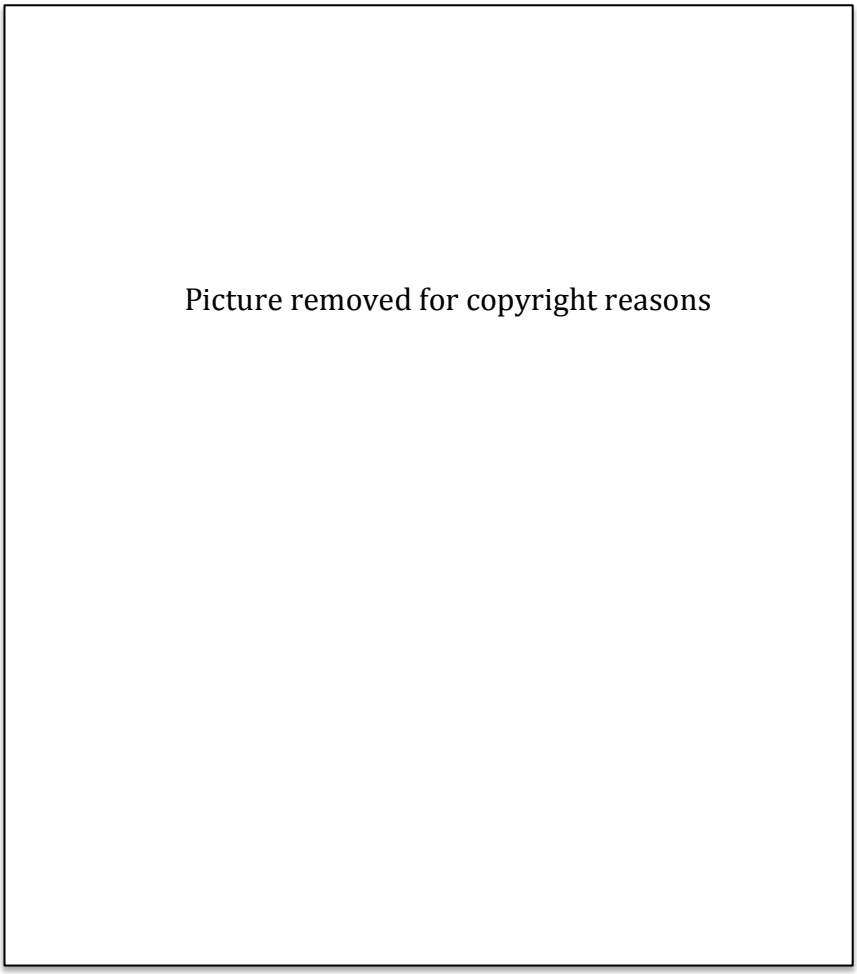


Figure 1.5 The amplitude spectra for 6 natural images (averaged across all orientations). The spectra are plotted on log coordinates to show the 1/f slope found in frequency analysis of natural scenes. Adapted from Field (1987).

Another demonstration of perception mirroring the environment has been shown at a neuronal level. Mante, Frazor, Bonin, Geisler and Carandini (2005) used microelectrodes to measure extracellular signals from neurons in area V1 and the lateral geniculate nucleus (LGN) in anesthetised cats. They showed that gain control for luminance and contrast work independently in the LGN. They also measured contrast and luminance in patches of natural images and found the same independence. This independence can easily be demonstrated in real scenes; for instance, there can be patches of high luminance and low contrast such as the sky, and equally patches of high contrast and low luminance in grass areas. The authors were interested in testing whether i) gain control mechanisms did strongly match the statistics of the environment, ii) luminance gain control could be affected by local contrast, or iii) contrast gain control be affected by local luminance. One of their methods used to test this was by creating images in which luminance and phase were no longer independent, such as by manipulating images to contain their original ($1/f$) amplitude but having randomised phase. If independence of gain mechanisms is not necessary, then just one mechanism should be able to predict LGN responses to stimuli (drifting gratings of differing luminance or contrast). Since luminance was not a good predictor of responses to different contrasts, the authors concluded that a model where luminance and gain are dependent is implausible. Thus, Mante et al. showed evidence of efficient encoding as processes in the visual system match the statistics of the environment.

Low-level evolutionary adaptations are not confined to visual perception and have also been found in the auditory system. A sound can be thought of as a pattern of waves of different frequencies, intensities and wavelengths. The signals can be thought of in similar terms to that of vision, with cochlear nerve fibres tuned to respond maximally to particular frequencies and amplitudes. Lewicki (2002) tested the coding efficiency of cochlear nerve fibres using natural sounds such as human speech, animal vocalization and environmental sounds (crunching leaves, snapping twigs). He used an algorithm to predict efficient codes and found the codes were a good match for physiological filters when a mixture environmental sounds and animal vocalisations was used. Thus the physiology of the ear (i.e. basilar membrane) and subsequent processing is ideally suited to the statistics of natural sounds. This is evidence for efficient coding in the auditory system as it is designed to work

optimally with main classes of natural (but not white noise) sounds. The authors do not explain why a mixture of sounds is needed rather than finding a good match for individual sounds, so there is still some mystery to how the brain encodes information optimally.

1.4 Image statistics

The drive to study image statistics is fuelled by the understanding that natural images contain great regularity; observers will only experience a small proportion of an almost infinite number of possible images. Thus, researchers explore the regularities within natural images in order to determine whether human observers have internalised these statistics. This section will first outline the statistical regularities that have been shown in different image categories and will then examine whether our visual systems are aware of this knowledge and are able to utilize it.

1.4.1 Regularity and Redundancy

Brunswik and Kamiya (1953) were among the first to explore statistical regularities in natural images and how these relate to human perception. They argued that Gestalt principles of perceptual grouping are an artefact of our experience with the way that objects project to the retinal image. In order to test this, they took still images from a movie and analysed the distance between pairs of parallel lines in the image. Though the effect was weak, they did show evidence for the Gestalt rule of proximity; elements that are close together do in fact tend to belong to the same group. More recently, Geisler, Perry, Super and Gallogly (2001) studied the co-occurrence of edges in images to understand whether this predicts contour grouping performance. They used several filters to pick out edges defined by local contrast energy, at any orientation. The likelihood that two edges were part of the same contour depended on three factors: i) distance between the centres of the two edges, ii) the difference in orientation between edge one and its reference, and iii) the clockwise angle between the centre of the second edge relative to the centre of the reference. Their results showed that for two edges to co-occur, they are most likely to be parallel. They also showed that edges are likely to be roughly co-circular (smooth), thus providing ecological validity to the Gestalt rule of good continuity.

Using a different method to study the ecological validity of Gestalt rules, Elder and Goldberg (2002) asked human participants to select contours in images, which were then analysed. Participants were aided by the use of computer software; they selected the start and end of a contour and the program filled in the line connecting the two. Participants were instructed to select both physical contours and those created by shading or shadows. Elder and Goldberg demonstrated that the proximity cue follows a power law, in that as the distance between tangents is greater, the likelihood of them being from the same contour decreases. They also provided evidence that good continuity is described by both parallelism and co-circularity and finally that similarity of luminance could be a cue to good continuity. Importantly, what these studies show is that there are rules or regularities within natural images that may be used by observers (and *are* used in the case of Gestalt rules) to gain information about properties of the scene.

A number of studies have investigated the structure of higher-order image statistics. Second-order statistics describe the relationship between pixels, for instance contrast, or spatial frequency, while the phase spectrum (a higher-order statistic) describes localised spatial frequency and provides the structure (contours) within an image (Gallant, 2003; Field, 1989). A useful property of these statistics is that they can be manipulated while keeping other image properties constant. This allows researchers to examine how important the information is by changing the reliability of each cue and seeing if this affects the observer's ability to extract information from the image (e.g. Joubert et al., 2009; Larson & Loschky, 2009).

One method of summarising images in terms of their second-order statistics is by analysing spatial frequency. Fourier analysis is used to summarise the contrast energy of component spatial frequencies at different orientations (the amplitude spectrum). Field (1987) conducted Fourier analysis on six natural images, and all showed a pattern of decreasing contrast energy as spatial frequency increased (Figure 1.5), and this mirrors what is known about human sensitivity to spatial frequencies, as discussed above (e.g. Banks, 1982). Torralba & Oliva (2003) also analysed second-order statistics in natural images and highlighted qualitative differences between different scene categories. For instance, city scenes are dominated by horizontal and vertical orientations while mountain scenes are more

isotropic. Behavioural studies aim to see whether humans are aware of these regularities and can actually use them to aid image recognition.

1.4.2. Internalised statistics

A common task in studies examining image statistics is to present participants with a stimulus for just tens of milliseconds and determine whether the information contained is sufficient for rapid categorisation (e.g. Fei-Fei, Iyer, Koch & Perona, 2007). Images are typically split into two superordinate categories: those that contain man-made objects and structures and others that do not. Alternatively, they can be grouped by their basic category where the focus is on common categorical representations (e.g. street or city centre). At a subordinate level the focus is on objects and local features within an image (Olympic pool versus paddling pool). Some researchers argue that amplitude information is diagnostic of image category and can be processed by observers very rapidly (Guyader et. al 2004; Kaping, Tzvetanov & Treue, 2007). Others provide evidence that amplitude alone is insufficient for image recognition and some phase information is necessary (Joubert, Rousselet, Fabre-Thorpe & Fize, 2009). The phase of an image is described by localised amplitude, and thus provides its structure. When phase is intact, luminance changes are aligned, thus providing sharp edges and definition. If phase is scrambled then the image becomes unrecognisable. There is an on-going debate as to whether second- or higher-order statistics are more important for scene categorisation; therefore this was addressed in Chapters 2 and 3.

1.5 Efficient coding

According to Information Theory (Simoncelli & Olshausen, 2001) an efficient system will be sparsely coded in a way that is sensitive to common properties of natural images. Several examples have already been discussed including colour (Maloney, 1986) and sensitivity to spatial frequency (Banks, 1982). It is also important to note that neurons themselves are limited in their processing capacities; capacity should be limited to levels that are necessary for the tasks we carry out. For efficient coding of complex natural scenes a sparse code would minimise redundancy and limit neuronal activation. Vinje & Gallant (2000) addressed the issue of sparse coding in V1 by attaching extracellular electrodes to awake macaques and recording

action potentials as they viewed photographs. They showed that neurons in classical and non-classical receptive fields interact, making sparse codes with neurons carrying independent information. Another example of efficient coding can be seen when comparing oriented lines in the real world to orientation sensitivity in V1. Coppola, Purves, McCoy and Purves (1998) analysed photographs of indoor, outdoor and natural scenes and found the distribution of orientations peaked at cardinal axes (particularly in man-made scenes; also see Girshick et al., 2011). This mirrors human sensitivity to orientation since research has shown greater neuronal activity at vertical and horizontal than at oblique orientations (Furmanski & Engel, 2000). In addition, studies show these orientations can be detected at lower contrast than others as well as being more easily discriminated (though see Essock, DeFord, Hansen & Sinai, 2003). It appears that the visual system is efficiently coded to deal with redundancy; more resources are allocated to respond to properties that are more prevalent in the real world. As more research is conducted, it seems likely that further examples will be found to highlight the efficiency of the visual system in mirroring the statistics of the environment.

1.6 Scene statistics

Laser scanners send out a narrow-beam pulse and calculate the time taken for the beam to be reflected off an object and returned to the scanner. They tend to be mounted on a tripod at roughly eye-level to obtain a view similar to what a human observer might experience. Some scanners have accuracies of ± 25mm at a distance of up to 300m (Yang & Purves, 2003b). They are capable of measuring depth at different angular distances above or below the horizon (elevation) as the scanner rotates to different angles about the horizon (azimuth; Figure 1.6). There are several uses for data obtained by using such technology. Range statistics can show relative distance and relative depth across different points within an image ('scene roughness'). Scanners could also be used to measure convexity (see Burge, Fowlkes & Banks, 2010) or the size of objects. Several notable scene studies have been conducted so far.

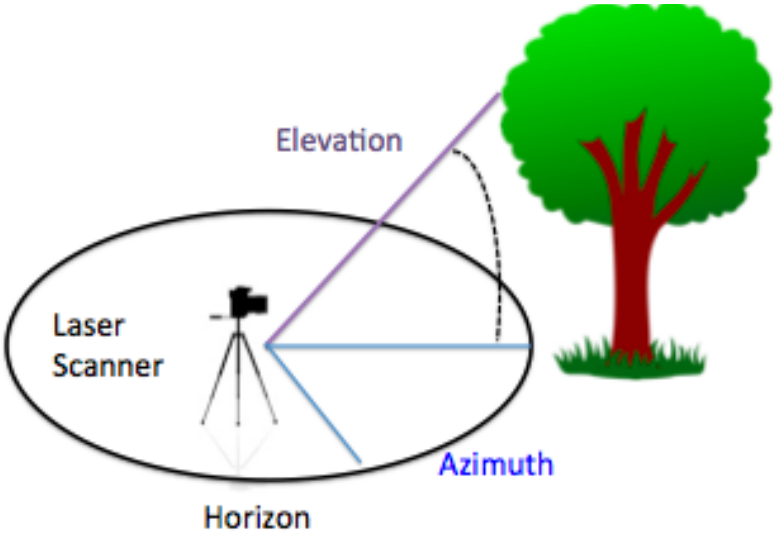


Figure 1.6 A demonstration of elevation and azimuth. The laser range scanner can rotate about the azimuth and elevation to obtain a 3D map of distances from objects in the scene.

Yang and Purves (2003b) scanned both natural and man-made scenes and calculated the probability distribution of distances for points that were not occluded. This was done in order to examine the relationship between perceived and actual distance with the aim of understanding why distance is often misperceived. One perceptual phenomenon investigated was the 'specific distance' tendency (Figure 1.7a); observers tend to estimate that distances are within a range of 2-4m[1]. Results obtained from laser scanning confirmed that the distribution of surface distances away from the scanner peaked at 3m, showing that the visual system's perceptual biases mirror properties of the real world. A second phenomenon investigated by Yang and Purves was the 'equidistance tendency' (Figure 1.7b) that describes how observers tend to assume neighbouring points in images are the same distance away. The authors found that the distribution of physical distances between two points in an image peaked at zero; neighbouring image features do tend to be equidistant, thus explaining the human perceptual bias. Finally, the scene statistics showed that distance from the scanner varied as a function of height above or below eye level. This is important, as it demonstrates that feature distributions can have several peaks and that we cannot assume human bias

[1] For instance, in one study participants were asked to judge the distance of a point of light, whilst wearing an eye patch over one eye. Despite the light being shown at distances between 0.3m and 8.8m, participants judged the distance to be between 2 and 4m on average (Gogel & Tietz, 1979).

will be equal in all parts of a scene (indeed, this is of particular interest for Chapter 5, as is discussed later on).

In another study, Yang and Purves (2003a) investigated several aspects of natural scene geometry by using a laser scanner to measure depth and distance. Their results showed surfaces were rougher along the vertical than horizontal axis perhaps reflecting greater density towards the ground plane that recedes as elevation increases. This is interesting as analysis also shows that image statistics (amplitude) are somewhat location-dependent and can change as a function of height within an image (Oliva & Torralba, 2001). Perhaps the visual system can use this relationship to infer physical properties of scenes directly from image properties. Another finding from the analysis of scene statistics was that size and distance of surface patch were interrelated; things tend to be bigger if they are closer (Yang & Purves, 2003a), and this is probably due to larger and more distant points being more likely to be occluded. Thus, when visual information is highly ambiguous it would make sense for observers to assume that large objects are relatively close. Yang and Purves' (2003a) measurements also revealed anisotropies (differences depending on the direction of measurement) in the joint distribution of slant and tilt, in particular patches tend to be slanted about a horizontal or vertical axis. This is an important finding as it may aid the understanding of perceptual biases in slant estimation, which is the topic of interest in the experiment found in Chapter 5.

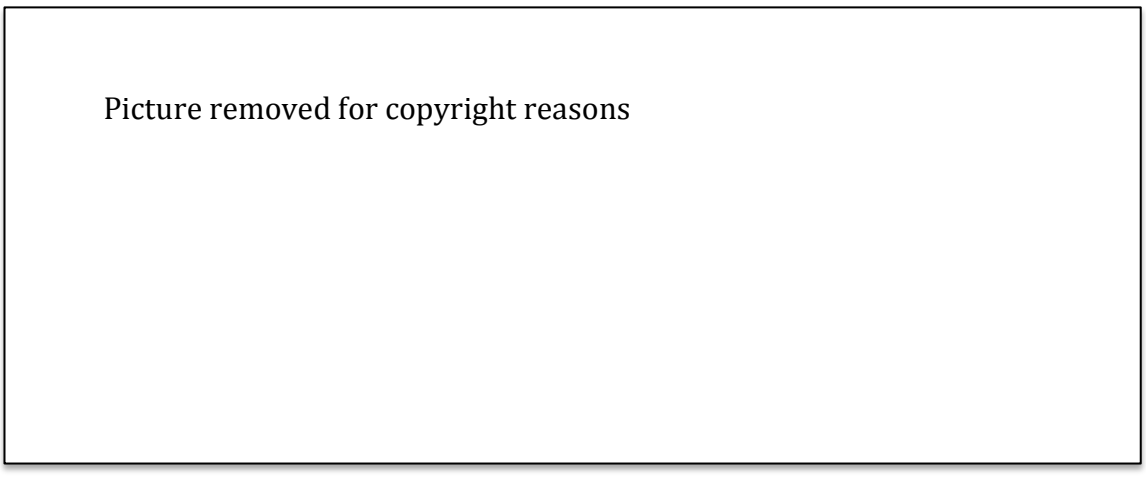


Figure 1.7 (a) The specific distance tendency describes how participants generally assume objects are about 2-4m away, this means they both over and underestimate physical distances. The equidistance theory (b) describes how observers tend to assume two neighbouring patches in an image are at the same distance away from

them, yet in reality their physical distances can vary. Adapted from Yang and Purves (2003b).

Another factor to consider when dealing with scene statistics is human ergonomics; realistically our perceptual experience is limited to the location of the head relative to the environment (the distance we tend to position ourselves from objects, the ground or landmarks). Indeed, Purves argues that it is not the statistics of the real environment itself that biases perception, but the experiences we have with respect to the actual outcomes (Purves et al., 2001). Thus the perceptual system and the environment will not be perfectly aligned if our behaviour affects the frequency at which we experience particular features in the environment. Indeed, one might even expect more efficient coding for more frequently completed tasks and regularly experienced stimuli. Indeed, it is important to investigate the joint probabilities of several scene statistics at once, though such calculations become difficult when many dimensions are involved. It is however undoubtedly easier to study observers' use of internalised scene statistics when one particular cue is isolated. For example, if observers have internalised the knowledge that the probability distribution of possible distances in the environment peaks at 2-4m, then given an ambiguous image where an object appears to be about 5m away, the observer might slightly underestimate the distance of the object. This is a hypothetical example of a systematic bias based on prior knowledge of the environment, or a 'prior'. The following section will outline several known priors and summarise evidence that demonstrates their relationship to natural scene statistics and image statistics.

1.7 Priors

Bayes' Theorem is a statistical framework that can be used to explain human perception, taking into account both sensory information and prior knowledge about the natural environment (see Knill & Richards, 1996; Mamassian, Landy & Maloney, 2001). In a hypothetical scenario where a participant is trying to judge the distance of a point of light, X, then they would acquire new information (the likelihood) about the distance from their visual system, however, the information is made more ambiguous by asking the observer to wear an eye patch over one eye, thus eliminating stereo cues to depth. According to Bayes' Theorem, the participant also has prior knowledge about the probability distribution of distances from completing earlier parts of the

experiments; this prior knowledge is called the prior distribution. In order to calculate the probability of outcomes (the location of the current light point, X), the observer can calculate a weighted combination of sensory information (likelihood) and prior knowledge in order to estimate the posterior distribution (Mamassian, et al., 2001). These concepts are outlined in Figure 1.8.

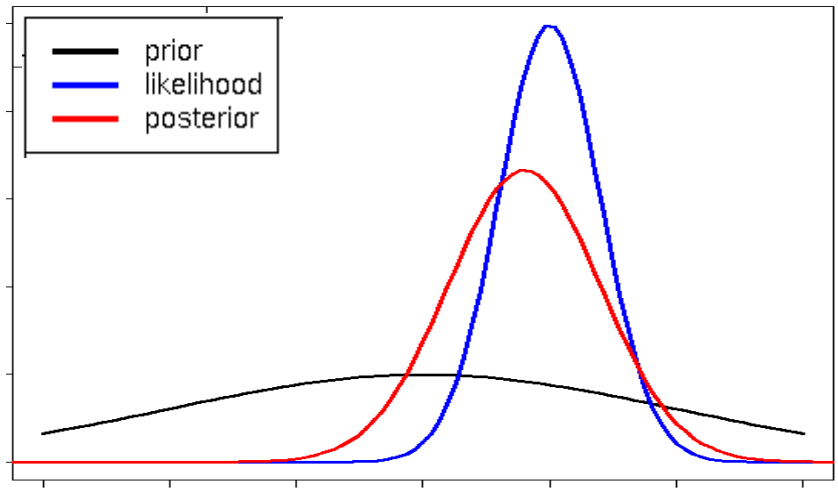


Figure 1.8 Demonstration of perception as Bayesian inference (see text for example).

As outlined above, when faced with a highly ambiguous image, the visual system may utilize prior knowledge about the probable physical source of that image, thus systematically biasing perception. For instance, shape can be inferred from shading in Figure 1.9; though shape is ambiguous, typically observers perceive the top row as convex and the bottom row as concave, using the rule that light tends to come from above. Other empirical tests of the 'light from above' prior have shown that the prior assumes a single light source and that discs must be vertically, not horizontally shaded (Ramachandran, 1988; Kleffner & Ramachandran, 1992).

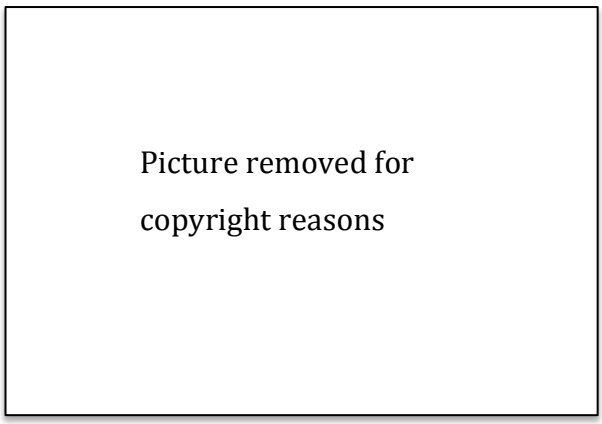


Figure 1.9 Demonstration of 'light from above' prior using shape from shading. Rotating each disc by 180° will reverse the sign of the shape, though the global image properties remain constant.

Another example of a visual prior is the tendency for human observers to be biased towards convex judgements when they are shown ambiguously shaded discs. Langer and Bülthoff (2001) empirically tested for a convexity prior by asking participants to state the shape of local patches within larger surfaces that were globally convex or concave. Since participants' performance was above chance for the convex shape and below chance for the concave shape it was inferred that participants demonstrated a convexity bias. If observers had based their perception on other information in the image then performance should have been equal for both shapes. This bias may be based on experience of internalised knowledge about the relationship between the retinal image and physical environment, and this relationship could be ecologically validated with use of scene statistics.

Indeed, some studies investigating perceptual biases are able to validate their results with image or scene statistics that have already been measured. For example, it is well known in vision research that the distribution of oriented contours in natural images peaks at the cardinal axes (0° and 90°; Coppolla, Purves, McCoy & Purves, 1998). Girshick, Landy and Simoncelli (2011) conducted a psychophysical study to investigate human bias in orientation perception. They asked participants to simultaneously view two arrays of oriented Gabor patches (that could vary in orientation) and report which array had a mean orientation that was rotated more clockwise. Their results showed a bimodal prior distribution that peaked at horizontal and vertical. Essentially, this meant that when the orientation stimulus was more ambiguous, then participants' estimations were shifted towards the closest peak in the prior. This demonstrates that context must be taken into account when

considering perceptual biases, since the direction of a bias might be influenced by its position relative to the prior (e.g. slant might be underestimated in the vertical plane but overestimated in the horizontal plane; Durgin, Li & Hajnal, 2010).

Researchers have also studied a prior for speed that may be utilised when visual information is more uncertain, such as when stimulus contrast is low. Sometimes when observers see a moving stimulus, there is ambiguity in the speed or the path, so they either assume a short path or a slow speed. Researchers have therefore described a prior for slow, smooth speed, with a distribution peak at zero or no movement (Weiss, Simoncelli & Adelson, 2002). This mainly applies to apparent motion stimuli, since there are other cues in real moving stimuli (motion parallax, accommodation), thus showing how prior knowledge becomes more highly weighted when visual information is highly ambiguous. Indeed, Weiss et al.'s results showed participants perceived lower contrast (and thus more ambiguous) stimuli to be moving slower than higher contrast gratings. This demonstrates how a shift in cue reliability can lead to greater reliance on prior knowledge. Furthermore, the speed prior explains real-life behaviour since evidence shows that car drivers tend to speed up in the fog, assuming that they are driving more slowly (Snowden, Stimpson & Ruddle, 1998).

1.7.1. The origin of priors

Researchers strive to learn where such priors originate and what they can tell us about how the visual system works. In fitting with the natural statistics approach, numerous studies have been cited that show a relationship between real-world statistics and behavioural biases. Nevertheless, it is still debatable whether priors are learned or innate, whether they are universal and whether they can change based on experience. The following studies have made some progress in answering these questions.

When describing the 'light from above' prior, one could argue that it exists because there is typically only one light source from overhead - the Sun. However, there is some debate about whether the light prior is directly above or slightly to the left (Mamassian & Goutcher, 2001). Indeed, Sun and Perona (1998) found their observers had a light prior that was above and left, though this was related to handedness. They argue that prior coordinates could be due to the way people

position themselves relative to the light source for tasks such as writing or drawing. Perhaps O'Shea, Banks and Agrawala (2008) would agree with this reasoning, since they found participants to be most accurate at setting a probe to match the local surface orientation of a 3D, shaded surface when the light source was 20-30° above the viewpoint (elevation); if it were directly above, then the observers' head might cause a shadow over the task they were working on.

It seems more likely that differences in the light prior are due to task requirements and perhaps even experience, as there is great variation in the measured light prior across different studies. For example, Adams (2007) reported the results of a number of previous papers, and found the average light prior reported in each to be: μ = -26.1°, σ = 15.4°, Mamassian & Goutcher (2001), μ = -1.3°, Adams, Graf & Ernst, (2004) and μ = -0.8°, σ = 28.3°, to μ = -13.9°, σ = 37.9°, Adams (2007). Adams (2008) demonstrated that the frame of reference for the light prior was retinal for a visual-search task, whereas for shape judgement a combination of gravitational and retinal frame of reference was used (average light prior shifted 60% towards the gravitational reference point). This showed that different information was used for tasks requiring rapid response (time pressure versus accuracy), whereas a more robust approach with additional information was used for explicit shape judgements. This also demonstrates how time constraints of a task affect perceptual outcome by using a 'quick and dirty' approach unless more time is available for more thorough processing (Adams, 2008).

When considering the etiology of a prior, it is interesting to determine how malleable the prior is. That is, are priors learnt, and if so, can they be modified by external feedback regarding the current state of the world? Adams, Graf and Ernst (2004) tested whether learning could alter the light prior by trying to alter their priors using haptic feedback and found that, following visual-haptic training, participants' light priors (in visual-only trials) were shifted towards the direction of the haptic feedback. The researchers concluded from this that priors can be affected by learning and may adapt to match properties of the environment that the organism is currently experiencing.

1.8 Correlations between images and physical scenes

In addition to an observer's knowledge of the physical properties of the world, there are situations in which they may have knowledge of how a physical property (such as line length or orientation) tends to project to the retinal image. If this knowledge is internalised then humans can utilise their understanding of image statistics to infer properties of features in the environment. There are some cases in which the physical properties of natural scenes are correlated with the image that those scenes make on the retina. We can study these instances in order to understand how the visual system works. Indeed, driven by the observation that the apparent length of a line in a retinal image depends on its orientation, Howe and Purves (2002) studied the relationship between the length of a line on the retinal image and its length in 3D space. Using range and image data the authors found that the orientation of a projected retinal image changes systematically with 3D line length. This may explain some line length illusions; the rule the visual system uses (that vertical lines are typically shorter than horizontal ones) might be correct in most situations, though because it uses this bias in ambiguous situations then it can lead to errors. It is also of interest that Howe and Purves find that the average spatial interval (depth) in the real world varies depending upon the orientation of its projection on the retina. This depth is greatest when the projection is close to vertical. Indeed, this matches the bias in line length perception found in behavioural studies (Craven, 1993).

Howe & Purves (2005) later investigated whether the relationship between retinal projections from physical angles can explain why observers tend to overestimate acute angles and underestimate obtuse angles. They used an image database with range and luminance information (Figure 1.10) of i) entirely natural scenes and ii) those that contained man-made structures. Using a straight-line template, they found regions in the scene at which the line was equidistant, and then rotated a second line, which was considered a valid physical source of the angle if they also made a line in 3D space. This was done with reference lines at 0, 45 and 90°. Their main finding was that the probability of a reference line and a second line forming an angle with it decreases as the pair becomes more orthogonal. Howe and Purves (2005) explain this is likely a result of the inverse relationship between patch size and probability, since smaller angles are less likely to be occluded.

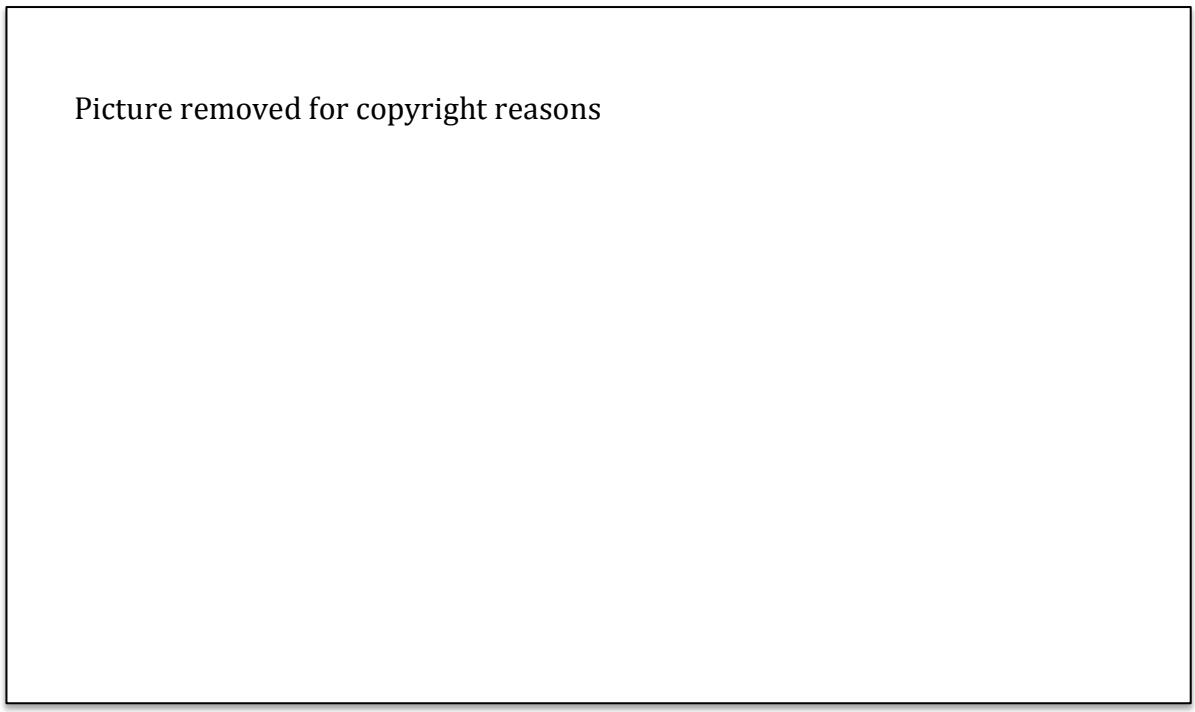


Figure 1.10 The bottom left image shows the range information for the corresponding full colour image above. The images on the right show how straight line templates were used to identify different angles within the image. Adapted from Howe & Purves (2002 and 2005).

Potetz and Lee (2003) took range and intensity measures from real scenes and their corresponding images in order to establish the extent to which intensity could predict range. First they found that range recedes into the distance, at a slightly steeper incline in rural than urban images for their chosen image set. Secondly, when analysing covariance between pixels in image and range patches, the authors found greater correlation with neighbouring range than luminance pixels, suggesting that luminance changes are more abrupt than changes in depth. They also showed a negative correlation between luminance and range for rural images; brighter pixels were also closer to the camera. Potetz and Lee argue that this correlation is an artefact of shadows on surfaces in natural images, caused by textures that are not as apparent in man-made scenes (self-shadowing). By analysing intensity and range the authors also showed that convex surfaces were more likely to be brighter at the top and concave surfaces brighter at the bottom; thus matching behavioural evidence that top-lit discs tend to look convex (Kleffner & Ramachandran, 1992; Langer & Bulthoff 2000). Though range cannot be predicted very accurately from intensity in this case (4%), some ambiguity can be reduced when considering bright patches since these

are correlated (r = 0.28) with filter responses[2] to vertical slopes receding away from the observer; bright patches tend to point upwards and away from the observer. They also found the trend for convex objects to be brighter on top was greater for natural than urban images. This is important as it suggests that the visual system may have to employ priors or weight them differently depending on scene context. Studies such as this could explain how the brain computes depth from luminance. Finally, though the relationships in Potetz and Lee's study do not have large predictive power, there are many other factors and relationships that are not considered. It is logistically very difficult to study the joint probabilities of many cues, though it is likely that the visual system itself can do this. Indeed, there is much research that shows how the visual system performs in a statistically optimal manner and must use a combination of cues at different weights, yet it is difficult to accurately model human performance. Behavioural studies are conducted to establish the human limits or abilities with natural scene perception and to try and piece together the visual cues that are most informative.

1.9. Behavioural studies in natural statistics research

As suggested above, one area of research that has been particularly active in recent years concerns the rapid categorisation of natural scenes, and much effort has been made to understand what information within images allows them to do this. There are several methods to study rapid scene categorisation, including: recording the briefest presentation durations at which images can still be classified (Greene & Oliva, 2009), testing whether categorization is still possible with limited attention (Fei-Fei et al., 2002) and determining whether images are still recognisable when amplitude or phase are altered (Joubert, Rousselet, Fabre-Thorpe and Fize, 2009). For instance, Fei-Fei, Iyer, Koch, & Perona (2007) studied the time-course of image processing by asking participants to give free recall descriptions of greyscale images that were presented for varying durations (27ms to 500ms). There was a tendency to report more sensory or low-level descriptions ("rectangular" or "dark") at short presentations but these were replaced by scene context and object descriptions at

[2] In this example, the filter was designed to detect features that point upward and away from the observer; the filter can then be systematically placed over every patch of the image to register whether or not the feature is present within that patch.

longer presentation times. Importantly, *scene categorization* occurred early on, suggesting it did not require explicit awareness of the scene content itself. It seemed that observers could tell whether the scene was indoors or outdoors before they could accurately name objects. Nevertheless, it is difficult to argue that scene and object recognition are independent, since even perceiving the shape of an object, without explicit identification, could prime the observer towards a particular image category. It is interesting to note that many studies of scene gist present their stimuli in greyscale. Research has shown that although colour can aid confidence in memory retrieval and speed responses when it is diagnostic of an object (Yao & Einhauser, 2008; Rousselet, Joubert, & Fabre-Thorpe, 2005) it is not necessary for gist recognition (Fei-Fei et al, 2002). Presenting images in greyscale allows researchers to control for colour cues and focus on the importance of other information.

1.10 Future research

As has been outlined in this review, there are several approaches that can be taken to investigate the ways in which vision is informed by the statistics of the natural environment. Within each approach, whether focusing on behaviour, image- or scene statistics, there are many methods that have been used, and as many perceptual or statistical biases that have been investigated. The empirical chapters that follow focus on examining human biases in visual perception and relate these to what is known about the environment and natural images.

In Chapters 2 and 3 I examine the role of amplitude spectra in rapid scene categorisation. As discussed in part 1.9 (and 1.4), there are statistical regularities in the distribution of contrast energy across orientation and spatial frequency that are characteristic of different scene categories (e.g. Oliva & Torralba, 2001). If the visual system has internalised the properties of characteristic amplitude spectra, then they might be diagnostic of natural or manmade scenes. Since we know that observers can rapidly categorise natural images when information is deformed or absent (e.g. Yao & Einhauser, 2008; Schyns & Oliva, 1994), then it may be diagnostic amplitude information that is retained and used for categorisation.

In Chapter 4 I will explore a hypothesised, but untested bias, that humans use the bounding contour of ambiguous 2D shapes to infer surface depth (Bertamini & Wagemans, 2013). Evidence does show that object identity can be inferred from

object silhouette alone (e.g. Wagemans et al., 2008) and that the proximity to the bounding edge of a contour does affect participants' judgements of surface depth (Norman & Raines, 2002). Indeed, existing evidence from the analysis of natural scene statistics shows that the shape of a bounding contour should be informative, since the sign of contour curvature is correlated with the distance between foreground and background (e.g. Burge, Fowlkes & Banks, 2010). Thus I examine whether the *sign* and *magnitude* of contour curvature affects perceived depth internal to a contour. If contour curvature does significantly affect depth judgements, then it would motivate future studies to more thoroughly analyse existing natural statistics databases, as this might provide ecological validation if a bias does exist.

Finally, I will examine the *known* bias of slant underestimation. Many studies have reported evidence that observers tend to underestimate slanted surfaces (e.g. Andersen, Braunsein & Saidpour, 1998; Warren & Mammassian, 2010) and some researchers argue that this is evidence of a 'frontoparallel prior', in which surfaces are perceived to be flatter than physically indicated, relative to the observer (Adams & Mamassian, 2004; van Ee, Adams & Mamassian, 2003; Caudek, Fantoni & Domini, 2011). Nevertheless, slant studies that present stimuli on computer monitors introduce a cue to flatness from the frontoparallel monitor (Watt, Akeley, Ernst & Banks, 2005), as such previously reported results might not be fully indicative of the biases that exist in natural viewing. Furthermore, there has been little emphasis on investigating slant biases that might be directly linked to the statistics of the natural environment, thus researchers tend to study slant perception about a single axis (e.g. horizontal or vertical), without taking into account anisotropies in the distribution of slanted surfaces in the real world (see Yang & Purves, 2003a). Therefore, in Chapter 5, I will examine potential biases in slant perception across different tilt axes, using a novel method that allows real 3D surface stimuli to be presented at different surface attitudes, using a haptic (touch) device to record observers' responses.

Chapter 2

Suppression, but not Facilitation, of amplitude statistics in natural image categorisation

2.1. Introduction

Humans are very good at categorising briefly presented natural images (e.g. Rousselet, Joubert & Fabre-Thorpe, 2005, Joubert, Rousselet, Fize & Fabre-Thorpe, 2007; Greene & Oliva, 2009). After presentations of only 30ms or so, observers can provide verbal descriptions of perceptual variables such as lightness or colour and may also recognise shapes or contours (Fei-Fei et al., 2007). At longer presentation times, image descriptions progress from coarse to fine detail (Greene & Oliva, 2009; Loschky & Larson, 2010); for instance, participants may identify objects or activities occurring within a scene (Fei-Fei et al., 2007). Nevertheless, scene gist can be obtained even when objects are difficult to identify, such as in filtered images that contain only low spatial frequency components (Schyns & Oliva, 1994) or images that are presented in peripheral vision, thus having low spatial resolution (Larson & Loschky, 2009). Researchers have recently sought to determine which information is diagnostic within images, and thus may underlie these rapid categorisation effects (e.g. Loschky, Hansen, Sethi & Pydimarri, 2010; Loschky & Larson, 2008; Guyader, Chauvin, Peyrin, Hérault & Marendaz, 2004; Gaspar & Rousselet, 2009).

There are characteristic differences in orientation statistics across scene categories that could drive rapid scene categorisation. For example, though both natural and manmade scenes tend to be dominated by vertical and horizontal orientations, natural scenes tend to contain relatively more oblique orientations (Oliva & Torralba, 2001). One can examine an image by deconstructing it down to its component sinusoidal gratings that different in spatial frequency, orientation, amplitude and phase. If one created a luminance profile of an image, the difference between high and low luminance patches can be described as the amplitude and thus is also referred to here as contrast energy. The phase of the component gratings describes the relative position of the sinusoidal gratings, thus when phase is aligned then it can describe contours within the image. For instance, a square wave can be

created by adding several aligned sinusoidal components (Palmer, 2002; Sekuler & Blake, 2002, see Figure 2.1). This decomposition of an image can be achieved by conducting a Fourier analysis, which provides the output: i) the amplitude spectrum[3] (the contrast or amplitude at each spatial frequency and orientation) and ii) the phase spectrum, which specifies the phase of each grating.

Figure 2.1 A square wave (F) can be constructed by adding several different components (A to E) that have the correct phase and amplitude. If one were to alter the phase of some of the sinusoidal gratings, then the contours in F would lose their definition. Note that this example only uses a single orientation, whereas images of natural scenes are much more complex and can be deconstructed into gratings of any orientation. Figure adapted from Sekuler & Blake (2002).

[3] Note: researchers sometimes use the term 'power spectrum' when describing the squared amplitude

By calculating the mean Fourier power spectrum (squared amplitude) of a large group of natural and manmade images, Torralba and Oliva (2003) showed that there are substantial differences in the distribution of contrast energy across orientation in both superordinate (natural/manmade) and basic scene categories (see Figure 2.2 for examples). Perhaps it is these differences that allow us to distinguish categories so rapidly. Indeed, Oliva and Torralba (2001) found a relationship between observers' descriptions of global qualities such as depth or naturalness, and the amplitude spectra of images, however, it is unclear whether this is a causal relationship that extends to rapid scene categorisation.

2.1.1. Methods of studying the use of amplitude in scene categorisation

A number of different psychophysical methods have been used to explore the role of amplitude and phase on scene gist. Most of these involve image manipulation by either adding in white noise (flat amplitude), scrambling phase, or morphing information from more than one image (thus creating a hybrid). For example, Joubert, Rousselet, Fabre-Thorpe and Fize (2009) first equalised the contrast and luminance of natural and manmade scenes, then replaced the Fourier amplitude spectra of individual images with the averaged amplitude of the entire image set (See Figure 2.3). It is clear that although the amplitude spectrum is identical for natural and manmade test images, scenes are still easily distinguishable due to edges and shapes defined by localised phase. Indeed, when participants were presented with the modified images for 26ms in a go/no-go task, their performance was still well above 90%, indicating that intact amplitude was not necessary to complete this task (though response times were significantly faster when amplitude was intact). Conversely, blending image phase information with phase from a white noise image severely disrupts scene categorisation (see Figure 2.4) and performance falls to chance at 50-60% phase noise. This suggests that amplitude alone is uninformative and must be localised for accurate categorisation (Loschky & Larson, 2008; Joubert et al., 2009).

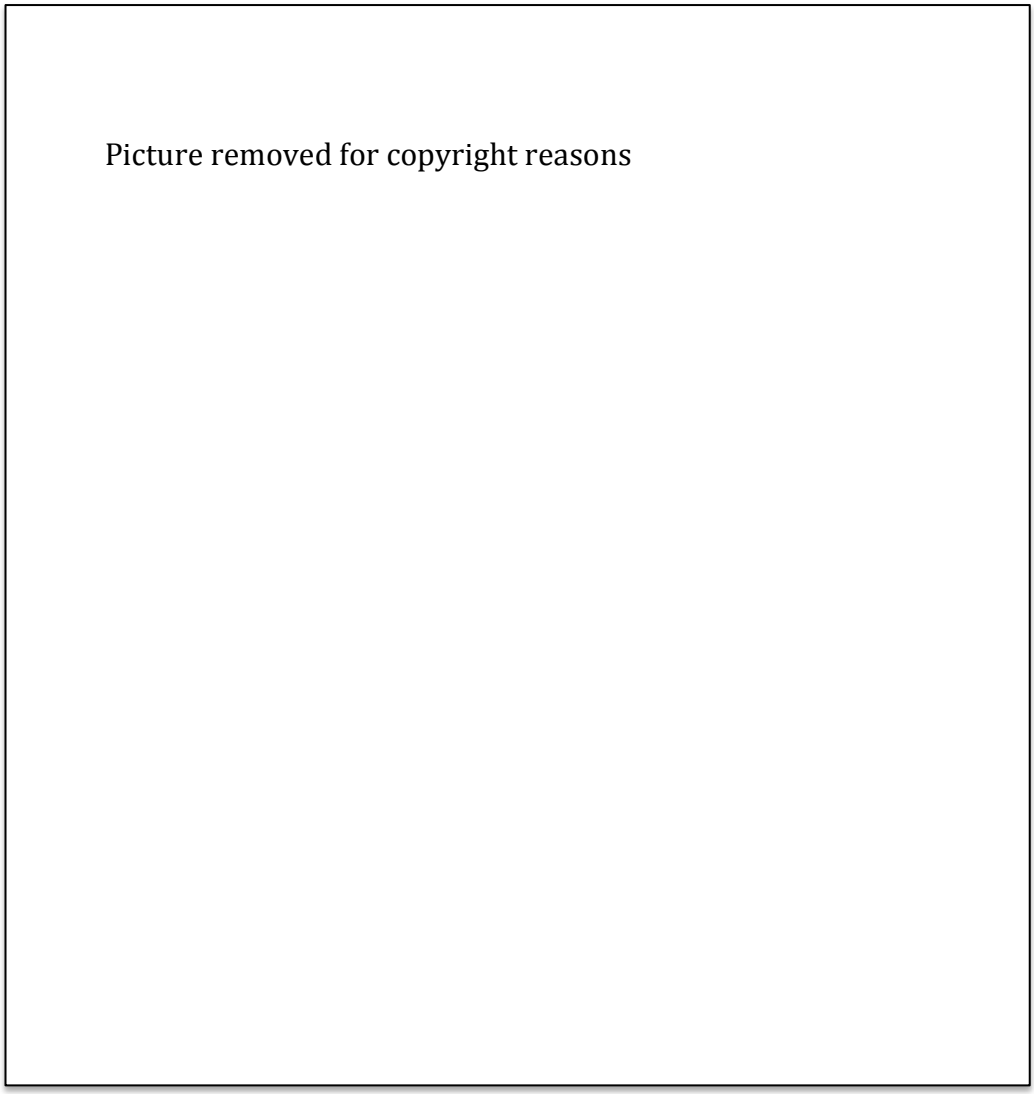


Figure 2.2 Figure adapted from Torralba & Oliva (2003), summarising average power spectra across multiple images from natural/manmade and basic scene categories. A) and C) show averaged power spectra for natural and manmade scenes respectively. Contrast energy is plotted along the z-axis, with larger peaks indicating higher amplitude; spatial frequency increases from the centre of the plot (units in cycles per degree of visual angle). B) and D) are contour plots representing this same information, with orientation plotted about 360° (though plots are rotated 90° so that contrast energy at the horizontal component is displayed vertically). Contours represent 50% (inner contour) and 80% (outer contour) of cumulative contrast energy; contours close to the central point indicate high contrast energy at low spatial frequencies. E) Spectral signatures representing average power spectra for several basic scene categories. Contours represent 60, 80 and 90% of contrast energy.

Priming and adaptation studies have also been used to determine the diagnostic role of amplitude information in rapid scene categorisation. In adaptation paradigms, long exposure to a prior stimulus tends to reduce sensitivity to a similar test image (e.g. Webster & Miyahara, 1997). Kaping et al. (2007) used adaptation to investigate the importance of amplitude information. Participants categorised images after extended exposure (1.17s) to abstract stimuli that mimicked the amplitude spectra of natural or manmade scenes. Observers' categorisation responses were biased away from the category of the adapting stimulus. In contrast, if a test image is preceded by a briefly presented matching prime, the test image is typically identified more quickly or more accurately (e.g. Cave, Blake & McNamara, 1998). Priming has thus been described as a memory process, whereby an initial memory trace remains active and can affect processing of similar succeeding stimuli (see Wiggs & Martin, 1998, for review).

In an image categorisation task, Guyader et al. (2004) primed participants using either the amplitude or phase component of a city or beach image (10ms) before showing an intact beach or city image (20ms) for categorisation (go/no-go). Although performance accuracy was at ceiling across conditions, response times were shorter when a test was preceded by an amplitude prime from the same category rather than a different category. In contrast, response times were unaffected when prime *phase* was manipulated to be either consistent, or inconsistent with the test category (where phase is defined as localised amplitude).

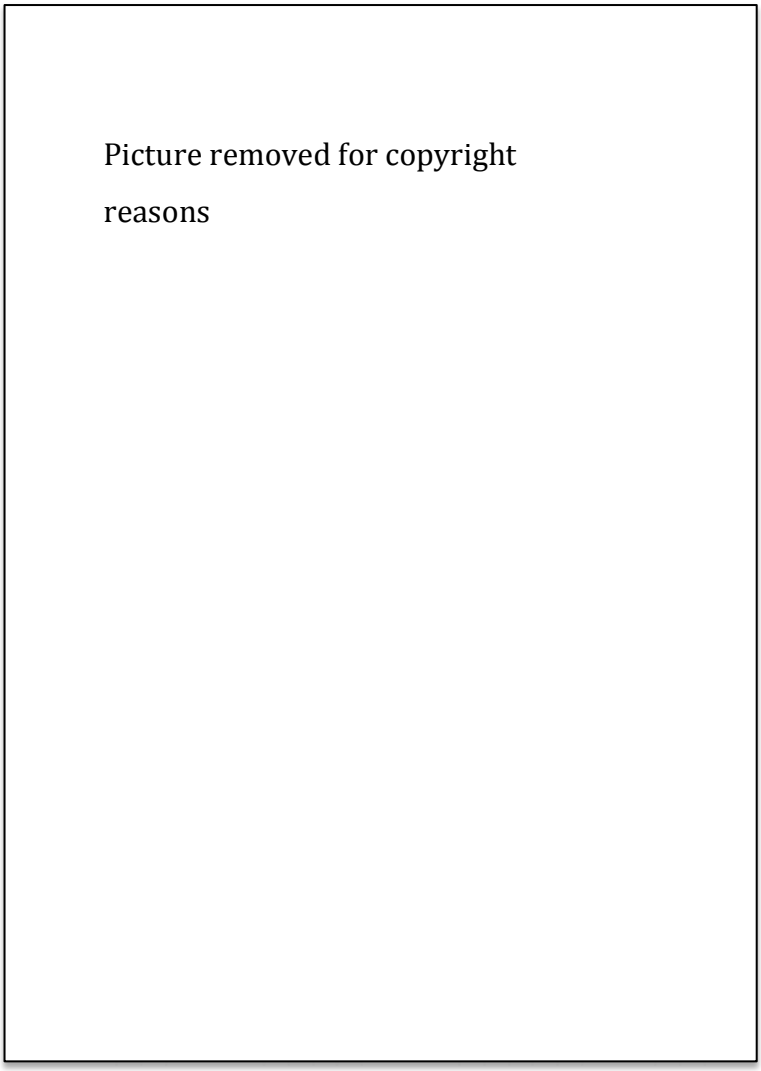


Figure 2.3 Top: original natural and manmade scene. Middle: images adjusted to contain the average luminance and global contrast across the entire 1152 image set. Bottom: Fourier analysis was used to calculate amplitude spectra of each image and the average across the entire set; ELA stimuli (equalized amplitude spectrum) were created by pairing image phase with the average amplitude and then applying an inverse Fourier transform Adapted from Joubert, Rousselet, Fabre-Thorpe & Fize (2009).

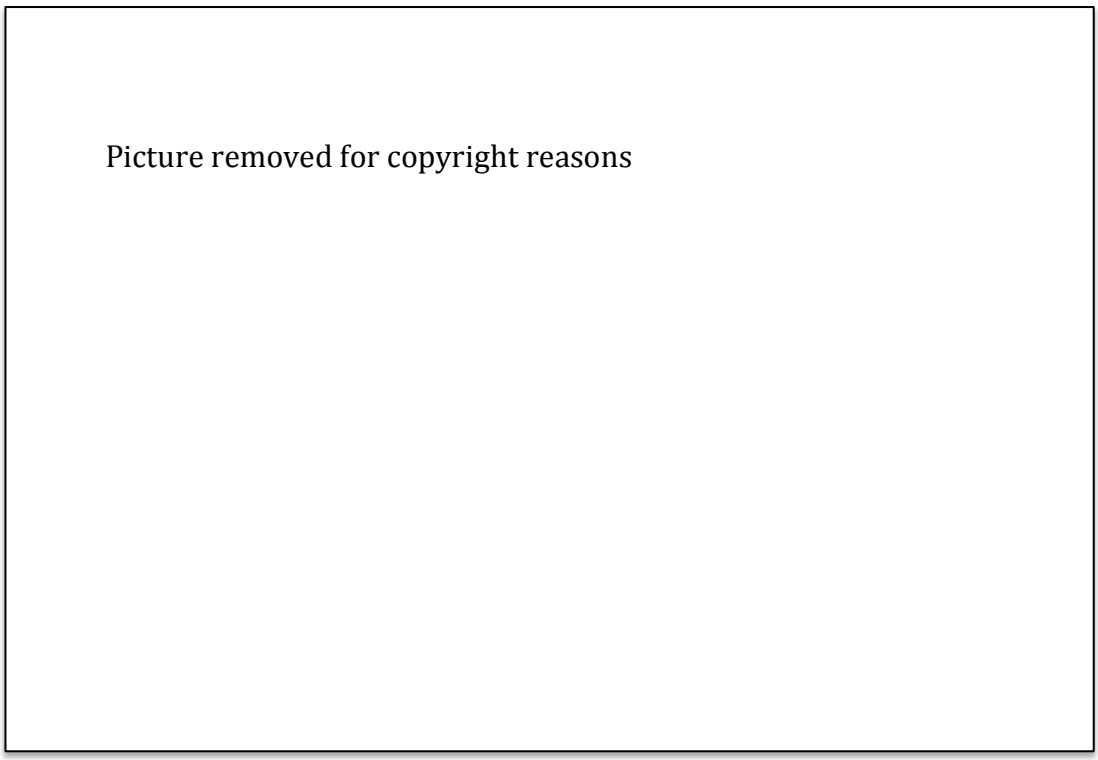


Figure 2.4 A city scene manipulated to contain different levels of phase noise from 0 (original image) to 1 (phase entirely randomised). Adapted from Loschky & Larson (2008).

Together, Kaping et al. (2007) and Guyader et al. (2004) provide evidence that amplitude information does play a role in rapid scene categorisation, such that amplitude information, presented prior to a test image, biases observers' categorisation of 'manmade' versus 'natural' scenes. Though these findings contradict evidence that observers are unable to categorise phase scrambled test images, one could argue that manipulated images are unnatural and their lack of coherent contours might outweigh any categorisation bias due to amplitude statistics**.** Despite this, amplitude information in phase-scrambled images should still be processed and thus priming provides a useful method for studying its potential effects**.** Kristjansson (2006) demonstrated that priming occurs when a memory trace remains following exposure to a behaviourally relevant stimulus. Since the visual system is sensitive to the statistics of natural images (e.g. Baker & Graf, 2009; Tolhurst & Tadmor, 2000) it makes sense that amplitude could be primed and subsequently influence scene categorisation. Indeed, Rajimehr (2004) showed that the direction of sub-threshold (perceptually invisible) orientation patches could bias perceived direction in a subsequent bi-stable motion stimulus. This demonstrates that although participants

were not explicitly aware of the prior-stimulus or its contents, it nevertheless influenced their perception of the test image and we might expect the same effect for amplitude priming.

Given the discrepancies within the literature discussed above, the current study investigates the effects of amplitude statistics on rapid scene categorisation, while also revisiting priming and adaptation paradigms to test whether the Guyader et al. (2004) and Kaping et al. (2007) studies represent two ends of a continuum from facilitation (short primes) to suppression (long primes/adaptors). A facilitative effect would imply that the visual system associates characteristic amplitude spectra with subordinate scene categories (natural/manmade), while a suppressive effect would demonstrate that disruption of amplitude information makes scenes more difficult to recognise, whether due to the suppression of amplitude alone, or the decreased salience of contours that may be affected by amplitude-phase interaction.

The key hypothesis for this study is that the visual system is able to distinguish between different scene categories based on their unlocalised amplitude information. Consequently, when a hybrid containing both natural and manmade information is presented at test, subsequent to a briefly presented natural amplitude prime, the test image should more often be categorised as 'natural' due to conceptual priming. Conceptual priming, in this sense, refers to a bias in responding to the test image that is due to the *semantic* meaning of the prime, rather that repetition of a spatially identical image at prime and test (perceptual priming). Typically, conceptual priming studies use lexical tasks, such as word completion (see Schacter & Buckner, 1998, for review), thus making the current task a rather novel one as it relies upon implicit categorisation of low-level information (amplitude) within the prime stimulus. Nevertheless, Guyader's (2005) results could also be described as conceptual priming, and several other studies appear to show conceptual, rather than perceptual priming of visual images, as observers are slower to respond to novel images than they are to test images that are i) mirror images of the prime (Cooper, Schacter, Ballesteros & Moore, 1992) or ii) different exemplars from the same category as the prime (Koutstall et al., 2001; Simons et al., 2003) and thus are conceptually related, but spatially non-identical.

By manipulating the test images (see below), task difficulty was increased to avoid the ceiling effects found by Guyader et al. (2004). Finally, in contrast with

Kaping et al. (2007) the current study used amplitude information from real scenes (rather than artificial images) for our adaptation/priming stimuli and created caricatures by exaggerating the difference between the adaptation/priming stimuli and category averages.

2.2 Experiment 1

2.3. Methods

2.3.1. Participants

Eight observers (2 male) took part in Experiment 1 after exclusions (mean age 27.13 years). Eleven observers (6 male) took part in Experiments 2 and 4 after exclusions (mean age 27 years). Seven observers took part in Experiment 3a (5 female; mean age 28 years) and there were 5 participants in Experiment 3b (4 female; mean age 23.6). All had normal or corrected vision and student participants received a small payment for their time. The local ethics committee granted approval and participants gave written consent.

2.3.2. Apparatus

Stimuli were created in MATLAB (Mathworks) and presented on a gamma corrected, 17" CRT Iiyama Vision Master 500 screen (resolution 1152 x 864, refresh rate 85Hz) using a Mac Pro (Apple Computing). A chin rest maintained head position at 85cm from the screen in a darkened room.

2.3.3. Stimuli

Images were photographs from urban and natural scene categories of the LabelMe Database (Russell, Torralba, Murphy & Freeman, 2008) and Oliva and Torralba's (2001) image database. Natural images included forests, open fields and beaches (none contained buildings); manmade images included street-views, highways and buildings (none contained pedestrians). All images were converted to greyscale and scaled to 256 x 256 pixels, subtending 6 x 6 degrees of visual angle. Image luminance and RMS contrast were equalised across the entire image set (mean luminance output of stimuli = 33.0 cd/m^2, (or in RGB, luminance mean = 127.5, RMS contrast = .53; see Bex & Makous, 2002 for RMS contrast equation)).

Amplitude spectra were obtained using a 2D Fast Fourier Transform (FFT) for each image. The average amplitude spectrum was calculated for each stimulus category (across 434 natural images or 351 manmade images).[4] Both spectra were then caricaturised to create prime stimuli with exaggerated 'natural' or 'manmade' characteristics; the difference between the category average and the average amplitude across all images was multiplied by 3[5]. For each trial, either the natural or manmade amplitude caricature was paired with randomised phase to create a unique prime. Finally, the average luminance contrast of prime stimuli was reduced by 20% relative to test stimuli in order to increase priming (Tanaka & Sagi, 1998) and avoid masking (Pearson, Clifford & Tong, 2008; Brascamp et al., 2007). All stimuli presented prior to the test images are henceforth referred to as primes, though at longer durations these were used to investigate adaptation effects.

A 'hybrid' test image was created on each trial by randomly selecting one natural and one manmade image (see Figure 2.5 for example). The phase information in test images was a blend of these two images, containing 25%, 37.5%, 50%, 62.5% or 75% manmade image phase, with the remainder determined by the natural image. Different proportions of phase blend were included to allow us to calculate the psychometric fit of data and thus determine the point at which a 50% threshold for manmade responses was reached; this allowed us to examine the effect of prime category and prime duration on biases in scene categorisation (shifts in threshold).

The amplitude information of test images always matched the average of the whole image set and was thus non-diagnostic, thus we were able to determine whether or not amplitude primes did have an effect on scene categorisation, as test amplitude was unbiased. Primes were presented in greyscale and test images were presented in green to facilitate perceptual segregation[6] of prime and test (a pilot

[4] Note: the number of manmade images was reduced due to stringent exclusion criteria (e.g. excluding those that contained foliage). The final number of images was deemed sufficient, as other studies such as Loschky et al (2010) used a similar number to create naturalistic or manmade texture stimuli (300 images).

[5] The decision to use a multiplier was made in order to maximize the difference between the uniquely natural and manmade components of amplitude primes. A factor of 3 was chosen after examining the resulting spectral templates of prime stimuli; the enhancement of natural versus manmade components was deemed sufficient without departing too far from the original amplitude.

[6] Pilot testing revealed that participants were often unsure about whether they had responded to the prime stimulus or the test image, due to the rapid presentation and short ISI. Test images were more salient when presented in a different colour.

study revealed that participants could not differentiate between prime and test in rapid presentation). See Figure 2.6 for schematic diagram.

Figure 2.5 Example manmade (left) and natural (middle) image with equalised luminance and global contrast and average amplitude across entire image set. Example test image (right), showing a hybrid of the two individual images, with 50% natural, 50% manmade phase.

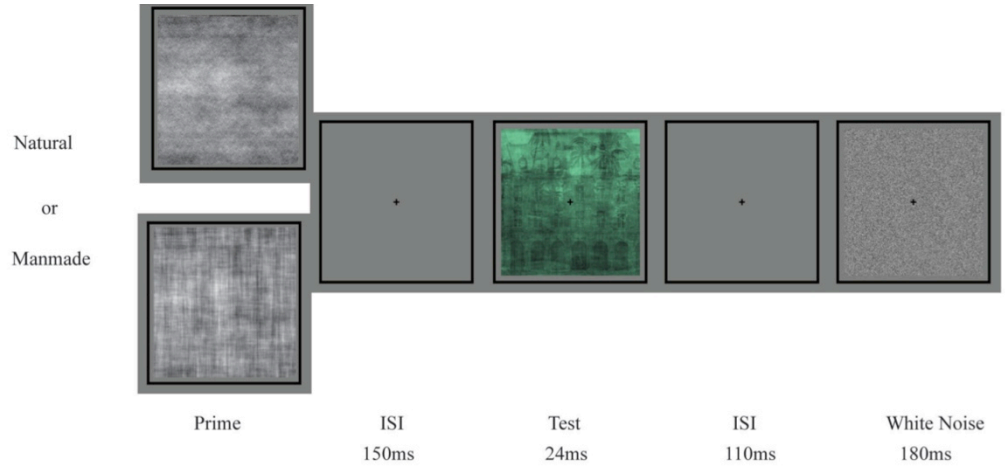


Figure 2.6 Schematic diagram of priming/adaptation trial from Experiment 1. An example caricature stimulus is shown for each prime category. The test image shown contains 50/50 phase blend and amplitude information averaged across the entire image set.

2.3.4. Procedure

A square frame containing a central fixation cross was visible until the participant pressed a button to start the trial. A natural or manmade prime (6 possible durations) was followed by a blank period (150ms inter-stimulus interval/ISI), a brief test stimulus (24ms), a blank duration (110ms ISI) and finally a dynamic, white noise mask (180ms); see Figure 2.6. Prime stimuli were presented for 35ms, 150ms, 300ms, 450ms, 1500ms or 4000ms. For trials in which the prime duration exceeded 150ms, the prime was updated every 150ms with a new random phase, resulting in a dynamic presentation that minimised local adaptation.

Observers classified each test image as predominantly natural or manmade. Trials were blocked by prime duration and prime category, but test stimuli and block order were randomised. Baseline trials in which no prime was presented were completed at the start and end of the experiment. In total, subjects completed 2 blocks of 180 baseline trials (30 reps per phase level but 60 reps for 50/50 phase blends) and 12 blocks (6 for each prime category) of 180 priming/adaptation trials (30 reps per phase level; 60 reps for 50/50 phase blends). Trials were interleaved with respect to the phase of test images; a greater number of 50/50 phase blends was included relative to other phase levels, as these were most informative for calculating 50% thresholds from psychometric functions.

2.4. Results

For each participant, the proportion of manmade responses was calculated for each test phase level in the baseline condition, and in each priming/adaptation condition (2 prime types x 6 prime durations). These data were fit with cumulative Gaussians with variable slope, lapse and guess rates; the proportion of manmade phase at which natural and manmade responses were equally likely was taken as the 50% threshold and used for further analysis. One participant was excluded at this point as they displayed a strong response bias: his / her responses in the baseline condition did not reach 50/50 responding at any phase level. For clear graphical presentation, psychometric fits were also obtained for group data (Figure 2.7a, b). Figure 2.7a shows that following a natural prime/adaptor, observers made more 'manmade' responses (the 50% threshold is shifted to the left), compared to baseline, for all prime durations. Conversely, observers were more likely to categorise images

as 'natural' (thresholds shifted to the right) when the prime was manmade (Figure 2.7b). In other words, participants' responses were biased away from the primed category (main effect of prime type: 2 factor ANOVA on 50% thresholds: $F(1,7) = 39.73$, $p<.01$, $\eta^2 = .85$). Previous research suggests that shorter prime durations result in positive priming (Wiggs & Martin, 1998; Gauthier, 2000) and longer durations result in adaptation (Webster & Miyahara, 1997). However, we found no interaction between prime duration and prime category; the negative effect (responses shifted away from the prime) was constant across durations (no main effect of duration, $p = .352$, $\eta^2 = .14$, no prime type/duration interaction, $p = .363$, $\eta^2 = .14$).

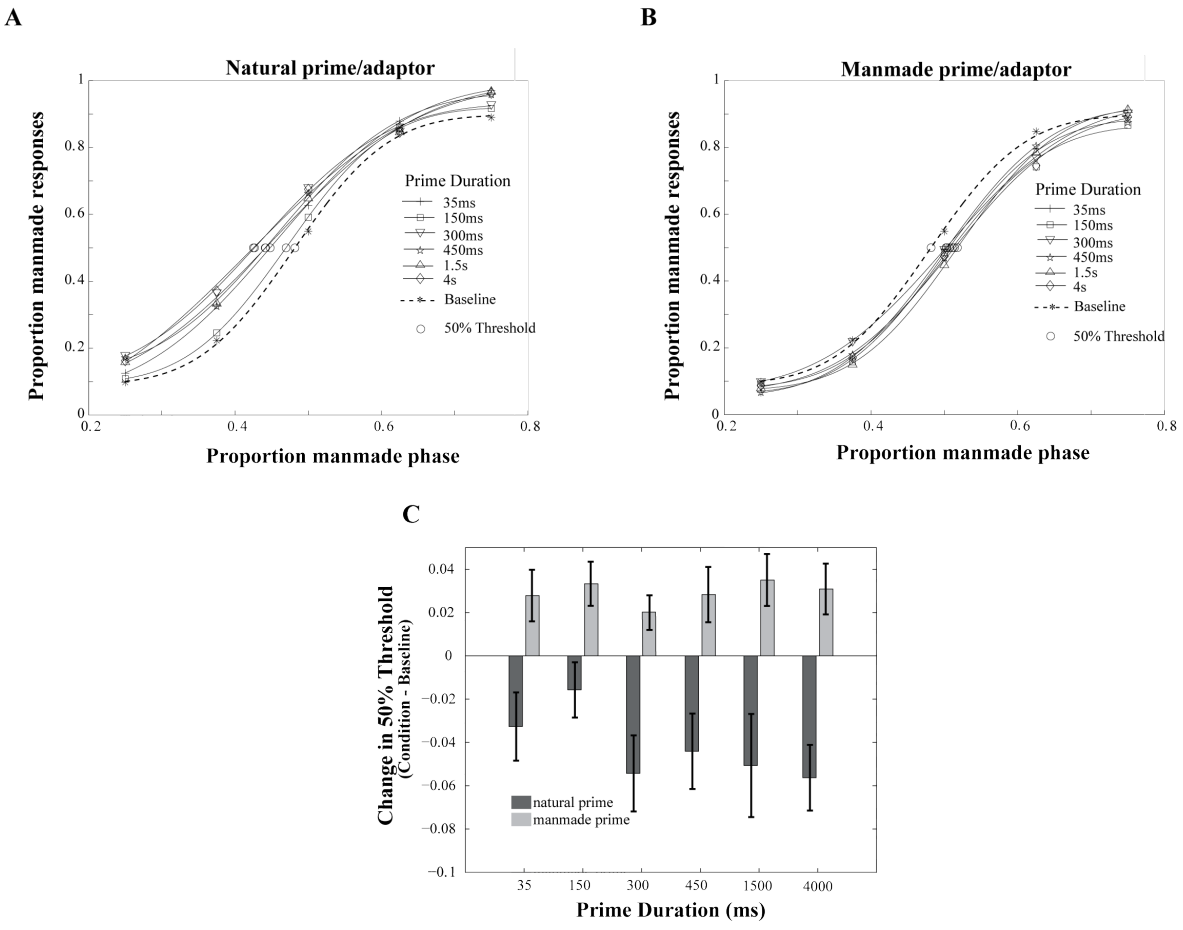


Figure 2.7 Psychometric fits across averaged data ($N = 8$) in Experiment 1 for baseline and priming conditions (6 durations) for (a) natural and (b) manmade primes; the x-axis represents the proportion of manmade phase in test images, from 25% manmade, 75% natural, to 75% manmade, 25% natural; circles indicate 50% categorisation threshold. (c) Bars represent the difference in threshold between each prime duration and baseline. Error bars indicate the standard error of the mean (±1 *SEM*).

Figure 2.7c summarises the data from Experiment 1 as changes in threshold (from baseline) for each prime category and duration. Both prime categories produced significant threshold changes relative to 50% baseline threshold (t-tests on 50% threshold, averaged across prime duration: natural primes: $t(7) = -3.05$, $p = .019$, $d = 1.08$, manmade primes: $t(7) = 4.67$, $p = .002$, $d = 1.65$; statistical significance adjusted for Bonferonni correction due to multiple comparisons: new alpha = .025[7]). In summary, I found no evidence of positive priming. Instead, thresholds were shifted away from the prime category at all durations. Our results are similar to Kaping et al.'s (2007) 'adaptation' effect, though it is unclear whether they should be described as adaptation, or masking effects, since in both cases the effects are known to build up over adaptor/mask duration (e.g. Greenlee, Georgeson, Magnussen & Harris, 1991; Breitmeyer & Ogmen, 2006), whereas our observed threshold shifts were independent of prime duration. Nevertheless, the results demonstrated are in the opposite direction to what would be expected from conceptual priming (the main focus of the study) and thus the negative priming effects are referred to as perceptual suppression; the prime stimulus appears to suppress category-specific components of the test stimulus, leading to increased influence of test components corresponding to the non-primed category.

I propose three possible mechanisms for the suppression effects shown: primes may reduce sensitivity to / recognition of:

(i) The amplitude spectra of test stimuli that are directly diagnostic of image category.

(ii) Diagnostic contours / shapes. Contours that are more common in manmade objects will be defined by spatial frequencies and orientations containing more energy in the manmade primes than in the natural primes, and vice versa. Thus, suppression may influence categorisation by reducing sensitivity to, and recognition of diagnostic contours / shapes, even if global amplitude per se is not used in categorisation.

(iii) Diagnostic objects, via suppression of their diagnostic features. Similarly to (ii), amplitude priming may affect categorisation via impaired recognition of

[7] Standard alpha $0.5/n$ comparisons. All further Bonferroni corrections for multiple comparisons are corrected in this way.

diagnostic objects, even if global amplitude or local contours do not drive categorisation directly.

Experiment 2 tested option (iii) by manipulating our test images to hinder object recognition. If our subjects were using object recognition to guide categorisation in Experiment 1, and if objects were suppressed by the prime stimulus, then one would predict (i) impaired categorisation in the baseline condition and (ii) reduced suppression, when test images do not contain recognisable objects.

2.5. Experiment 2

Experiment 2 followed a similar procedure to Experiment 1, except that a single prime duration was used (24ms; shorter primes were used to eliminate the possibility of very rapid adaptation) and test images were manipulated to disrupt object recognition by either inverting their contrast or inverting contrast and rotating by 180˚. Previous research has shown that object recognition is disrupted by contrast inversion (e.g. Vuong, Peissig, Harrison & Tarr, 2005) and spatial inversion (e.g. Kelley, Chun & Chua, 2003; Shore & Klein, 2000). These additional manipulations leave global amplitude statistics and local phase relationships unaffected (i.e. contours remain intact). However, one would expect that at such a brief test presentation, recognition would be severely impaired for both manmade and natural objects (e.g. buildings, trees). Observers' baseline categorisation for the different test image manipulations will show the extent to which rapid categorisation relies on object recognition; if categorisation is entirely dependent on object recognition, then the slopes of observers' psychometric functions would approach zero in the contrast inverted, rotated condition.

What effects of facilitation/suppression (changes in observers' natural/manmade category boundaries) do we expect with our manipulated test stimuli? Given that object recognition will be severely impaired, categorisation can only be informed by local phase information and global amplitude spectra. If the effects in Experiment 1 were due to the suppression of global amplitude statistics directly, or local phase information, then we should see similar pattern of suppression in Experiment 2. In contrast, if our Experiment 1 effects were due to the impaired recognition of diagnostic objects or structures, then I should find little or no effect of primes, since the contribution of object recognition to the task will be small

in both baseline and prime conditions.

2.6. Method

2.6.1. Procedure

Participants viewed a prime (24ms) followed by a test image that they classified as natural or manmade. The durations of ISI, test and mask remained the same as in Experiment 1. Test images were either normal, contrast inverted (CI), or contrast inverted then rotated by 180° (CIR). As before, the phase information in test images ranged from mostly natural to mostly manmade, while amplitude matched the average across the whole image set. Observers completed 540 baseline trials (3 test types: normal, CI, CIR x 4 phase levels x 30 reps per phase level; 60 reps for 50/50 phase blends) and 720 priming trials (2 test types: CI, CIR x 4 phase levels x 2 prime categories x 30 reps per phase level; 60 reps for 50/50 phase blends); trials were blocked by prime duration and prime category but block order was randomised. Trials were interleaved with respect to the phase of test images.

2.7. Results

As in Experiment 1, psychometric functions were fit to each observer's data for each test and prime type (see Figure 2.8a and b; slope, guess and lapse rate variable); the proportion of manmade phase at which natural and manmade responses were equally likely was taken as the 50% threshold and used for further analysis. Two participants were excluded, as threshold could not be reliably calculated in one of the baseline conditions. Our test manipulations had a significant effect on baseline categorisation; the more distorted the test image, the more 'natural' responses were made (mean proportion of manmade phase required in test images to reach threshold was 0.47, 0.53 and 0.54 for normal, CI and CIR test images respectively; this effect was significant: F (2,20) = 9.72 p=.001, η^2 =.49 with the normal thresholds significantly different from the CI and CIR thresholds: t (10) = 3.27, p=.009, d = 1.37; t (10) = 4.72, p=.001, d =1.60, from Bonferroni corrected post-hoc t-tests, adjusted alpha = 0.0167). In contrast, paired samples t-tests revealed that these stimulus manipulations (CI and CIR) did not produce significant reductions in discrimination (i.e. shallower slopes) relative to the normal baseline condition (CI – normal, p = .344, CIR – normal, p = .471, CI-CIR, p = .856). Together, these results suggest that the presence of recognisable man-made objects biases observers toward a 'manmade' categorisation, and their absence toward a 'natural' categorisation; as objects became

less recognisable (under spatial and contrast inversion), observers were more likely to classify scenes as natural. Though the results from previous studies imply that object recognition is not *necessary* for rapid scene categorisation (e.g. Schyns & Oliva, 1994; Larson & Loschky, 2009) this does not mean that impaired object recognition would have no detrimental effects. Furthermore, we did not distinguish here between recognisability of objects versus the recognised configural layout of natural scenes. Indeed, it might be possible that our 'object' effects were in part due to changes in scene structure, such as the location of characteristic structures or patches of high or low intensity (sky versus trees) which are also affected by image inversion, though further examination would be necessary to make such a distinction.

In conclusion, the results from Experiment 2 imply that categorisation can be accomplished solely via local phase relationships (shapes or contours) and amplitude spectra.

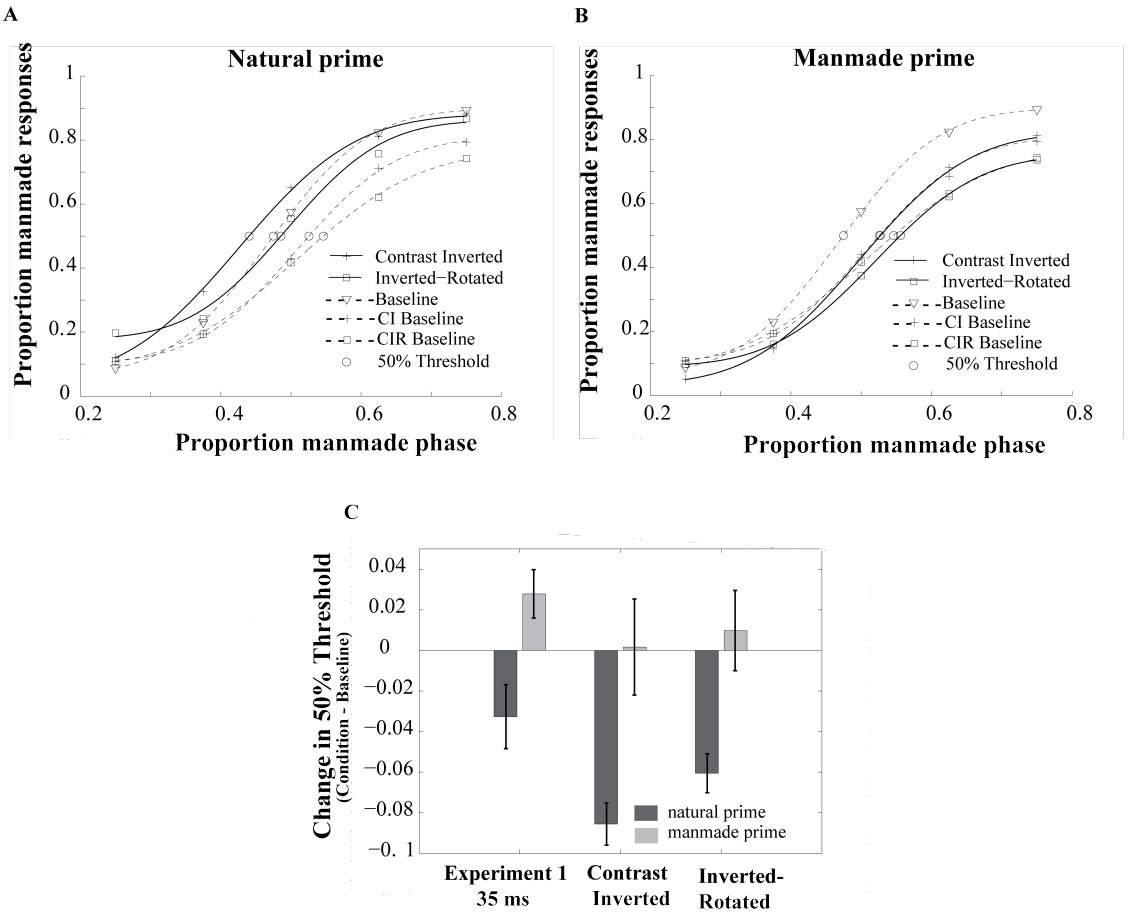


Figure 2.8 Psychometric fits across averaged data (N =11) from Experiment 2 for baseline and priming conditions for (a) natural primes and (b) manmade primes. (c) Bars represent the difference in threshold between each inversion condition and its

baseline, with data from the shortest prime condition (35ms) from Experiment 1 (*N* = 8) shown for comparison. Priming condition is plotted on the x-axis. Error bars indicate ±1 *SEM.*

In the priming conditions there was a significant effect of prime category on categorisation 50% threshold ($F(1,10) = 34.08$, $p<.01$, $\eta^2 = .77$), a marginal effect of test manipulation ($F(1,10) = 4.56$, $p = .059$, $\eta^2 = .31$) and no significant interaction ($p = .81$, $\eta^2 = .00$). The test image manipulations should have reduced or eliminated the contribution of object recognition to image categorisation. Thus, if the effects in Experiment 1 were due to the impairment of object recognition, one would predict reduced, or no effect of primes in the current experiment. In fact, natural, but not manmade primes had a significant effect on categorisation of CI and CIR test images, relative to baseline (See Figure 2.8c; Natural: CI $t(10) = -8.08$, $p<.001$, $d = 1.89$, CIR $t(10) = -6.45$, $p<.001$, $d= 1.26$. Manmade: CI, $p = .81$, d = 0.11, CIR, $p = .32$, $d = 0.138$, alpha value = .0125, adjusted for multiple comparisons). A post-hoc power analysis was conducted in order to test whether the lack of effect for manmade primes was due to insufficient power with the modest sample size ($N = 11$). A paired t-test on the means in this study ($d = 0.11$ and $d = 0.138$) would have required over 100 participants to detect an effect at the recommended statistical power of 0.8 (Cohen, 1988; calculation conducted using ESCI, Cumming, 2001-2011). Since this number is far greater than needed to detect an effect for natural primes, it is assumed that the lack of effect was not due to insufficient power.

Thus, the reduced suppression following manmade primes in the current experiment (when compared to Experiment 1) implies that the effect in Experiment 1 was primarily due to object suppression (with little or no additional effect of disruptions to amplitude or localised phase, i.e. contours). In contrast, natural primes *did* have a significant effect on categorisation of CI and CIR images, shifting observers' thresholds towards manmade, despite object recognition being reduced. I therefore conclude that masking effects following a natural prime with normal (Experiment 1) and manipulated images (Experiment 2) are either due to suppression of a) global amplitude or b) contours, or c) local configural properties, but not due to impairment of object recognition.

2.8. Experiment 3

In Experiments 1 and 2 I found suppression following exposure to natural amplitude stimuli that could be explained by either (i) a direct effect on global amplitude, or (ii) decreased salience of category-specific contours. In Experiment 3 the contributions of these two factors (a and b from above) is directly tested in a simple categorisation task, without priming. As in Experiments 1 and 2, phase information in the test stimuli varied from natural to manmade, but here test amplitude could also be diagnostic. The latter was determined by either (i) the category average (Expt. 3a), or (ii) the amplitude of one of the two component test images (Expt. 3b). If detection of category-specific contours is more important than global amplitude, then one would expect a larger effect of amplitude manipulation in Expt. 3b, where the energy will be strongest for the particular diagnostic contours in the image. In contrast, if global amplitude were the main factor driving categorisation then one would expect larger amplitude effects in Expt. 3a, where the amplitude matches the category average, rather than just one (less representative) category exemplar.

2.9. Method

2.9.1. Procedure

Participants (N = 7 in Experiment 3a, N = 5 in 3b) viewed a test image for 24ms and classified it as natural or manmade. As in Experiments 1 and 2, phase information for each test image was a weighted blend of two randomly selected images (one manmade, one natural). In Experiment 3a test image amplitude was the average across the natural category, the manmade category or across the entire image set. In Experiment 3b the amplitude spectrum for the test was taken from one of the two test images (natural or manmade) or an average of the two individual images. See Figure 2.9a and b for example test stimuli.

2.10. Results

Test amplitude had a significant effect on 50% categorisation thresholds in both Experiments 3a and 3b (see Figure 2.9c and d; ANOVA main effect Expt. 3a: $F(2,12) = 24.09$, $p<.001$, $\eta^2 = .80$, and ANOVA main effect, Expt. 3b: $F(2,8) = 51.63$,

$p<.001$, $\eta^2 = .93$; paired samples t-tests for experiment 3a showed a significant difference in threshold between natural and average amplitude, $t(6) = 3.33$, $p = .001$, $d = 1.14$, and a significant difference between manmade and average amplitude, $t(6) = 3.99$, $p = .007$, $d = 1.81$), both significant after Bonferroni correction (alpha = .025). Paired samples t-tests for experiment 3b showed a significant difference in threshold between natural, $t(4) = 17.11$, $p = 0.00$, $d = 5.75$, but not manmade amplitude, $t(4) = 2.17$, $p = .096$, $d = 1.61$ (alpha = .025).[8]

Comparing across the two experiments, the amplitude manipulation had a larger effect on categorisation when amplitude matched the individual images (Expt. 3b) than when it matched the category averages (Expt. 3a). This effect was marginal for natural amplitude (natural – average threshold difference in Expt. 3a vs. Expt. 3b: $t(10) = 2.79$, $p=.036$, $d = 1.52$; ns with adjusted alpha = .025) but was not significant for manmade amplitude (manmade – average threshold difference: $t(10) = .37$, $p=.72$, $d = .235$; independent samples *t*-tests). The difference between category and individual image amplitude shows that the amplitude-phase relationship (i.e. the salience of diagnostic contours) is a more important factor in scene categorisation than global amplitude per se. It suggests that the suppression following natural primes found in Experiments 1 and 2 may have been due to a reduction in visibility of diagnostic contours, rather than of natural global amplitude.

[8] It is of note that only 6 participants would be required to detect a significant effect at the recommended level.

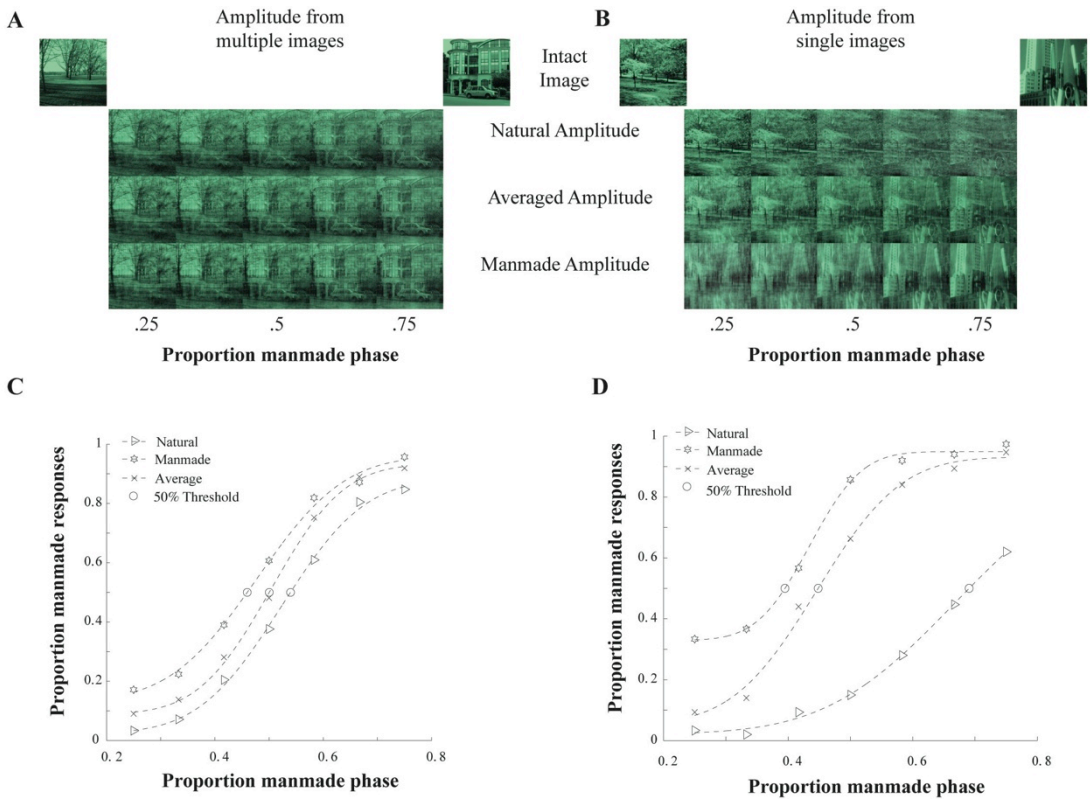


Figure 2.9 A and B: pair of intact images, followed by example test images from Experiments 3a and b, with phase blend from 75% natural, 25% manmade, to 25% natural, 75% manmade, paired with natural, averaged or manmade amplitude. Amplitudes in 4a are category averages, while amplitudes in 4b originate from either the natural or manmade intact image, or an average of the two. C and D: Psychometric fits across averaged data from Experiments 3a (N = 7) and 3b (N = 5) respectively. Fits are shown for average, natural and manmade amplitude categories.

2.11. Experiment 4

In Experiments 1 and 2 I failed to find a positive effect of priming on performance and instead found suppression. Although this contradicts the positive priming effect reported by Guyader et al. (2004), their results were based on response time (RT) data rather than on shifts in the natural/manmade categorisation boundary; performance accuracy was at ceiling in their task. In Experiment 4, therefore, I investigated potential priming effects on RTs.

2.12. Method

2.12.1. Procedure

The procedure was identical to Experiment 1 with two exceptions: (i) there were 2 prime durations (1 or 3 frames: 11.8 or 35ms) and (ii) participants were asked to respond as quickly and accurately as possible. Prime durations were chosen to be similar to Guyader et al. (2004; 10ms) and the shortest prime from Experiment 1 (35ms). Test image amplitude always matched the average of the whole image set and was thus non-diagnostic, while test image phase contained 25%, 37.5%, 50%, 62.5% or 75% manmade image phase, with the remainder determined by the natural image.

2.13. Results

2.13.1. Threshold analysis

As before, cumulative Gaussians were fit to each participant's data to obtain 50% performance thresholds for baseline and priming conditions (Figure 2.10a and b). Two participants were excluded, as categorisation thresholds in one or more conditions could not be reliably estimated. Observers' categorical responses followed a very similar pattern to the data from Experiment 1; there was a significant effect of prime type ($F(1,10) = 40.68$, $p<.001$, $\eta^2 = .803$) but not of prime duration ($p = .61$, $\eta^2 = .03$) and no significant interaction ($p = .92$, $\eta^2 = 0$). The natural prime produced a significant shift in categorisation away from the prime category, relative to baseline (natural: $t(10) = 4.61$, $p = .001$, $d = 0.216$) and the effect following manmade primes was marginal but non-significant after Bonferroni correction (manmade: $t(10) = 2.35$, $p=.041$, $d = .106$; adjusted alpha = .025).[9]

2.13.2. Response Time Analysis

RTs for correct responses were analysed for each participant; correct responses were defined as those that matched the dominant phase of the test image, while in in 50/50 phase trials all responses were classed as correct. RTs were log-transformed, before excluding outliers (RTs more than ±2.5 standard deviations from the participant's condition mean); the mean proportion of trials excluded per

[9] Note that around 50 participants would be required to detect a significant effect of manmade primes to reach the recommended statistical power.

condition was 1.41% and the maximum in any condition was 2.2%. RT data were transformed back to milliseconds before further analyses and are displayed in Figure 2.11a and b; the results are organised by correct response (natural/manmade) and by prime consistency with test (C / IC).

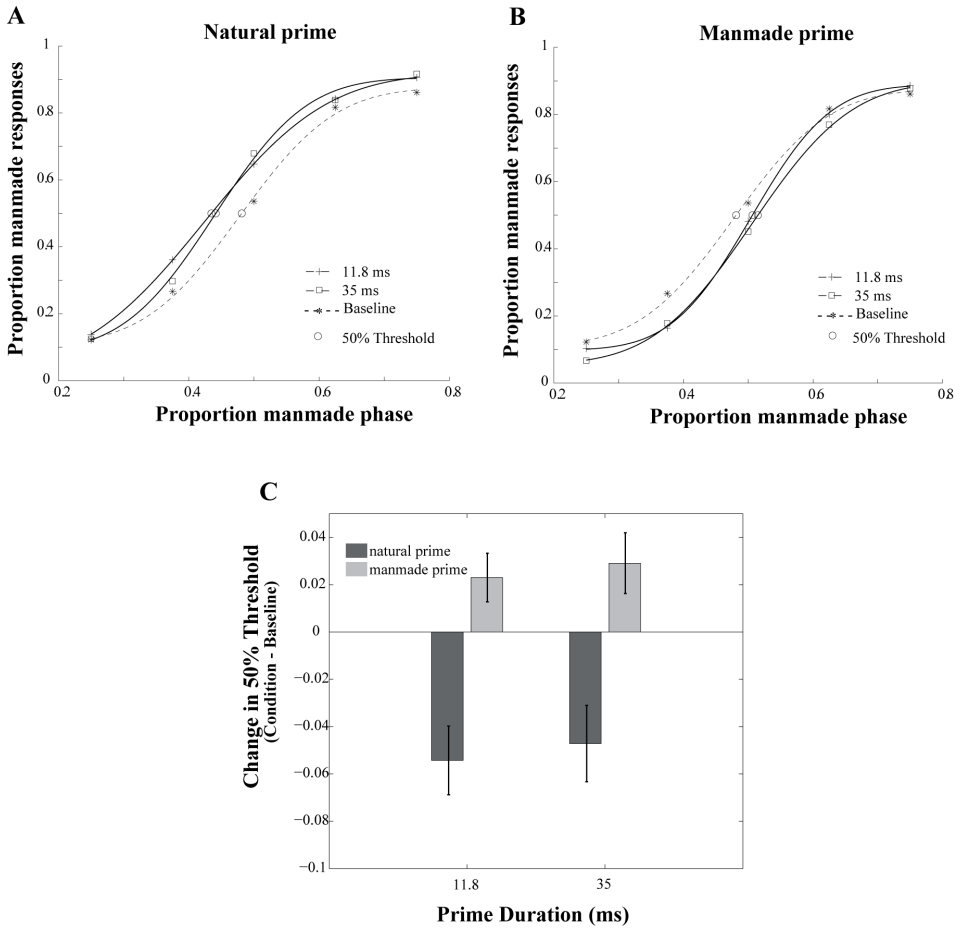


Figure 2.10 Psychometric fits across averaged data for Experiment 4 (N = 11). Fits for baseline and priming conditions for (a) natural primes and (b) manmade primes. (c) Bars represent the difference in threshold performance in the reaction time task for each experimental condition relative to the baseline. Error bars indicate ±1 *SEM.*

In the baseline condition, a repeated measures ANOVA found that mean response times were not affected by test category, proportion phase blend or their interaction (p = .33, p = .30 and p = .76 respectively). Incidentally, Joubert et al. (2009) also found that RTs were unaffected by the percentage of phase noise blended with their test images, despite its effect on participants' accuracy. They did, however, find that responses were significantly faster when test images contained their original

amplitude than when test amplitude was matched to the average across manmade and natural images, analogous to our categorisation effects in Experiment 3.

To explore any facilitation (priming)/suppression effects, trials were classified as consistent (C: prime and test category matched) or inconsistent (IC: prime and test were from different categories). However, in contrast with Guyader et al. (2004), who found a benefit of prime-test consistency on RT, we found no evidence of such an effect (4-factor (prime duration, phase level, prime category, consistent/inconsistent) repeated measures ANOVA on RTs, effect of prime-test consistency $p>.05$). There was a significant effect of phase level ($F(2,20) = 16.54$, $p<.001$, $\eta^2 = .62$), but no main effect of prime ($p = .66$), prime test-consistency ($p = .77$) and no significant interactions (*ps* all > .3). Further investigation into the effect of phase on reaction time showed that RTs decreased as the proportion of the dominant phase increased (RTs for high (75%) phase tests were significantly faster than for the intermediate (67.5%) phase tests, $t(10) = 3.89$, $p=.003$, $d = 2.46$, intermediate vs. mid (50%) comparison not significant ($p = .083$, $d = 1.2$); from post-hoc t-tests with Bonferroni correction; adjusted alpha = .025).

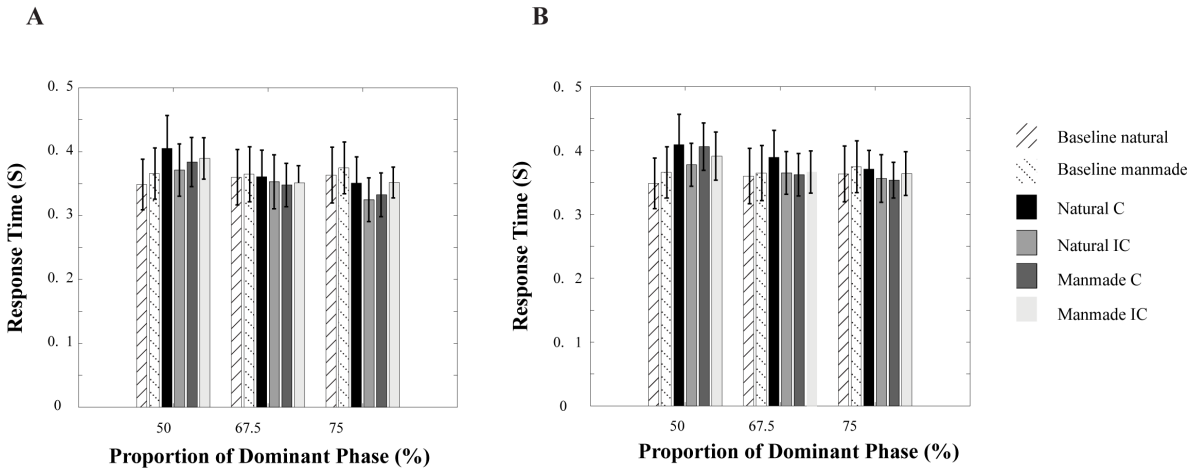


Figure 2.11 Response time data from Experiment 4, averaged across participants (N = 11) for (a) 1-frame prime condition and (b) 3-frame condition. Bars represent correct responses in baseline conditions (no prime), and prime conditions in which the prime was either consistent (CS) or inconsistent (IC) with the dominant phase of the test. Bars are labelled by correct response (natural/manmade) and by prime/test consistency. Error bars indicate ±1 *SEM.*

2.14. Discussion

The current studies investigated the effects of amplitude information on manmade vs. natural categorisation of briefly presented scenes. Across three priming studies I found no evidence of conceptual priming or adaptation on categorisation responses (Experiments 1, 2, 4), or response times (Experiment 4). I did, however, find category-specific suppression, such that, irrespective of prime duration, observers' responses were biased away from the prime category. Experiment 2 showed that suppression following manmade primes in Experiment 1 was primarily due to impaired object recognition, while Experiment 3 revealed the importance of naturalistic and manmade contours for scene categorisation. The implications of these results are discussed in relation to the proposal that global amplitude is diagnostic of scene category.

2.14.1. Amplitude Suppression

The initial hypothesis was that positive priming (brief primes) and negative adaptation effects (longer primes) of global amplitude would be found. However, the data were more consistent with perceptual suppression and are thus comparable to previous evidence of backward masking of amplitude on scene categorisation. For example, Loschky et al. (2010) presented participants with intact images, followed by an amplitude mask. Masks containing amplitude information from real images decreased scene gist accuracy by 12% relative to white noise masks with flat amplitude spectra. Although this suggests that scene amplitude *is* important for scene categorisation, tests and mask images were not from the same basic category in this task. Furthermore, Loschky et al. (2007) found that the magnitude of masking did not depend on whether the mask amplitude was consistent or inconsistent with the test image. Together, these studies suggest that the visual system is not sufficiently sensitive to category differences in unlocalised amplitude for the global amplitude spectrum to be diagnostic.

Why then, did these results show there to be a category-specific suppression in all three of the 'priming' experiments, though Loschky and colleagues did not? There are several possible explanations for this: i) primes in the current study were caricatures, designed to exaggerate the uniquely natural or manmade characteristics of the averaged amplitude spectra. Suppression may, therefore, have reduced the

salience of the most diagnostic information for one scene category without affecting the unique components of the non-suppressed category. ii) Loschky et al. (2010) used intact scenes as test images, whereas our tests contained an equal blend of both natural and manmade amplitudes. Suppressing one amplitude category in our study would leave the unique components of the other category's amplitude relatively intact. iii) Our test images were hybrids containing a blend of natural and manmade phase; the amplitude information remaining after suppression could bias categorisation due to a change in the salience of naturalistic and manmade contours, which are defined by the interaction between amplitude and local phase.

The difficulty with evaluating the contribution of these three components to category-specific amplitude suppression is that it is difficult to separate the effects of global amplitude from those of an interaction between amplitude and phase. Phase is specified by localised amplitude and thus any weakening of amplitude following suppression will inevitably weaken the strength of phase information also. This means that neither the results from our studies (1, 2, 4) nor those from adaptation paradigms (Kaping et al., 2007) can conclusively provide evidence that amplitude alone is diagnostic of scene category, since phase is necessarily a confounding factor in each study.

2.14.2. Object Recognition

In Experiment 2, I introduced stimulus manipulations that eliminated, or severely disrupted recognition of objects[10]. Contrary to the large suppressive effects following natural primes with normal, upright test images (Experiment 1), suppression was greatly reduced for both manmade CI and CIR test images. I conclude from this that the suppression in Experiment 1 following manmade primes was due to the effect of amplitude suppression on object recognition; when objects were effectively made unrecognisable by our test manipulations, the addition of amplitude suppression had no effect on categorisation. This implies that object recognition was the primary factor in the recognition of manmade scenes.

[10] As discussed in section 3.2, the inversion and rotation of scenes would affect both object recognition and scene layout, though further studies would need to be conducted to distinguish between the two.

Natural primes did lead to suppression despite the inhibition of object recognition. The loss of sensitivity to natural amplitude increased the relative weighting of manmade information, and this led to a significant shift in categorisation, though it is unclear whether this was due to suppression of global amplitude information, or its interaction with local phase, i.e. contours. Interestingly, there was little or no suppression following exposure to manmade amplitude primes. One would expect suppression of manmade amplitudes to decrease the salience of manmade contours. Since this was not the case, it implies that manmade shapes or contours are more resilient to amplitude suppression than natural ones. This is plausible if we consider the differences between the spectral signatures of manmade and natural scenes. Oliva and Torralba's (2001, 2003) model of scene gist recognition contains a set of perceptual dimensions that could provide a meaningful description of a scene. In their model, 'naturalness' captures the information critical to determining whether an image should be categorised as natural or manmade, where scenes that exhibit less naturalness have a higher proportion of vertical and horizontal orientations at medium and high spatial frequencies. In contrast, close-up views of images with high naturalness have low spatial frequency horizontal contours (i.e. horizon) as well as diagonals covering the range of spatial scales. It is known that brief periods of adaptation to natural images lead to a marked loss in sensitivity at low spatial frequencies, both for static images (Webster & Miyahara, 1997) and movie sequences (Bex, Solomon & Dakin, 2009). Thus, the selective suppression effects in this study could be explained by the decreased salience of low spatial frequency information that is diagnostic for natural scenes.

2.14.3. Amplitude-Phase Interaction

Experiment 3 sought to investigate the role of amplitude-phase interaction on scene categorisation. Experiment 3a demonstrated that both test phase (proportion of manmade vs. natural phase) and test amplitude biased categorisation. The results from Experiment 3b implied that the effect of amplitude was somewhat stronger when the amplitude originated from one of the individual test phase images, rather than the category average; contours and shapes are more salient when the amplitude matches the dominant phase (compare top left and bottom right images in Figure 2.9a and b). I argue that is this balance of naturalistic and manmade contours, rather

than the global amplitude spectra (which is more diagnostic in Figure 2.9a than Figure 2.8b) that appears critical for classification.

2.14.4. Amplitude and Response Time

Guyader et al. (2004) found that observers were significantly faster to categorise test images following an amplitude prime from the same basic category (city or beach) than a different category. However, the results from Experiment 4 failed to replicate this effect. One reason for this difference could be the generality of categories used: our prime images were caricatures of average natural or manmade amplitudes, those in the Guyader et al. study were exemplars from the same basic category (beach or city). In their study, therefore, the priming effects may be due to the greater similarity of the amplitude prime to the test image. Indeed, Guyader et al. found that their priming effect disappeared when city scenes contained more horizontal orientations, as amplitudes became less distinct from beach scenes that are dominated by horizontal contrast energy. It is also of note that the Guyader et al. study did not include baseline (no prime) RTs, so it is not possible to determine whether RTs overall were faster or slower following exposure to amplitude information. An interesting finding in this particular study was that there was a significant suppressive effect for categorisation of natural scenes following natural primes; this is likely due to be linked to the demands of the task, as participants were asked to respond as quickly as possible, which they were not asked to do in the other three studies.

2.14.5. Conclusion

In conclusion, although there may be statistical differences in amplitude spectra between scene categories (e.g. Torralba & Oliva, 2003), I found no convincing evidence that observers use unlocalised amplitude to differentiate natural and manmade images under brief presentation. Instead, the evidence across our experiments supports the notion that the distribution of contour and object information in an image is far more important in rapid scene categorisation. Further investigation into the diagnostic role of amplitude information must avoid the effects of negative amplitude-phase interaction; designs that can reliably lead to facilitation would be superior in this case, as one would be able to test whether the visual system

associates a prime consisting of unlocalised naturalistic or manmade amplitude with a test that is ambiguous. Here one could study the addition of potentially diagnostic information prior to categorisation, rather than decreasing its salience.

Chapter 3

Can prior exposure to characteristic scene amplitudes tip the balance in binocular rivalry?

Chapter 2 summarised four experiments that investigated the role of Fourier amplitude statistics in rapid scene categorisation. The results highlighted the importance of scene layout, recognisable objects and diagnostic contours in making a rapid natural/manmade distinction. However, the primary aim of the studies was to examine whether unlocalised amplitude could be diagnostic of scene category. As exposure to amplitude primes led to the suppression of same-category images rather than facilitation as predicted, it became difficult to isolate any unique contribution of amplitude information; amplitude suppression would also decrease the salience of contours defined by phase. The objective of the present experiment was, therefore, to use an alternative experimental paradigm that would allow us to test for the effects of amplitude facilitation on subsequent categorisation of natural/manmade scenes.

3.1. Introduction

As summarised in detail in Chapter 2, there is a wealth of literature showing that humans are remarkably good at rapidly categorising natural scenes (e.g. Rousselet, Joubert & Fabre-Thorpe, 2005, Joubert, Rousselet, Fize & Fabre-Thorpe, 2007; Greene & Oliva, 2009). Indeed, categorisation is still possible even when objects are difficult to identify due to being shown in low-resolution peripheral vision (Larson & Loschky, 2009), bandpass-filtered so as to only contain low spatial frequencies (Schyns & Oliva, 1994) as well as when images are luminance contrast inverted and spatially rotated (Chapter 2, Experiment 2). Thus, several research groups have focused on determining exactly what perceptual information contained within natural scenes allows such efficient, rapid categorisation. One theory is that humans have internalised characteristics of the natural world that are associated with particular scene categories (i.e. forests and cities), and use this characteristic structure to discriminate between categories. For instance, horizontal and vertical edges are characteristic of manmade scenes that are dominated by buildings. Natural

scenes also tend to be dominated by cardinal orientations, there is also relatively more contrast energy at oblique orientations, as might be expected within foliage. Image analysis of natural and manmade scenes supports the idea that characteristic differences exist between scene categories. For instance, Oliva and Torralba (2001, 2003) demonstrated that the distribution of contrast energy across spatial frequency and orientation is computationally and phenomenologically different between natural and manmade scenes and between basic scene categories. In addition, they also showed evidence of a relationship between amplitude spectra and observers' descriptions of scenes (e.g. depth or naturalness; Oliva & Torralba, 2001). Furthermore, the results of two psychophysical studies examining the effects of amplitude in priming (Guyader, Chauvin, Peyrin, Hérault & Marendaz, 2004) and adaptation (Kaping, Tzvetanov & Treue, 2007) indicate that prior exposure to amplitude spectra can influence subsequent scene categorisation, and may thus be diagnostic.

Nevertheless, an alternative theory postulates that phase must be intact (amplitude must be localised) in order to inform about scene category, since phase alignment is necessary to form the contours that make up recognisable shapes and edges. Indeed, evidence shows that images cannot be categorised when phase has been disrupted by more than about 50% (e.g. Loschky & Larson, 2008; Joubert et al., 2009), and this suggests that amplitude alone is insufficient for accurate categorisation. However, the argument put forward in Chapter 2 was that phase scrambled images are poor representations of real scenes, since they contain no contours or shapes. It is possible then, that amplitude information could be diagnostic of scene category, but that detection of such a phenomenon is sensitive to experimental design. The purpose of Chapter 2, therefore, was to revisit priming and adaptation paradigms in order to investigate whether prior exposure to an amplitude stimulus would bias subsequent perception of ambiguous natural/manmade images.

The results of the experiment described in Chapter 2 showed that prior exposure to scene category amplitude actually *decreased* the proportion of subsequent trials that were reported to be from the same category. Thus, there was no evidence of conceptual amplitude priming and the results were more consistent with amplitude suppression. Since amplitude is a key component of phase, then we could not conclude that amplitude alone, and not phase was crucial to the effects

observed, thus the question of whether amplitude is diagnostic remained unresolved. The motivation for the experiment outlined in the present chapter was, therefore, to find a paradigm that would reliably lead to facilitation of previously seen images or information. A method utilising both facilitation and suppression of dominance in binocular rivalry was chosen for several reasons that are discussed below.

3.1.1. Binocular Rivalry

In binocular rivalry, each eye is presented with a distinct image. Since only one object can occupy a space at any one time, the images compete for dominance with only one interpretation seen at any given time, with the dominant percept changing intermittently between the two eyes images. Early rivalry studies focused on competition between the two eyes (e.g. Blake, 1989), but more recent studies show evidence that dominance is stimulus- rather than eye-dependent (Logothetis, Leopold & Sheinberg, 1996; Kovács, Papathomas, Yang & Fehér, 1996). For instance, Kovács et al. showed that when patchwork stimuli (e.g. containing red and green dots, see Figure 3.1) were presented to each eye, the overwhelming percept was of either red or green dots, indicating that *pattern coherence* can also determine the outcome of binocular rivalry. This demonstrates that there is a hierarchy of visual competition and in this case the visual system selects the image that is more meaningful or relevant as a whole, rather than the input from a single eye. Patterns of dominance in binocular rivalry can be evaluated in order to understand the process of perceptual selection by the visual system. There are many other factors that can affect patterns of stimulus dominance and we can examine these to better understand the importance of certain stimulus properties.

Several low-level stimulus properties have been shown to increase dominance in binocular rivalry, such as high contrast (e.g. Mammassian & Goutcher, 2005), flicker (e.g. Tsuchiya & Koch, 2005) and naturalistic spatial frequency (spectral slope; Baker & Graf, 2009). Evidence also shows that dominance can be affected by pattern coherence across the two rival stimuli (Kovács et al., 1996) and by prior exposure to the same stimulus (e.g. Pearson, Clifford & Tong, 2008). Importantly, these studies demonstrate that rivalry can be resolved at multiple levels and that top-down feedback from higher cortical areas can also affect rivalry. The current study aims to explore high-level effects of scene category amplitude on rivalry dominance.

Figure 3.1 When the patchwork stimuli shown below are presented to different eyes during binocular rivalry, instead of seeing just one eye's image as would be expected in eye rivalry, sometimes participants perceived only red or only green spots, indicating that rivalry was resolved by pattern coherence. Adapted from Kovács et al., 1996.

3.1.2. Motivation Behind Binocular Rivalry Approach

A binocular rivalry priming paradigm improves upon the priming method used in Chapter 2 (experiments 1, 2, 4) for several reasons: Firstly, it was difficult to find prime durations and ISI durations that would be appropriate for both priming and adaptation in Chapter 2. In contrast, rivalry priming studies have reliably shown both facilitative and suppressive effects that are dependent on the contrast of prime stimuli and thus could guide experimental design (e.g. Pearson, Clifford & Tong, 2008; Brascamp et al., 2007). For instance, Pearson et al. (2008) demonstrated that perception of a previously suppressed grating stimulus could be both facilitated and suppressed when shown at a low- or high contrast in the intervening period between rivalry trials. A similar effect of stimulus contrast, also using grating stimuli, was demonstrated by Brascamp et al. (2007), though they additionally showed that the duration of the prime stimulus (from short to long) caused a similar effect, indicating that it is the strength of the prime stimulus that leads to biases in rivalry. Furthermore, Brascamp et al. tested with several prior exposure- and blank interval durations and reported their effects on facilitation and suppression. Thus, such studies not only provide a comprehensive guide to planning aspects of the experimental design (e.g. choosing inter-stimulus interval and stimulus contrast), but

with careful selection of experiment parameters, also make it reasonable to predict both facilitation and suppression following exposure to complex natural images. Importantly, the results from Chapter 2 lacked the critical facilitation component that is necessary to show the effects of amplitude without being confounded by phase, as is the case in masking; the method chosen here may rectify this problem.

A second benefit of the binocular rivalry method is that, as rivalry stimuli are often presented in contrasting colours (e.g. one red, one blue), participants can simply report which colour stimulus they have seen, thereby avoiding any cognitive bias involved in stating the content of the image. The prevention of cognitive bias in this design is an improvement relative to the previous study, in which participants had to state whether a morphed image appeared to be more natural or manmade. Cognitive bias could have been introduced here, for instance, if participants reported any test image containing a tree to be natural, and anything with distinct vertical lines to be manmade, though morphed images contain both natural and manmade components. Indeed, vertical lines could project from both a tree and a building, thus it is clear that decision rules could have introduced error or noise in the measures.

3.1.3. Study Outline

In the current study, participants will be presented with a priming stimulus, followed by a binocular rivalry pair, consisting of one city and one forest scene. One test image will be presented behind a translucent red filter, and the other behind a blue filter. Participants will indicate by button press the colour (and thus category) of the image they saw during rivalry. Prime contrast will be manipulated to alter its influence on subsequent rivalry, in line with research discussed above (Brascamp et al., 2007; Pearson, Clifford & Tong, 2008). The main independent variable is the amplitude of the prime stimulus; the current experiment will examine whether the specificity of the amplitude within the prime is sufficient to facilitate or suppress dominance of the rivalry image that matches prime category.

Why might one expect a facilitation effect of amplitude using this paradigm, when priming was not found in Chapter 2? Firstly, following parameters from previous studies (Brascamp et al., 2007; Pearson, Clifford & Tong, 2008) means that the procedure itself should be reliable; the main deviation is that the stimuli used in the current study will be complex natural scenes rather than simple gratings.

Secondly, test duration in the rivalry task will be 1500ms, but was only 24ms in Experiment 1. Masks in natural scene categorisation tasks typically follow (or precede) short test stimuli (e.g. Gaspar & Rousselet, 2009; Fei Fei, VanRullen, Koch & Perona, 2002; Greene & Oliva, 2009), and masks become less effective as the test stimulus is seen for longer (Bacon-Macé, Macé, Fabre-Thorpe & Thorpe, 2005), thus one would not expect masking effects using this design, as test duration is much longer. Furthermore, the task itself is different, not only because participants are asked to respond to the colour of a unique, intact image, rather than consciously interpret one that has been morphed between two separate categories (a hybrid), but also because any bias due to prior exposure might result in only one of the rivalry pair being perceived at all. In addition, there is existing evidence that the visual system is biased toward perceiving images that are close to a natural ($1/f$) amplitude spectrum relative to other slopes during binocular rivalry (Baker & Graf, 2009). If the amplitude statistics of one image of the test pair are pre-activated during the prior exposure stage, then it may bias subsequent rivalry. If the amplitude energy in the prime was not associated with the intact image, then one would expect no effect. Finally, in this study the image sets were simplified to include only forest (natural) and city scenes (manmade), thus maximising the difference in amplitude and structure between categories. This should increase the experimental power too and is helpful since binocular rivalry can be noisy. For instance, attention, eye movements or blinking can all affect dominance periods in binocular rivalry (see Blake, 2001, for review).

3.1.4. Conditions and Predictions

The four main conditions of the rivalry experiment are outlined below, along with an explanation of what they test for, and their predicted outcomes. Forest and city primes (or white noise in baseline condition) will be shown at four contrast levels, selected so as to produce both facilitation and suppression, as informed by previous research (Brascamp et al., 2007; Pearson, Clifford & Tong, 2008).

3.1.4.1. Condition 1: Baseline

The prime in the baseline condition will be a white noise stimulus. Here, I predict no effect of the prime on subsequent rivalry, as a white noise prime should be

no more similar to forest or city images, making the contrast of the prime inconsequential. This condition should reveal any baseline bias in dominance between forest and city test images.

3.1.4.2. Condition 2: Match

The prime stimulus will be an intact (no modification to amplitude spectra or phase information) forest or city scene that is identical to one image in the rivalry pair. The aim of this condition is to replicate the expected pattern of facilitation and suppression, which will be evident if there is an interaction between prime category and prime contrast, as they affect the dominance of the forest versus city image. It is worth noting that our stimuli are broadband and orientation is unconstrained, whereas similar studies have mainly used simple grating stimuli (Brascamp et al., 2007; Pearson et al, 2008), thus there is some uncertainty as to whether the effect will be replicable. Nevertheless, I predict that primes will facilitate/ suppress test image dominance here due to i) matching amplitude information, ii) matching phase (structures/contours), or iii) repetition of conceptual category (e.g. natural/manmade or basic category).

3.1.4.3. Condition 3: Mismatch

In this condition, the prime and one of the test images will be different exemplars from the same category (forest/city). Here, prime/test amplitude will be similar but not identical, while also sharing common objects or features and conceptual category. A main interest of this condition is to examine whether the visual system treats similar amplitudes as being from the same category (conceptual, high-level effects) or whether low-level features are the crucial components.

Potential sources for prime-test facilitation and suppression are comparable to those for the matching condition: i) similarities in low-level amplitude information, ii) similarities in phase (structures/contours), and iii) repetition of conceptual category. Indeed, several previous studies have shown that responses can be influenced when observers are presented with a similar, but not identical, visual stimulus prior to testing. In a priming study for instance, Koutstall et al. (2001) found participants were faster to judge the size of novel objects (relative to repeated objects) when a different exemplar of the same image had been shown previously (e.g. coloured

umbrella). In addition, Brascamp et al. (2007) showed their facilitation/suppression pattern in binocular rivalry was somewhat resilient to changes in grating phase between prime and test image. Though, as with the above study, Pearson, Clifford and Tong (2008) found that the effect of prime on rivalry decreases when deviations in orientation between prime and test were introduced. Furthermore, as mentioned above, our stimuli are complex, rather than single-orientation gratings, thus similarities or differences between prime and test will be more complex. It seems plausible that some facilitation or suppression should occur due to any of the three points outlined above, and this would present as an interaction between contrast and prime category. However, I predict a reduced effect relative to the match condition; if amplitude is informative then the visual system is likely to be sensitive to amplitude differences between stimuli, rather than the effect being purely conceptual.

3.1.4.4. Condition 4: Amplitude Component

Prime stimuli will contain the amplitude component of one of the rivalry pair, but with randomised phase. Importantly, this final condition tests whether the amplitude component image is associated with its intact counterpart, as facilitation (perceptual priming) should occur if the visual system recognises having previously processed the same information. Prime-test facilitation or suppression could occur due to i) matching amplitude information (low–level), or iii) repetition of image identity or conceptual category. This condition lacks outcome ii) as the prime stimulus contains no contours.

Together, these four conditions address two questions: 1) Does the visual system recognise that an amplitude component image contains the same information as its intact counterpart? 2) Does the visual system associate characteristic amplitude spectra with scene categories?

3.2. Experiment 1

3.3. Method

3.3.1. Participants

Eight participants (4 male) took part in Experiment 1 (mean age 29.25 years); a subset of three participants (1 male) from Experiment 1 took part in Experiment 2 (mean age 29.33). All had normal or corrected vision and student participants

received a small payment for their time. The local ethics committee granted approval and participants gave written consent.

3.3.2. Apparatus

Stimuli were created in MATLAB and presented on a 17" CRT ViewSonic Graphic Series G90FB screen using an AppleMac Pro (Apple Computing). The screen resolution was 1152*870 with a 75Hz refresh rate. A chin rest was used to maintain fixed head position 85cm from the screen and stimuli were viewed through a mirror stereoscope, in a darkened room.

3.3.3. Stimuli

Images were photographs of forests and city scenes taken from the LabelMe Database (Russell, Torralba, Murphy & Freeman, 2008) and Oliva and Torralba's (2001) image database. All images were initially matched for RMS contrast. Primes were intact forest or city scenes (match/mismatch conditions), their amplitude component with randomised phase (amplitude component condition), or white noise (baseline). Primes were presented at 7, 20, 50 or 90% of the mean RMS contrast; contrast values were selected so as to span the range of values likely to incorporate both facilitation and suppression effects (Brascamp et al., 2007), though very low contrasts were avoided to prevent perceptual stabilization (see Pearson & Brascamp, 2008, for review). Test images were intact and presented at 100% contrast, though a transparent red or blue filter (created by modifying the alpha channel in an RGBA image) was placed over each image for distinction; the luminance of the red filter was set for each participant using colour pre-test data. Stimuli subtended 2.16 degrees of visual angle and were surrounded by a box containing randomised blocks of regularly spaced black and white pixels; this was shown in both eyes to aid fusion (D.V.A. = 4.31).

3.3.4. Procedure

A pre-test was run prior to the main experiment in order determine the luminance at which red/blue stimuli were equally dominant, thus avoiding colour bias in the main experiment. Each eye was presented with a red or blue disc and participants were asked to hold down a key to report the dominant colour percept.

Discs were shown continuously for 60 seconds, followed by a blank interval of 45 seconds, which was introduced to prevent afterimages. The luminance of the red disc was 25, 50, 75 or 100% of the luminance of the blue disc (as in Baker & Graf, 2009) and the left/right position of the discs was counterbalanced. All trials were completed twice, making a total of 16 trials with 4 repetitions per red value. From these results, the red luminance value at which 50% threshold was reached was estimated for each participant. Participants with over 90% eye dominance in this pre-test were not invited to complete the main experiment.

In the main experiment, participants saw a fixation cross, followed by the binocular presentation of the prime stimulus for 800ms, followed by a 70ms ISI, a 1500ms presentation of the rivalry stimuli and 150ms blank period (see Figure 3.2; a relatively long rivalry period was chosen as the focus was on transient preference rather than initial dominance (Mamassian & Goutcher, 2005)). Participants made a button press to report the colour of the 'most dominant' stimulus and were asked to report the most salient colour in cases of piecemeal rivalry. Three conditions were run initially: match, where prime and one test image were identical, mismatch, where the prime and one test image were different exemplars from the same scene category, and the amplitude component condition, in which the amplitude of the prime matched the test, but prime phase was randomised (see Figure 3.2). Colour and position (left/right) of test stimuli were counterbalanced and trials were interleaved, with 30 reps per condition, making a total of 2880 trials (4 * contrast, 2 * prime category, 3 * prime condition, 2 * colour, 2 * test position), completed in blocks of 60 trials. An additional prime condition (baseline, white noise) was run separately from the original three conditions, on the same participants (480 trials).[11]

[11] A baseline condition was not run initially, as white noise can still lead to masking (Loschky et al. 2010) and is thus not a true measure of baseline performance. Nevertheless, unexpected results (outlined in section 3.3.) made it necessary to include a baseline in order to fully consider the implications of our data.

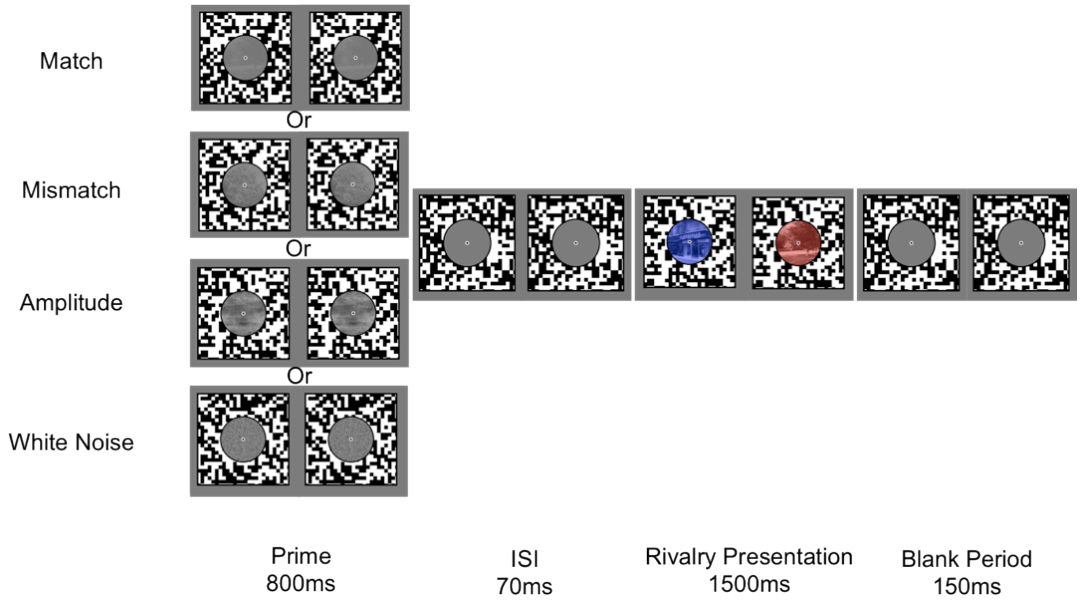


Figure 3.2 Schematic diagram of procedure. Here, primes (all natural) are depicted at 7%, 20% and 50% contrast for match, mismatch and amplitude component primes respectively. Bottom prime stimulus shows white noise (baseline) at 90% contrast.

3.4. Results

The proportion of trials in which participants reported seeing the colour equivalent to the city test image was calculated. A two-factor ANOVA conducted on data for each participant confirmed that none had a significant eye or colour bias (all $p>.05$); data were subsequently pooled across counterbalance conditions. Figure 3.3a to c show the proportion of city scene dominance as a function of RMS contrast for a) match, b) mismatch and c) amplitude conditions, alongside the baseline comparison; the difference in city scene dominance between city prime and forest prime conditions is shown in Figure 3.3d.

The results for the matching condition (Figure 3.3a) demonstrate a pattern consistent with facilitation and suppression; city scenes were more dominant at low contrast when the prime stimulus was a matching city scene, and became much less dominant as contrast increased. City scenes were less dominant when the prime stimulus was a low contrast forest scene, though the effect of low versus high contrast was notably less pronounced for the forest prime than city prime.

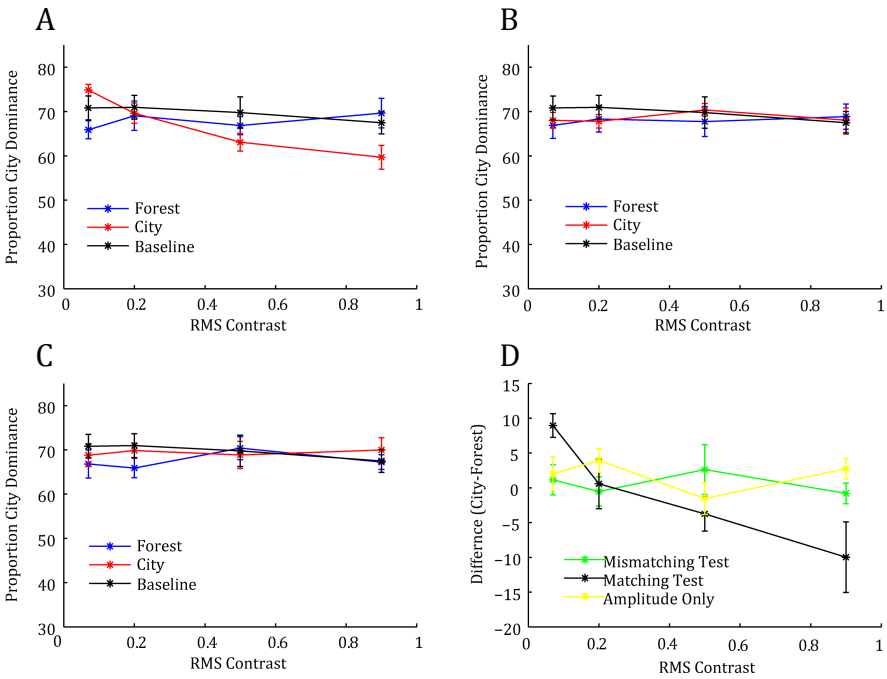


Figure 3.3 A to C: City scene dominance plotted as a function of RMS contrast for forest and city scene primes for each of three conditions (A: Match, B: Mismatch and C: Amplitude Component) and baseline, averaged across all participants (*N*=8). D: Difference in city scene dominance between city prime and forest prime for each condition. Error bars represent ±1 S.E.M

The results for the mismatch, amplitude and baseline conditions (Figure 3.3b, c) do not show a clear pattern of facilitation or suppression, which suggests that neither the contrast, nor content of these prime stimuli had great influence on rivalry dominance. Figure 3.3d summarises the results from the match, mismatch and amplitude component conditions; data points represent the difference in city scene dominance following city versus forest primes. A value of zero difference indicates that the primes did not differentially affect binocular rivalry, thus the only clear trend here is for the match condition, where some points do appear to deviate from zero.

For the baseline condition, a repeated measures ANOVA found no significant effect of contrast on rivalry dominance, $F(3,21) = 1.07$, $p = .38$, $\eta^2 = .13$. After collapsing across contrast, the mean dominance of city scenes was 69.75% and this was significantly different from chance (50%), $t(7) = 7.72$, $p<.01$, $d = 2.73$. This indicates that city scenes were more salient and thus more dominant than forest scenes in this task.

A two-factor repeated measures ANOVA comparing the difference in rivalry

dominance between forest and city primes (city minus forest) revealed a significant main effect of contrast, $F(3,21) = 5.07$, $p < .01$, $\eta^2 = .42$, and a significant interaction between contrast and matching condition, $F(6,42) = 4.27$, $p < .01$, $\eta^2 = .38$, but no significant main effect of matching condition, $F(2,14) = .69$, $p = .52$, $\eta^2 = .09$. To find the source of the condition/contrast interaction, post-hoc t-tests were conducted, comparing city scene dominance at 7% contrast relative to 90% contrast (for each prime category and non-baseline condition); there was a significant difference for city primes in the matching condition, $t(7) = 6.67$, $p = .001$, $d = 2.36$, but not for forest primes, $t(7) = 1.19$, $p = .27$, $d = .42$; t-tests also confirmed that prime contrast had no effect on rivalry dominance in the amplitude and mismatch conditions, $p > 0.3$ (alpha adjusted for multiple comparisons = 0.0083). The results show that prime contrast had little effect on rivalry dominance, with exception of the matching city condition, for which there was a very strong facilitation-suppression effect.

Essentially, this demonstrates that none of the primes caused a bias in rivalry dominance above the effect of the baseline, with the exception of city primes in the match condition, where low contrasts facilitated city responses, and high contrast primes suppressed them.

3.5. Discussion

We examined the effects of matching, mismatch and amplitude-component prime stimuli on subsequent rivalry dominance of intact city and forest scenes. As predicted, the results from the match condition showed a significant interaction between prime category and contrast. Importantly, this successfully demonstrates that facilitation and suppression in binocular rivalry is possible using complex natural scenes, and thus the effect is not limited to simple grating stimuli. Furthermore, the significant interaction means that the match condition can act as a benchmark to compare the results of the mismatch and amplitude conditions. An additional finding was that the contrast of a white noise prime (baseline condition) had no effect on rivalry dominance. Firstly, this implies that any bias in subsequent rivalry in this condition was due to characteristics of the test stimuli themselves, rather than the prime. Secondly, it demonstrates that the white noise prime was no more similar to the city or forest scenes, otherwise an effect of contrast would have been found.

Upon examination of all three main conditions, the only evidence that the contrast of prime stimuli affected subsequent rivalry dominance above baseline level was in the match condition, where prime and test were identical. Upon first glance, this would imply that the dominance of forest or city scenes in binocular rivalry was unaffected by i) matching conceptual category (either implicitly from amplitude information, or explicitly from recognisable scenes or objects), or ii) low-level similarities in amplitude information between prime and test. Nevertheless, the fact that *forest* primes in the matching condition had no effect on subsequent rivalry makes it difficult to draw strong conclusions from the data.

There are several possibilities that may explain why any effect that the matching forest primes had on rivalry dominance was non-significant. Firstly, it could be that the strength of facilitation and suppression are dependent on local phase structure within *prime* images; differences in the salience of contours or recognisability of objects within primes could have led to strong effects for the city- but not forest category. This is unlikely, since experiments 3a and 3b of Chapter 2 showed that amplitude-phase interaction had a strong disambiguating effect on phase-morphed scenes, and this was likely due to the increased salience of characteristic contours, which are clearly evident in both natural and manmade scenes.

A more compelling explanation for the null effect of forest primes is that *test* image dominance was significantly biased even at baseline, thus any influence of the forest prime might have been outweighed by the underlying dominance of city test images. Indeed, facilitation of forest images during rivalry may have been limited due to a floor effect of city scene dominance. If one were to test the hypothesis that effects following matching forest primes were outweighed by the strong baseline dominance of city scenes, it would first be necessary to understand *why* city test images were preferred stimuli; this would allow us to modify the paradigm and attempt to create test stimuli that were more equally matched at baseline.

One explanation for the strong city bias in Experiment 1 could be due to the relative salience of contours within *competing* test images. For instance, the results from Chapter 2, Experiment 3, indicated that the balance between naturalistic and manmade contours was important when distinguishing between morphed scenes; perhaps the same contour information is what competed during rivalry. An alternate,

and more plausible explanation for the baseline results is that the visual system was more sensitive to the spectral slopes of city scenes than forest scenes in this study. Indeed, previous research has shown that spectral slope can significantly affect binocular rivalry (Baker & Graf, 2009), with dominance favouring images with slopes of $1/f$. Thus, it is possible that baseline dominance could vary depending on the basic categories chosen to provide test images.

Importantly, the discrepancy in the results between matching forest and city scenes makes it difficult to discount the possibility that facilitation and suppression effects in binocular rivalry are dependent upon strong contours within the prime stimuli themselves. If this were the case, then one would not predict any effect of amplitude primes on binocular rivalry, as they contain no identifiable structures.

To resolve this ambiguity, two additional conditions were designed to test i) whether intact (unmodified amplitude and phase information) city *and* forest scenes can facilitate or suppress subsequent matching scenes when test image amplitudes have been equalised and ii) whether facilitation and suppression of test images is possible when both prime and test images contain no identifiable structures, but are structurally identical in both amplitude and phase.

3.6. Experiment 2

3.7. Method

3.7.1. Procedure

The basic procedure was almost identical to that in Experiment 1: participants viewed a prime stimulus, followed by an inter-stimulus interval, rivalry presentation and finally a brief blank period (see Figure 3.2 for durations). Prime stimuli were presented at one of four contrast levels, as in Experiment 1. Participants indicated the dominant percept of the rivalling test images (red or blue) by button press. There was no pre-test condition in this experiment, as this information was collected from the same participants in Experiment 1; red values for test images were adjusted for these participants accordingly.

There were two main conditions, the i) phase-match condition and the ii) identical-amplitude condition. The former was designed to be comparable to the match condition from Experiment 1, whilst avoiding any strong underlying

dominance of city versus forest test images. Here, prime images were created from intact forest or city scenes (see section 3.2.3); whereas each test image contained the phase information from one of the prime stimuli, and amplitude information that was a hybrid of the two individual test images (and thus should not affect rivalry). In the baseline measure for this condition, participants were presented with a white noise prime stimulus, followed by a single forest and city test image that retained their phase, but had amplitude information that was a hybrid of the two individual test images, thus we predicted baseline dominance of test images to be close to 50%.

The second main condition (identical-amplitude) was designed to be comparable to the amplitude component condition from Experiment 1, but here prime and respective test image were identical. Prime stimuli and test images consisted phase-randomised versions of individual city and forest scenes (thus prime and one of the two respective test were identical but neither contained objects or shapes).

Participants completed 2400 trials (counterbalanced across eye and colour) and conditions were interleaved.

3.8. Results

Data were pooled across counterbalance conditions as analysis from Experiment 1 confirmed that participants did not have a significant eye or colour bias. The averaged proportion of city scene dominance for each condition is presented in Figure 3.4.

Upon examination of Figure 3.4, it is clear that the bias in baseline dominance is greatly reduced here, relative to that in Experiment 1 (see Figure 3.3). This implies that the manipulation of test image amplitude was successful in reducing the bias in dominance at baseline.

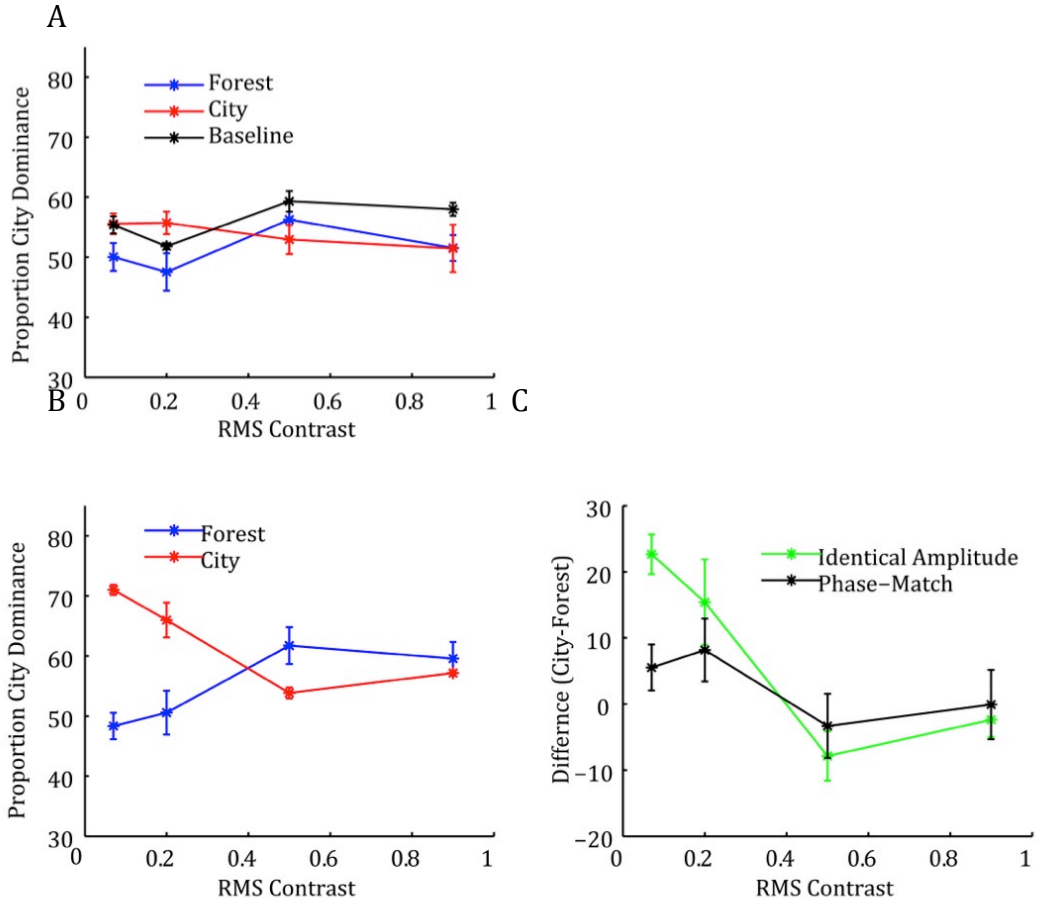


Figure 3.4 The figures show the proportion of city scene dominance as a function of RMS contrast for a) phase-match and baseline, and b) identical-amplitude conditions. Figure 3.4c shows the difference in city scene dominance between city prime and forest prime for each condition. Error bars represent ±1 S.E.M, N = 3.

In the phase-match condition, any visible effects of primes on subsequent dominance are minimal. In contrast, primes in the identical-amplitude condition do seem to show facilitative/suppressive effects on subsequent binocular rivalry. Figure 3.4c summarises the results from the phase-match and identical-amplitude conditions; each data point represents the difference in manmade dominance following city versus forest primes. The data suggest a facilitation to suppression shift in the condition where the amplitude of the test image was identical to the prime. In contrast, difference values for the phase-match condition do not appear to deviate from zero. These results indicate that prime stimuli do not affect binocular rivalry if prime and test do not contain identical amplitude information.

A repeated measures ANOVA found no significant effect of prime contrast on rivalry dominance in the baseline condition, $F(3,6) = 2.45$, $p = .16$, $\eta^2 = .55$. After collapsing over contrast, the mean dominance of manmade images was 57.63% and this was significantly different from chance (50%), $t(2) = 7.00$, $p=.02$, $d = 5.69$. This indicates that city images were still more salient than forest images, even though both rivalry images contained the same amplitude.

A two factor repeated measures ANOVA comparing the difference in rivalry dominance between forest and city primes (city minus forest) revealed a significant main effect of contrast, $F(3,6) = 7.62$, $p =.018$, $\eta^2= .79$, but no significant main effect of condition, $F(1,2) = 2.99$, $p=.23$, $\eta^2= .60$ and no significant interaction $F(3,6) = .77$, $p= .55$, $\eta^2= .28$. We would expect the largest difference to exist between the two extreme ends of our contrast range. Thus, post-hoc t-tests were conducted comparing manmade dominance at 7% contrast relative to 90% contrast (for each prime category and non-baseline condition); in the *identical-amplitude* condition there was a significant difference for city primes, $t(2) = 22.89$, $p=.002$, $d = 13.03$, but not for forest primes, $t(2) = 2.43$, $p = .14$, $d = 2.58$. Differences between the highest and lowest contrast primes were non-significant in the *phase-match* condition for city: $t(2) = 2.21$, $p = .16$, $d = .74$, or forest primes: $t(2) = .17$, $p = .23$, $d = .39$ (adjusted alpha = 0.0125). In contrast to Experiment 1, where prime contrast only had a significant effect for matching city primes, here the results show that only amplitude component primes from city scenes had a significant effect on rivalry dominance, when city test images were identical to the prime.

3.9. Discussion

The additional conditions run in this experiment were designed to expand on the results from Experiment 1, by 1) using test images with equalised amplitude, so as to avoid any strong underlying bias due to test-image amplitude (phase-match, baseline conditions) and 2) examining whether primes can influence rivalry when they contain no identifiable structure (identical-amplitude condition).

In the baseline measure of this study, the prime was a white noise stimulus, while the test images contained their original phase but their amplitude information was equalised. This manipulation was successful, as the baseline dominance dropped to 58%, which is greatly reduced relative to the 70% city test dominance found

following the baseline in Experiment 1. Though this result would suggest that the bias following the original baseline prime was predominantly due to test image amplitude, city scenes were significantly more dominant than forest scenes even after amplitude equalisation, which implies that either the phase, or amplitude-phase interaction within test images played some role in baseline rivalry dominance. Importantly, the decrease in the baseline dominance should have prevented any influence of a city scene floor effect, thus making it easier to find subtle effects of forest as well as city primes.

By blending the test image amplitudes in the phase-match condition in Experiment 2, I was able to examine potential facilitation and suppression effects of primes while also avoiding the strong city-scene dominance that was evident in Experiment 1. The resultant city prime effects were visibly reduced relative to Experiment 1, and the interaction between prime contrast and prime had non-significant effects on subsequent rivalry. This demonstrates that the strong city scene effects from Experiment 1 were likely due to the underlying city test bias, rather than the relative strength of contours between the city and forest primes.

Indeed, significant results from the identical-amplitude condition further confirm that salient contours are not necessary in order for primes to affect subsequent rivalry dominance. Crucially, this confirms that the amplitude condition from Experiment 1 was a valid measure to test whether the visual system associates unlocalised amplitude with its intact counterpart image.

3.8. General Discussion

The aim of this chapter was to answer two questions: 1) Does the visual system recognise that an amplitude component image (with randomised phase information) contains the same information as its phase-intact counterpart? 2) Does the visual system associate characteristic amplitude spectra with specific scene categories (e.g. forest or city)? A method utilising priming and binocular rivalry was chosen to address these questions, as previous research has shown that both facilitation and suppression of rivalry stimuli can be achieved by manipulating the contrast of prime stimuli (Brascamp et al., 2007; Pearson, Clifford & Tong, 2008).

3.8.1. Overview of findings

The results from the matching condition in Experiment 1 and the identical-amplitude condition from Experiment 2 confirmed that exposure to complex natural or naturalistic stimuli could bias subsequent binocular rivalry when prime and test were structurally identical. Furthermore, the identical-amplitude condition demonstrated that these effects were not reliant upon identifiable structures within prime or test image, thus confirming that inclusion of an amplitude-component prime condition was a valid test to examine question 1.

The premise behind question 1 was that the visual system might have an internalised, low-level representation of natural scene categories based on the neuronal activity elicited by image amplitude spectra. If so, then previous findings of rapid scene categorisation (e.g. Rousselet et al., 2005; Greene & Oliva, 2009) could be, in part, due to implicit conceptual categorisation based on image amplitude spectra. Indeed, the results from Chapter 2 (experiment 3a) implied that the amplitude of phase-morphed images might be informative even when averaged across multiple images within the same scene category (thus amplitude was non-specific). Nevertheless, due to the design of the study, the unique influence of amplitude was unclear in Chapter 2 due to amplitude-phase interactions, which may have increased the salience of diagnostic contours. The design of the current study allowed for potential facilitation of amplitude in the absence of such interaction.

The amplitude component condition was designed to test the hypothesis that the visual system might associate image amplitude spectra with specific scene categories (i.e. an amplitude template). If this were the case, then we would expect a pattern of facilitation/suppression of subsequent test images that contained the same amplitude as the prime, but with phase information intact, since the same low-level information is repeated from prime to test. Nevertheless, the results from the amplitude-component condition showed no significant interaction between prime contrast and prime category on binocular rivalry, thus demonstrating that the visual system *did not* associate the unlocalised amplitude prime with its phase-intact counterpart. The identical-amplitude condition, however, did demonstrate a pattern of facilitation to suppression as prime contrast increased. Crucially, this shows that the visual system is able to process and recognise the amplitude component primes when they are repeated, yet local structure must be consistent between prime and

test in order to affect binocular rivalry. What this implies is that the visual system does not use 'characteristic' low-level amplitude spectra to identify scene category, rather, the higher order relationship between localised image structure, primarily quantified as Fourier phase information is key to recognising the identity of a scene.

The negative conclusion to question 1 pre-empts the answer to question 2, since the visual system would not use amplitude spectra to identify scene category if it cannot recognise an intact image from its own amplitude spectrum. Nevertheless, the mismatch and phase-match conditions also investigated the role of i) localised phase and ii) conceptual categorisation in the facilitation/suppression paradigm. In the phase-match condition, test images retained their phase information, but their amplitude was a blend between a single forest and city image. The contrast of neither city, nor forest primes in the phase-match, nor mismatch condition had any significant effect on subsequent binocular rivalry. This implies that there must be a strong local correspondence between prime and test in order to observe facilitation and suppression in this paradigm. In sum, we find no evidence of repeated shapes, recognisable objects, or conceptual scene category (explicit or implicit) on rivalry dominance subsequent to structurally non-identical primes. This is, perhaps, not surprising as both Pearson et al. (2008) and Brascamp et al. (2007) reported that using simple grating stimuli, deviations in phase and orientation tasks significantly disrupt the typical pattern of results. It is, nonetheless, surprising that the phase-match condition led to non-significant effects of rivalry, as although the local contrast (amplitude) within test images was disrupted, local contours defined by phase remained intact. Perhaps a greater number of participants would be required to detect any subtle effects in the phase-match condition.

In sum, from the results discussed here, we can conclude that the visual system does not recognise that an amplitude component image contains the same information as its intact counterpart, and thus it is unlikely that unlocalised amplitude plays a direct role in rapid scene categorisation.

3.8.2. The indirect role of unlocalised amplitude in scene perception

The results reported here are inconsistent with those found in previous priming (Guyader et al., 2004) and adaptation (Kaping et al., 2007) studies, whose results have been interpreted as proving evidence that unlocalised amplitude might

be sufficient to allow rapid scene categorisation. Instead, they add to a wealth of evidence suggesting that, whether characteristic or not, the distribution of orientation and spatial frequencies described by the amplitude spectrum is not diagnostic of scene category (e.g. Joubert et al., 2009; Loschky et al., 2007). Nevertheless, masking studies do show that amplitude information has at least some role in rapid scene categorisation. For instance, Loschky et al. (2007) used a backward masking paradigm to show that amplitude masks were equally effective whether from the same, or different category to the test image. This implies that the visual system was insensitive to any 'characteristic' amplitude information within masks. Indeed, even masks containing uninformative white noise led to a decrease in categorisation accuracy, though significantly less so than scene amplitude masks. Together, these results show that masking of any sort can make image features less salient; shared features between mask and test should experience greater masking.

It is however interesting to note that white noise masks with flat amplitude spectra are less effective than natural amplitude masks, which have spectral slopes of roughly 1/f. Loschky et al. (2007) argue that the low spatial frequencies that dominate natural scenes are more important than higher frequencies for rapid scene categorisation (supported by research such as Schyns & Oliva, 1994), thus explaining why natural masks are more effective. Again, this does not imply that *unlocalised* low spatial frequency information is diagnostic, as this would overlook the influence of phase. Indeed, Loschky et al (2007; 2010) showed phase *is* important, as intact scenes proved to be the most effective masks, and masking becomes less effective as phase is increasingly randomised (2007), implying that the loss of contours or structure has a strong effect on scene categorisation.

A parallel can be drawn here with experiments 1, 2 and 4 from Chapter 2, in which prior exposure to amplitude negatively biased scene categorisation, while intact amplitudes (Chapter 2, Experiment 3) caused a positive bias. Both effects imply an interaction between amplitude and phase, making contours more or less salient, thus showing that amplitude can affect scene categorisation indirectly via phase. Nevertheless, what Chapter 2, and masking experiments failed to do was show unequivocally that amplitude can or cannot affect categorisation in the absence of phase interactions. The current study was able to test this, and confirmed that amplitude does not facilitate binocular rivalry unless the test images are identical;

therefore, it must be information held within the phase spectrum, or an amplitude-phase interaction that allows rapid scene categorisation.

One final investigation into the properties of natural scenes that allow rapid categorisation is to question what it is about phase that is diagnostic, such as object recognition or global layout and shapes. Loschky and colleagues (2010) probed this question when asking whether it is the recognisability of intact scenes that makes them more effective as masks. In their study, images were manipulated to obtain 'scene textures' that retained phase structure but were not recognisable. The researchers found that unrecognisable scene textures caused greater masking than phase randomised versions of the same image (with identical amplitude). This demonstrates that i) useful information is retained within scene structure, but not amplitude spectra, and ii) that effects were not conceptual, as masks were unrecognisable. Furthermore, intact scenes were found to be more effective masks than scene textures, which suggests that global information lost in image manipulation also has some influence on scene categorisation.

In conclusion, though unlocalised amplitude is insufficient for rapid scene categorisation, it is necessary as a component of phase, which can be diagnostic of scene category via contours, shapes or global configuration.

3.8.3. Why is the visual system sensitive to amplitude statistics?

Despite the evidence suggested that phase, rather than amplitude is important for scene categorisation, overwhelming evidence still exists that shows the visual system is sensitive to natural, $1/f$ amplitude spectra. Not only does the human contrast sensitivity function approximate the distribution of spatial frequencies in natural scenes (e.g. Field, 1987; Banks & Salapatek, 1978; Campbell & Green, 1965; Campbell & Robson, 1967), but evidence also demonstrates a preference for natural stimuli. For instance, stimuli with $1/f$ amplitude dominate in binocular rivalry (e.g. Baker & Graf, 2009) and $1/f$ surrounds maximally suppress central images relative to surrounds with other spectral slopes (McDonald & Tadmor, 2006). This sensitivity illustrates how the visual system is adapted to the statistics of natural scenes and this can be observed at multiple levels of processing, from the retina to complex cells in the visual cortex. For instance, an efficient system would have to adapt to the high level of redundancy within typical visual input, which is due to strong correlations

between neighbouring patches within scenes (eg. Potetz & Lee, 2003). Indeed, Atick and Redlich (1992) showed that output from retinal ganglion cells was whitened (had a flat amplitude) when exposed to natural scenes, but not when exposed to white noise. This whitening effect means the output was decorrelated and thus demonstrates that retinal ganglion cells are well adapted to reduce redundancy in natural scenes. In addition, Olshausen and Field (1996) showed that natural scenes could be sparsely represented by a limited number of receptive fields that are bandpass, orientation-selective and spatially localised. They argue that since receptive fields of simple cells in the human visual cortex share the same properties then this is another example of the visual system's adaptation to natural statistics.

At a higher level of processing, Felsen, Touryan, Han and Dan (2005) showed that response amplitudes of complex cells were enhanced when shown their preferred feature in natural- but not randomised scenes, yet this effect was only apparent when the phase structure of the natural scenes was intact. This implies that the complex cells were dependent on natural phase, but not natural amplitude. Essentially, this confirms that the visual system is well adapted to process natural scenes at multiple levels, even to the extent of increasing the salience of complex features. In addition, it implies that focus on scene categorisation research should be more focused on studying phase and image features, rather than low-level amplitude.

3.8.4. Conclusion

In the collection of studies explored here and in Chapter 2, the main focus has been on the potentially diagnostic role of amplitude spectra in rapid scene categorisation. In Chapter 2 participants were asked to rapidly categorise hybrid images that were part natural and part manmade in amplitude and phase. Compared to the current study, the experiment in Chapter 2 addressed a higher level of processing, since observers were faced with object rivalry (which scene is most dominant?). In comparison, the current experiments were less harsh a test of whether the visual system associates an unlocalised amplitude image with an intact counterpart, since the task was one of eye rivalry, and the task examined whether dominance would be facilitated in the eye that corresponded with the prime, rather than necessarily requiring any higher levels of contextual processing (though this was also tested in the mismatch condition).

Based on the theory that vision is informed by the statistics of the natural environment (e.g. Brunswick and Kamiya, 1953; Geisler, 2007), it would be logical to assume that common scene properties might be indicative of scene category. The fact that studies of natural images have shown characteristic differences in the distribution of spatial frequencies and orientations between scene categories (e.g. Oliva & Torralba, 2001; Torralba & Oliva, 2003) does suggest that amplitude alone might be informative, especially since object recognition is not necessary for rapid categorisation (Schyns & Oliva, 1994; Larson & Loschky, 2009). Nevertheless, the conclusion to be drawn from the literature discussed in this chapter, alongside the results from Chapters 2 and the lack of amplitude facilitation in the current study, is that the visual system does not utilise 'characteristic' information, only that categorisation is disrupted when amplitude is suppressed, and that this is likely due to an interaction with phase.

This can be explained more intuitively by using the analogy of a greyscale jigsaw puzzle: if a natural scene is cut up into infinitely small pieces and then rearranged, the identity of the scene will be lost, though all of the information remains present. Indeed, this demonstrates how 'the whole is greater than the sum of the parts', since pieces (and local amplitude and phase) must be aligned in order to inform about the content within the scene; low-level amplitude alone does not make an identifiable scene.

Further research could examine exactly which features make a scene, though this is likely due to involve complicated semantic rules, rather than a 'quick and dirty' summary that was once predicted based on amplitude spectra.

Chapter 4

Inferring 3D surface shape from 2D contour curvature

4.1. Introduction

When fixating on a well-lit object with both eyes, there are many visual cues as to the 3D shape of that object, whether from colour, specular highlights, surface texture, or the disparity between two retinal images, to name but a few. Nevertheless, studies have shown that a mere outline or silhouette can be sufficient to recognise a familiar object (e.g. Wagemans et al., 2008; see Figure 4.1), despite all other cues being stripped away. More intriguing, is that the silhouette of even unfamiliar objects can often lead to a compelling sense of 3D depth, as demonstrated in Figure 4.2 (see Norman & Raines, 2002; Tse, 2002). Since the 2D retinal image cannot provide metric depth information, then 3D shape from silhouettes must be inferred from internalised knowledge about common object properties. The goal of this chapter is therefore to examine properties of silhouettes that might affect 3D shape perception.

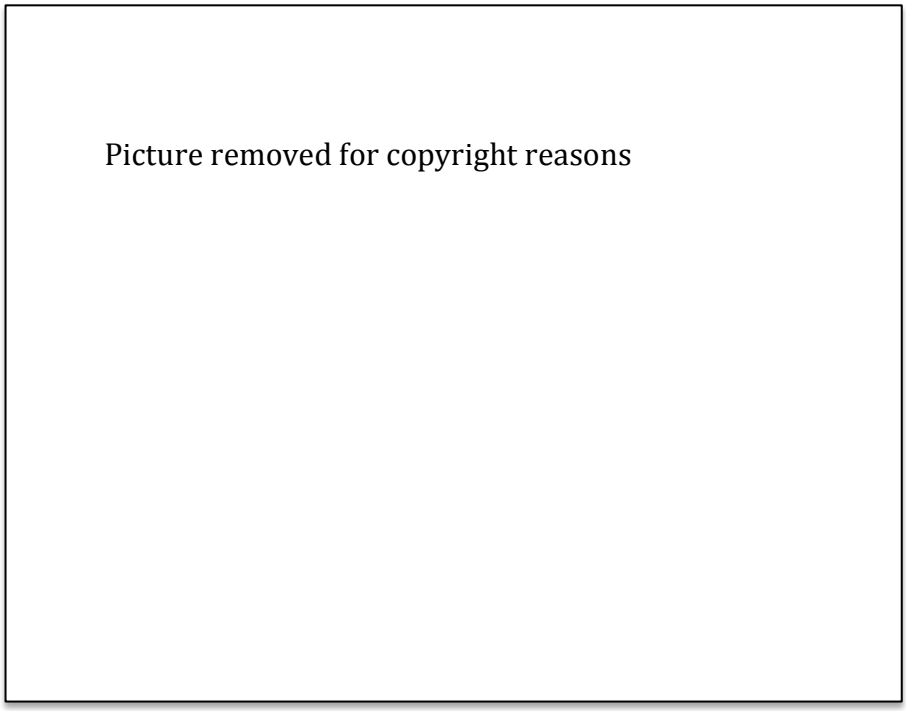


Figure 4.1 Familiar objects can be recognised from their silhouette (left) or outline (right). Here, the bird (top) and bunch of grapes (bottom) are easily identified when shown at a non-accidental viewpoint. Adapted from Wagemans et al., 2008.

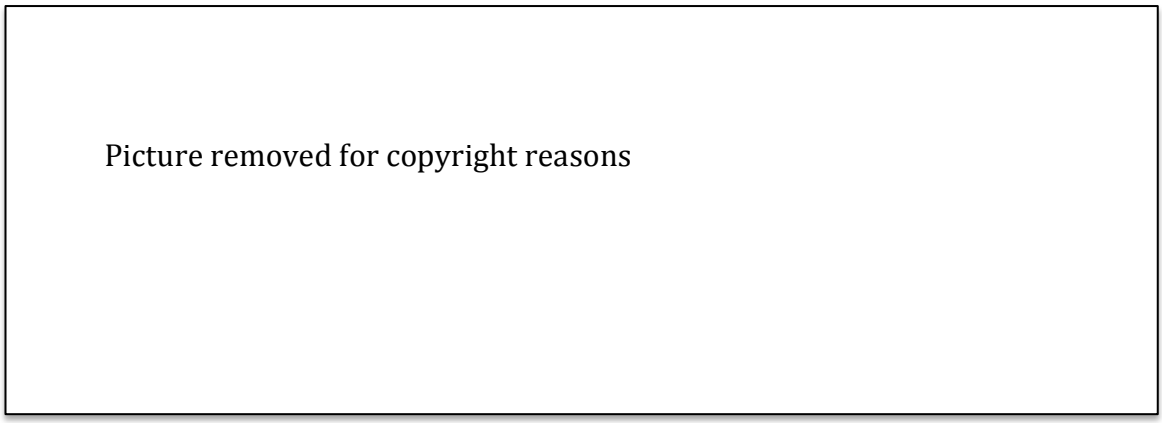


Figure 4.2 Points A and B above can easily be ordered in depth, since the object is familiar and its 3D shape is known. Likewise, D appears closer to the observer than C within the recognisable apple silhouette. The rightmost figure is unfamiliar, yet it is still possible to infer 3D shape and order the probe points, perhaps due to internalised knowledge of depth regularities within real objects (rightmost figure adapted from Norman & Raines, 2002).

For the purpose of this investigation, the bounding contour can be defined as the 2D retinal projection of a 3D surface, or its silhouette, while the rim is defined as the visible part of the 3D surface (the remainder of the surface is self-occluded; Koenderink, vanDoorn, Kappers & Todd, 1997). For smooth, solid objects, surface slant generally increases towards the rim, asymptoting at 90 degrees. Thus, locally, distance from the rim can be a powerful depth cue. Indeed, Figure 4.3 demonstrates how surface slant increases, and depth recedes to infinity at the bounding contour. Norman & Raines (2002) tested whether this knowledge was internalised by asking participants to judge which of two points within a silhouette was closer to them in depth (as demonstrated in Figure 4.2). Participants judged probe points on rotating stimuli that were randomly shaped, smooth objects and were presented stereoscopically (but had no shading nor texture). After fitting psychometric functions to the data and calculating discrimination thresholds, the results showed that not only could participants judge ordinal metric depth, but that they were better at doing so closer to the bounding contour. Importantly, a second experiment confirmed that, though thresholds increased, proximity to the edge still affected discrimination thresholds when stimuli were static and monocular, thus demonstrating that proximity to the bounding contour alone can be a salient cue to

3D shape. It is worth noting that in addition to the proximity from the edge, the separation between points was an important factor in ordinal depth judgements, with discrimination thresholds being smaller for closer points. This correlates with observations related to natural scenes, as the analysis of co-registered range-intensity images shows that shapes tend to be smooth and that range is highly correlated between neighbouring pixels (Potetz & Lee, 2003), but as the separation between points increases, so might variations in depth within the shape.

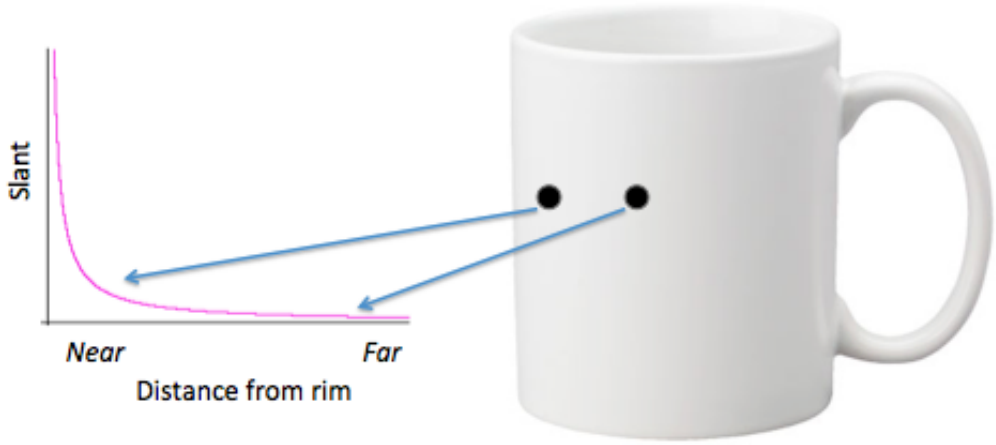


Figure 4.3 Demonstration of how slant, and thus depth, increases towards the rim of an object.

Though it is clear that proximity to the bounding contour can affect perceived depth, the focus of the current experiment is to examine whether the *sign* (convex/concave) and *magnitude* of the bounding contour can be a cue to metric depth, interior to the contour. Indeed, the importance of curvature along the bounding contour is not a new concept, as Attneave (1954) demonstrated that the maxima and minima of contours on a shape's outline are most salient, as demonstrated in Figure 4.4. Attneave asked participants to mark 10 points along bounding contours that best represented the object's shape. An example is given in Figure 4.4 (left). Concluding that curvature maxima and minima (extremes of convexities and concavities) were most informative, Attneave used this rule to plot an outline of his cat, which is clearly identifiable after joining the dots (Figure 4.4, right). This was further investigated and empirically tested nearly 50 years later, leading to the same conclusion that high magnitude convexities and concavities are

exceptionally informative about object shape (Norman, Philips & Ross, 2001; DeWinter & Wagemans, 2008). Furthermore, Feldman and Singh (2005) argue that concave curvatures on a silhouette should be more informative as they are less common; the sum of curvatures must be positive to make a closed contour. Indeed, there is much research investigating the different roles or effects that convexities and concavities have in areas such as figure-ground organisation and attention, though surprisingly little focus on bounding contour shape in relation to perceived depth (see Bertamini & Wagemans, 2013, for review).

Figure 4.4 Left: Participants marked salient positions along the bounding contour; the lengths of radial bars indicate the cumulative frequency of points chosen for that location. Right: Attneave's cat, created by drawing lines between the dots that mark curvature maxima and minima (adapted from Attneave, 1954).

Before investigating the effects of bounding contour curvature on human depth perception, it makes sense to first consider both mathematical and theoretical constraints that explain how 3D shapes project to 2D retinal images (whether such constraints are internalised or not). For instance, Koenderink (1984) showed mathematically that convex points on the boundary project from convex surface points, whereas concave points project from surface saddle points. A saddle point is locally concave, but all 3D objects must have surface completion, thus concave points are bounded by points, and shapes are globally convex. Richards, Koenderink & Hoffman (1987) add to this, stating that we must not assume an accidental viewpoint, and thus we should not infer bumps or dips unless indicated by the 2D shape (also see Knill, 1992 for further reading). Some further understanding can be gained by

exploring the relationship between depth statistics, as calculated from the natural environment, and co-registered photographic images that contain the potentially informative 2D contours (e.g. Potetz & Lee, 2003; Yang & Purves, 2003; see below). Indeed, Burge, Fowlkes and Banks (2010) used such information to demonstrate that depths between convex curves (2D) and the surfaces they occlude are significantly greater than depths between concave curves and their occluded surfaces. Though such findings are more enlightening about the relationship between convexity and figure/ground organisation, they highlight the fact that there is huge scope to utilise natural scene databases to better understand the statistics of our environment and how these relate to human perceptual biases.

Indeed, though researchers are aware that the shape of the bounding contour could be a cue to metric depth interior to the boundary (see Bertamini & Wagemans, 2013), little or no research exists to test this psychophysically, nor confirm that such a relationship does exist (between curvature and metric depth) in the natural environment.

In the current study we take a psychophysical approach to examine the theory that the sign and magnitude of contour curvature may be a cue to metric depth. Participants will view randomly shaped, smooth 'potatoes' that are defined in depth by binocular disparity (described below) and thus should appear to extend in depth in front and behind the surface of the screen. Half of the potato stimulus will be removed and replaced by a convex, concave or flat contour, with uniform depth set at the same distance as the screen. Observers will be asked to adjust the depth of a stereoscopic probe dot so that it sits just on the surface of the contour side of the stimulus. Since the contour itself is rendered to lie flat in the screen plane, any systematic deviations in depth probe settings from zero will be due to i) sign or magnitude of the curvature of the bounding contour, ii) proximity of the probe from the bounding contour, or iii) inference based on the depth at a corresponding point on the potato half of the stimulus.

Binocular disparity is usually a strong cue to depth (e.g. Adams & Mamassian, 2004; Bülthoff & Mallot, 1998) as the disparity between the left and right eye's images indicates the positions of objects relative to one another in the environment (see Figure 4.5 for example). It is possible to create the binocular disparity information experienced in normal viewing by presenting disparate images to the eyes using a

mirror stereoscope and carefully controlling for the focal distance of the observer. The use of binocular disparity allows participants to make depth judgements that relate to metric depth, rather than relying on ordinal judgements as previous studies have done (e.g. Todd & Reichel, 1989; Norman & Raines, 2002).

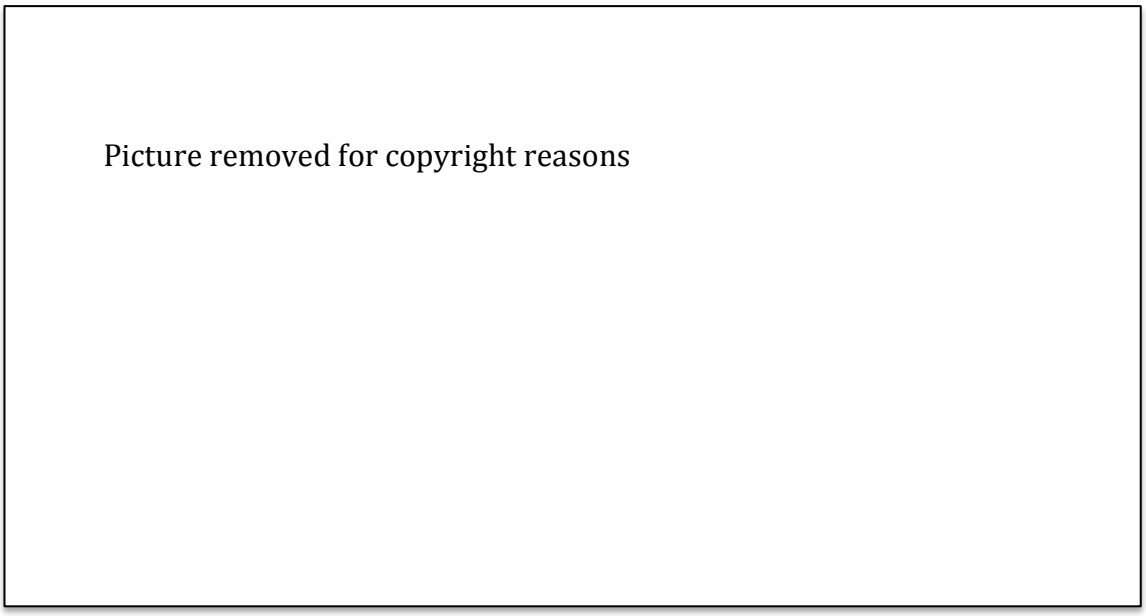


Figure 4.5 The two images above can be cross-fused to gain a sense of 3D depth. Note how the objects are positioned slightly differently in the left versus right eye's image. Adapted from Held, Cooper & Banks, 2012.

4.2. Method

4.2.1 Participants

Seven observers (2 male) took part after exclusions (mean age 30.14 years, s.d. 4.02); all had normal or corrected vision. Student participants received a small payment for their time. The local ethics committee granted approval and participants gave written consent.

4.2.2.Apparatus

Stimuli were created in MATLAB (Mathworks) and presented on a 17" CRT ViewSonic Graphic Series G90FB screen using an AppleMac Pro (Apple Computing). The screen resolution was 1152*870 with a 75Hz refresh rate. A chin rest was used to

maintain fixed head position 85cm from the screen and stimuli were viewed through a mirror stereoscope, in a darkened room.

4.2.3. Stimuli

A set of natural looking, 3D 'potato' stimuli were generated in OpenGL by applying 3D simplex noise to the vertex positions of a sphere (see Perlin, 2002). Potatoes were excluded from the stimulus set if they contained self-occlusions (depth steps), as the assumption was made that surfaces are smooth for computational simplitcity. All potatoes in the remaining set were resized to have the same width (X), height (Y) and depth (Z) at their cross section and were aligned to the centre of the screen. The potato was divided in half vertically, with only one half being rendered, while the other half was replaced by a partly occluded smooth contour.

Contours were generated by rendering circles of different radii (see Norman et al. 2006; Figure 4.6) and then positioned so that the edge of the bounding contour fell at point -X, thus locally mirroring, and 'completing' the potato. Areas of the circle that protruded past the occluder were then cropped.

The distribution of contour curvature magnitudes was selected to be similar to that calculated from a set of naturalistic potato stimuli from a previous pilot study (mean radius ~ .35). In this study, circles with radii of ± .25, .35 and .6 were chosen (where smaller radii make steeper curvatures), in addition to a contour with zero curvature (a square). Concave contours were generated by cropping a circle of a given radius (.25 .35 or .6) from a larger circle (See Figure 4.7) The depth of the contour was set to be equal to the depth of the screen when viewed stereoscopically.

On each trial, a pink occluder was stereoscopically rendered, so as to be perceived to be well in front of the depth of the potato, at a depth of Z plus two standard deviations of the original set of potato radii. The occluder was set to replace the cropped half of the potato stimulus, and was positioned laterally so as to prevent a gap between occluder and potato when viewed stereoscopically; its width and height were set to be greater in width and height than the cropped half of the potato. Finally, a 45° transparent wedge was rendered at the centre of the occluder; this was the viewing window for the rendered contour (see Figure 4.8).

Stimulus orientation was randomised on each trial, such that the potato half of the stimulus could be presented on the right or the left of the contour half. The

potatoes were lit by a single virtual light source, positioned at zero elevation on the side of the stimulus where the potato was visible; this meant that lightness was greatest at the bounding contour of the potato, decreasing towards black near its centre. Shading provided a cue to depth on the potato stimuli, while the contours themselves were unlit and remained uniformly black.

Picture removed for copyright reasons

Figure 4.6 The magnitude of a smooth contour curvature is calculated by fitting a circle with radius (r) that closely fits the curve (C). In the current study segments of perfect circles were used as contours. Figure adapted from Norman et al., 2006.

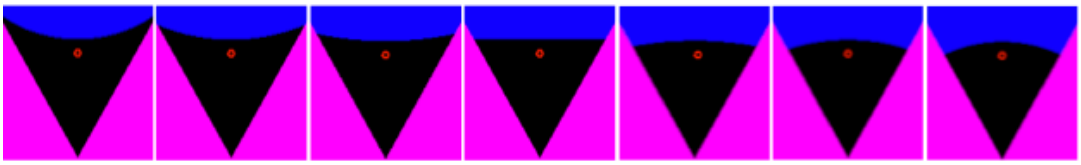


Figure 4.7 Demonstration of curvature magnitudes (black) used in the study. Left: maximum concavity, centre: flat contour, right: maximum convexity. Note that stimuli have been rotated for viewing purposes and would normally be presented with the contour to the left or right.

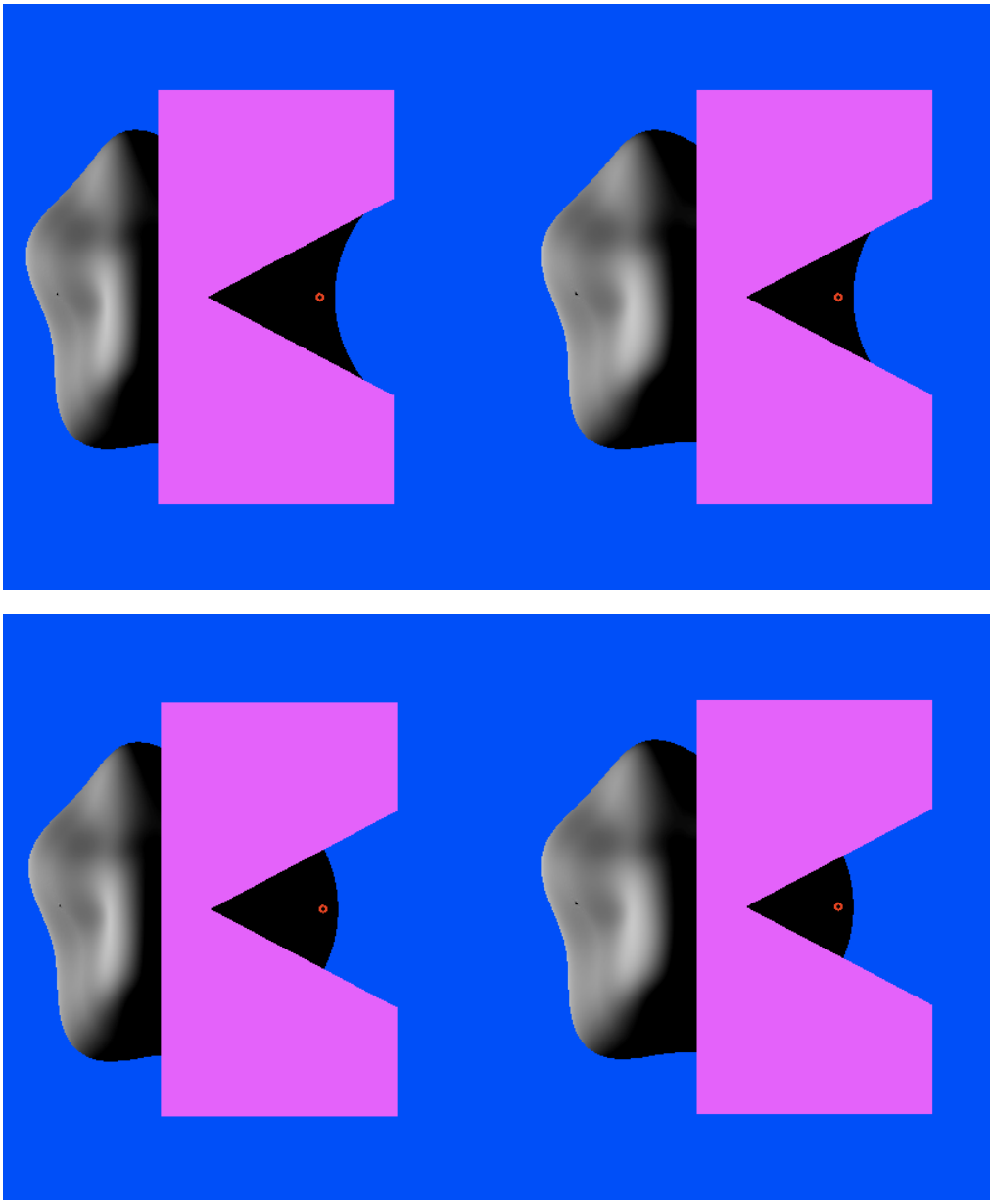


Figure 4.8 The top (concave contour) and bottom (convex contour) are examples of a single trial from the experiment; left and right images can be cross-fused for a 3D percept. Participants were asked to move the red dot (shown) until it appeared to rest on the surface of the potato.

4.2.4. Procedure

Participants viewed smooth, partially occluded stereoscopic 'potatoes' and adjusted the depth of a red stereoscopic probe dot to lie on the potato's surface. In each trial, the probe dot was placed at 80, 85 or 90% of the distance along the radius to the bounding contour, in the X direction. Participants were able to move the probe

closer or further in depth between ±Z. At the beginning of a trial, the probe started at a randomised starting depth between ± 2/3 * Z, and, with a button press, could be moved between ± Z at steps of either 1.25% or 2.5% of the total moveable distance. A beep indicated that the participant had reached the maximum or minimum possible depth to set the probe. There were no time constraints; participants pressed space to move on to the next trial. Twenty potatoes were shown with the contour on the left and right, making a total of 840 trials (20* potato, 2 * side, 7 * curvature, 3 * probe position). Trials were randomly ordered and completed in 4 roughly half hour sessions.

4.3 Results

An initial regression analysis was used to exclude participants who were not able to reliably use the stereo information to set the probe dot. The regression determined the degree to which the position of the probe point predicted set depth for each participant. Since proximity of a probe to the edge of the bounding contour has previously been shown to significantly affect participants' depth settings (Norman et al., 2002) then a lack of effect would suggest that participants' stereo settings were not sufficiently reliable to complete the task. Two participants were subsequently excluded from further analysis, as probe position did not predict set depth (P1: p = .48, CI slope included zero; P2: p = .48, CI slope included zero), leaving seven participants (P3:P9, p<0.001 CI slope did not include zero).

Data were pooled across curvature presentation side (left/right), resulting in seven curvature levels, each with three levels of probe position. Responses were also averaged across the repetitions for 20 potatoes, making 40 trials per data point in Figure 4.9. Though it is unclear whether there is any effect of contour magnitude on set depth. Figure 4.10 shows the same data as Figure 4.9; here data are organised to show the effect of contour *sign* on set depth, rather than magnitude. There is a slight trend for depth to be set from flattest to highest in depth, in the order concave<flat<convex curvature. Data were then pooled across probe position for statistical analysis, as a lack of probe proximity effect was used to exclude participants, and would thus not be informative. One-sample t-tests confirmed that depth was set significantly greater than zero (screen depth) for each level of contour magnitude (p<0.05), thus confirming that participants perceived the stimuli in depth.

We are interested in the effect of bounding contour curvature magnitude and sign on depth settings near the contour. A repeated measures ANOVA investigating the effect of curvature *magnitude* on set depth revealed a significant effect, $F(6,36) = 2.73$, $p = 0.027$, $\eta^2 = .31$. Paired sample t-tests ($t(6)$) comparing set depth between neighbouring contour magnitudes (e.g. flat contour relative to slightly convex or concave contours) revealed no significant differences, *ps*> 0.05.

A repeated measures ANOVA investigating the effect of curvature *sign* on set depth revealed a non-significant effect, $F(2,12) = 3.44$, $p = 0.066$, $\eta^2 = .37$. Paired sample t-tests comparing set depth for different curvature signs were non-significant, (concave-flat: $t(6) = 1.42$, $p = .21$, flat-convex, $t(6) = 1.82$, $p = .118$, convex-concave, $t(6) = 1.99$, $p = .094$).

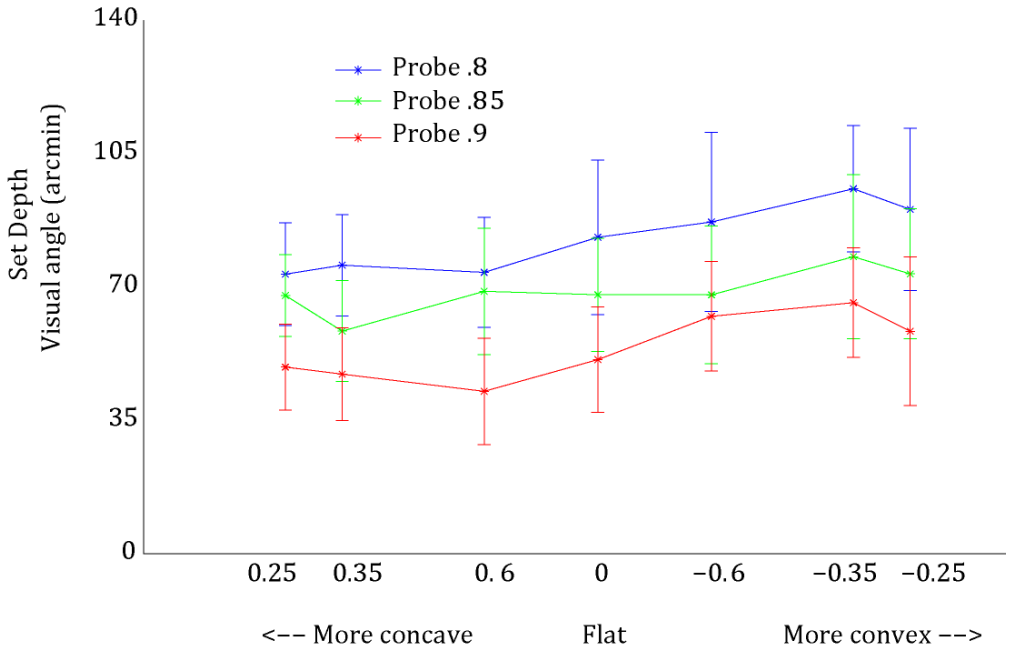


Figure 4.9. Plot showing set depth of probe points, for each of 3 probe distance conditions and seven curvature magnitudes. Data was averaged across participants (*N*=7); bars represent ±1 SEM.

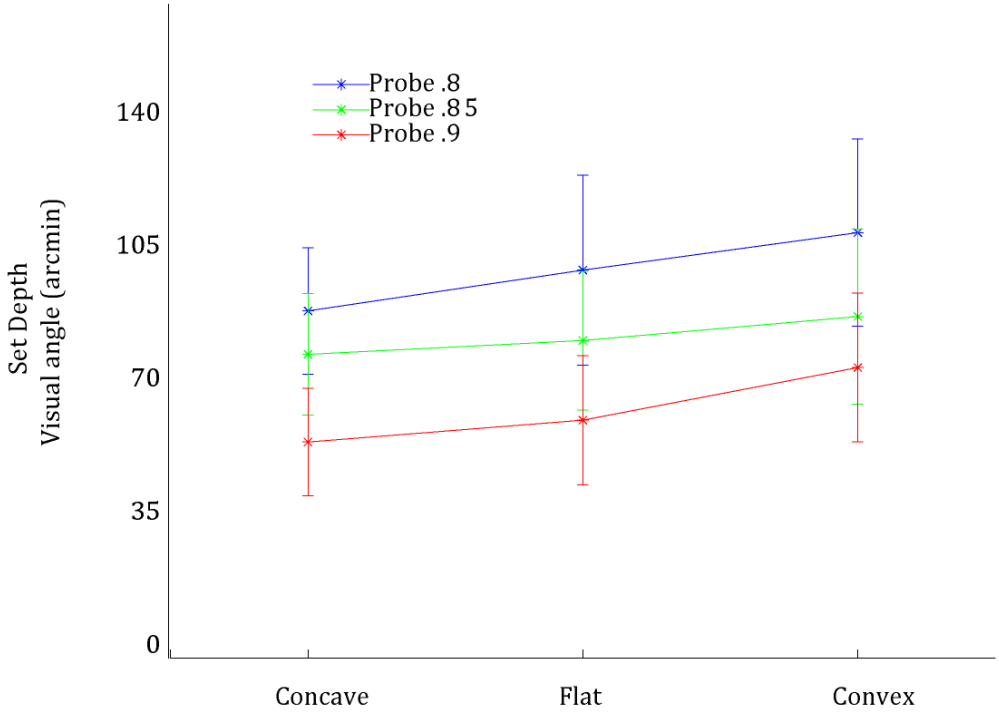


Figure 4.10. Plot showing set depth of probe points, for each of 3 probe distance conditions and 3 curvature signs. Data was averaged across participants (*N*=7), bars represent ± 1 SEM

4.4 Discussion

Though it has long been known that contour curvature provides useful information about object shape (e.g. Attneave, 1954; Koenderink, 1984), and it has even been suggested that the shape of a bounding contour might be a cue to depth within that shape (Bertamini & Wagemans, 2013), ours is the first study to test this hypothesis empirically. In this experiment we examined whether the sign and magnitude of a 2D contour curvature can be a cue to 3D shape interior to the contour. Participants were tasked with setting the depth of a stereoscopic probe dot until it appeared to sit on the surface of a shape that was bounded by a convex, concave or flat curve. Though the contour shape itself was flat in depth, it was joined with a natural looking, 3D potato, with the intention that the two would be amodally completed behind an occluder and that the flat contour would then evoke a convincing sense of 3D shape. Indeed, the fact that participants set the probe at depths significantly greater than zero (screen depth) confirms that depth was reliably

perceived. There were several potential cues that could affect perceived depth on the contour side of the stimulus: i) inference based on the disparity-defined depth at the equivalent location on the potato side of the stimulus, ii) proximity from the edge of the bounding contour, iii) sign of contour curvature (convex, concave or flat), iv) magnitude of contour curvature.

Since there was a significant effect of curvature magnitude on set depth, we can conclude that the shape of the bounding contour did significantly affect perceived depth, above any effect from depth propagation from the potato half of the stimulus. Our results (Figure 4.9 and Figure 4.10) also show that proximity from the bounding contour can affect perceived metric depth, which extends on previous knowledge that it is a factor in ordinal depth judgements (Norman & Raines, 2002).

Though the results showed a significant main effect of curvature magnitude on set depth, and thus advances our understanding of how 3D shape can be inferred from 2D contour curvature, the effects could not be attributed to specific levels of magnitude as paired comparisons of metric depth between specific curvatures were non-significant. There are several reasons why our experiment may not have been sensitive to what might be relatively subtle effects and these are discussed in relation to i) confines of the experimental method, and ii) potential refinements to the experimental method.

Though the potato stimuli in this experiment were designed to appear natural, their dimensions were not taken from real objects and thus there is no guarantee that stimuli approximated the real distribution of surfaces in the environment. Furthermore, the hypothesis of this study relied upon the assumption that surfaces are smooth, thus both 3D stimuli and 2D contours were designed accordingly. All contours in our set were segments from perfect circles, as this was computationally more straightforward. Furthermore, participants generally reported seeing the stimulus as a whole (thus it must have appeared realistic) and more importantly, probe depth settings near the bounding contour were reliably greater than the depth of the screen, indicating that observers did perceive the contour to have some 3D shape. Though it could be argued that ideal stimuli would be derived from real objects, such a database is not currently available, and it might make the design unnecessarily complicated.

Another barrier to detecting significant effects in such a study is that the task itself is not a natural one, for several reasons. Firstly, visual cues to shape in stereoscopic computer displays are artificial, providing cues to flatness from the frontoparallel monitor (e.g. from blur and accommodation; Watt, Akeley, Ernst & Banks, 2005) that conflict with the cues to depth from the stimulus. Secondly, in our study, the visual system was tasked with amodal completion of two shapes, each with different, ambiguous depth profiles. Nevertheless, the paradigm was sufficiently sensitive to detect main effects of contour curvature and thus more detailed analysis might be possible after refining the experimental parameters.

Indeed, there is much scope to make changes to our experimental paradigm for future research. For instance, in our display, the width of the contour section was set to equal that of the potato stimulus at its cross section (as can be seen in Figure 4.6 and Figure 4.8), so as to match the radius of the contour at Y = 0 with that of the potato stimulus. There are several alternative methods of choosing contour position, each with benefits and limitations. For instance, the chosen method does not take into account i) the visible area of the contour surface, ii) the length of the visible contour, iii) the start and finish points of the visible contour at the occluder. Indeed, there are two important implications that arise from the design of the contour surface. Firstly, when the potato and contour are amodally completed and perceived as a whole, the position of the contour influences the implied size of the completed stimulus. Since smaller 2D shapes are likely to arise from smaller 3D objects, then it is clear that the inferred depth of the probe might be influenced by the size of the stimulus. Indeed, the stimuli in this experiment were not controlled for in terms of implied total size, as can be seen in Figure 4.6. It is, however, interesting that depth near convex contours was in general set closer to observers than for flat or concave stimuli, despite their smaller implied size. Perhaps greater effects between contour magnitudes might be discovered when the contour positions are controlled differently.

A second implication of contour location is that the shape at the edge of the occluder can be a factor in amodal completion (Fantoni, Bertamini & Gerbino, 2005); if the stimulus as a whole was not convincing, then it is less likely for depth to propagate from the potato half of the stimulus, or for the contour to be seen in depth. However, if this were a serious issue within the current study one would expect the depth of probes to be set close to zero for concave contours, where the distance

across which amodal completion must take place is greater due to the design of our stimuli (see Figure 4.8).

A second aspect of the experiment that could be examined further is the range of contour values chosen. Only seven values (three positive, three negative, one flat contour) were investigated in the current study, though greater and lesser magnitudes could also have been presented. Indeed, as mentioned above, the contour values selected did not necessarily reflect the distribution of 2D contours in the real environment. Of particular note is that equal magnitude values were used for each sign, yet steeply curving convexities may not be realistic for the stimuli chosen (see Figure 4.8; Feldman & Singh, 2005). Again, ideally it would be best to choose contour values based on real world surfaces; the data for such analysis is already available (Potetz & Lee, 2003; Yang & Purves, 2003) and has been used for a similar purpose to ours (Burge et al., 2010).

The results discussed here do show that 2D contour shape is a significant cue to depth within the bounding contour and this information may be useful to integrate with models of contour propagation (Tse, 2002). Further research examining relationships between the environment and the retinal image would prove useful for guiding subsequent experiments.

Chapter 5

Measuring biases in slant and tilt perception

Chapter 4 investigated a perceptual bias that had previously been hypothesised but not empirically tested. Both theory and observation of real world surfaces suggested a relationship between 3D shapes and the 2D contours they project to, while experimental evidence demonstrated that silhouettes could evoke a convincing sense of 3D shape. Upon examination, we were able to confirm that the shape of the bounding contour does indeed affect depth perception interior to the contour. Further statistical analysis of real world scenes and how they project to the retinal image will be necessary to ascertain whether such biases arise due to the relationship between 3D shapes and 2D contour curvatures in the natural world.

Adhering to a psychophysical approach, the current chapter focuses on examining a *known* bias: slant underestimation. In contrast to what is known about the relationship between contour curvature and perceived depth, we already have evidence of slant underestimation (Andersen, Braunstein & Saidpour, 1998; Proffitt, Bhalla, Gossweiler & Midgett, 1995) and some knowledge about the distribution of surface slant in the real environment, both from statistical analysis of real scenes (Yang & Purves, 2003) and theoretical (geometrical) understanding of how slanted surfaces project to the retinal image. Nevertheless, many studies of slant perception are limited due to i) uncontrolled cues to zero slant (see Watt, Akeley, Ernst & Banks, 2005) and ii) only measuring slant about a single axis, which would predict a frontoparallel prior. Here, I examine biases in slant and tilt perception at multiple tilt axes, with the aim of describing perceptual priors for slant.

5.1. Introduction

The notion that vision may be informed by the statistics of the natural environment is partly driven by the observation that human perception is often biased. One such observation is that surface slant tends to be underestimated (e.g. Durgin et al., 2010a, Durgin, Li & Hajnal, 2010; Gruber & Clark, 1956; Porrill et al., 2010), though researchers tend to only examine such a bias at a single tilt axis,

without questioning whether this makes sense based on the possible distributions of surfaces within the environment. The aim of this chapter was, therefore, to create an experiment that would allow measurement of perceived surface attitude (the direction of the surface normal) and analyse how any perceptual errors map on to known, or theorised distributions of slanted surfaces within the environment itself.

Slant can be defined as the degree to which a surface is rotated from the frontoparallel plane (Figure 5.1), while tilt is defined as the axis of rotation for slant (Figure 5.2). In addition, slant can be described as optical (relative to gaze), geographical (relative to the horizon), or relative to another surface (Gibson & Cornsweet, 1952; Proffitt, Bhalla, Gossweiler & Midgett, 1995). For the purpose of this review, the literature cited mostly focuses on optical slant as there is much evidence that observers tend to underestimate the angle of rotation from the frontoparallel (optical) plane (e.g. Andersen, Braunstein & Saidpour, 1998; Girshick, Burge, Erlikhman & Banks, 2008, abstract; Porrill, Duke, Taroyan, Frisby & Buckley, 2010; Todd, Thaler & Dijkstra, 2005). In addition to evidence of slant underestimation, research has also shown that observers underestimate the depth of vertical ridges (Adams & Mamassian, 2004) and together these suggest that observers might utilise a 'frontoparallel prior', where surfaces are perceived to be less slanted (see Figure 5.3; van Ee, Adams & Mamassian, 2003; Caudek, Fantoni & Domini, 2011) or flatter than physically specified in the frontoparallel plane (Adams & Mamassian, 2004).

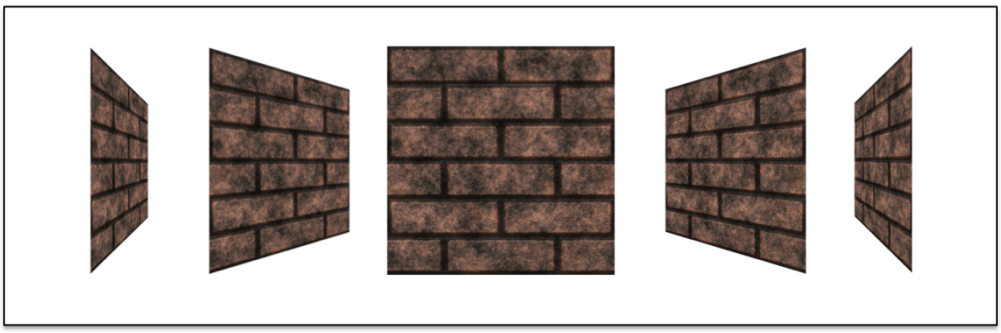


Figure 5.1 In this demonstration of slant, the centre image has a slant of zero and is thus frontoparallel. The images either side have been rotated about a vertical axis by various angles, so that the surface is receding in depth.

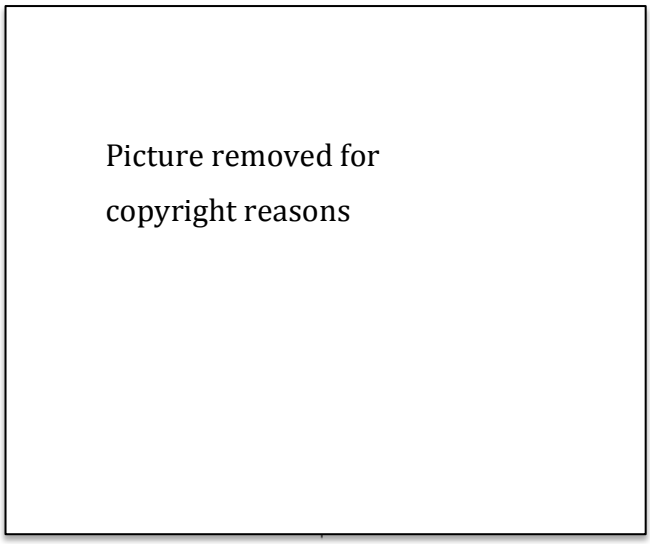


Figure 5.2 In this demonstration of slant and tilt, slant increases across the radial axis, while the angular location represents tilt (the axis of rotation). Note that tilt cannot be defined when there is zero slant. Adapted from Norman, Todd, Norman, Clayton & McBride (2006).

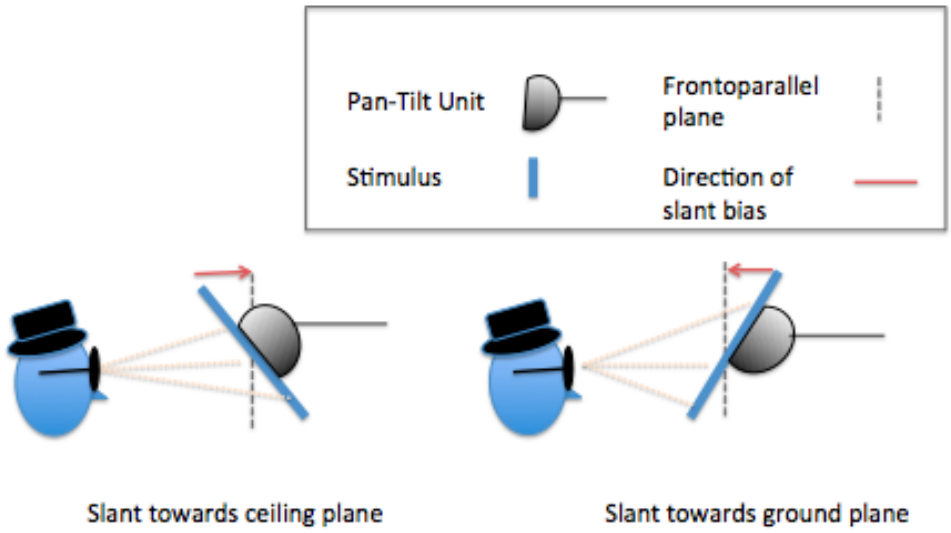


Figure 5.3 Demonstration of slant underestimation as predicted by a frontoparallel bias. The Pan-Tilt unit is a device that can mechanically change the slant and tilt of a planar stimulus.

In one experiment, for example, participants were shown textured surfaces that were rotated 40˚ and 80˚ away from the frontoparallel plane about a horizontal or vertical axis (Andersen, Braunstein & Saidour, 1998). Observers' mean judgements of slant were 14.5˚ and 12.1˚ for a stimulus slant of 40˚, and 53.5˚ and 48.4˚ for a

stimulus slant of 80˚ (about a horizontal and vertical axis respectively). This demonstrates a tendency to underestimate slant from the frontal plane at both horizontal and vertical tilt axes. One can model the probability of encountering slants in real environments in order to question whether such a bias is predicted by priors that mirror experience.

If we only consider surfaces in the world that are rotated about a single tilt axis, and that all world slants are equally likely, then a frontoparallel prior is logical; frontoparallel surfaces take up a greater proportion of the visual field compared to the same surface when slanted. This principle is demonstrated under monocular viewing conditions in Figure 5.4. At zero degrees slant, the surface projects to the greatest proportion of the visual field (or retinal image). When the surface has a slant of 65˚ it takes up a smaller proportion of the visual field, becoming invisible at a 90˚. The relationship between surface slant and its projected area on the retina is given by area $\propto \cos(\theta)$, where θ is the slant of the surface. Thus, as slant increases, a surface projects to a smaller area of the retinal image.

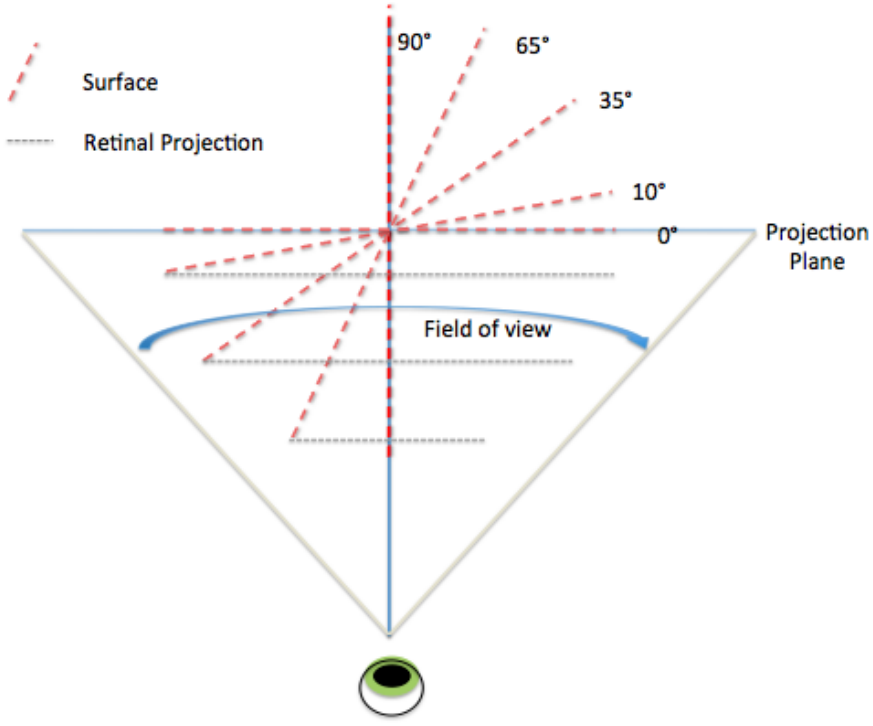


Figure 5.4 The diagram shows a cyclopean eye viewing the same surface at different degrees of rotation about a vertical axis. This demonstrates the relationship between surface slant and its projected area on the retina. When the surface is not slanted, it projects to the largest space in the retinal image. As the slant increases the projected area becomes smaller. The surface is not visible at 90˚.

In reality however, surfaces can have any tilt axis and therefore the prevalence of oriented surfaces across both slant *and* tilt should be taken into consideration. The principle of orientation prevalence can be demonstrated using a tessellated hemisphere as an approximation of all possible slant and tilt combinations, though the actual number of potential attitudes is infinite. The tessellated hemisphere in Figure 5.5 is a graphic representation of this idea. Only the single, central patch is frontoparallel. Patches equidistant from the centre all have the same slant but different tilts; as the distance from the centre increases, patches become more slanted and greater in number. Therefore the most prevalent slant is at 90˚; 0˚ slant is the least common. The distribution of slant is proportional to sin(θ). Although zero slant surfaces project to the largest retinal area, they are least prevalent. In contrast, 90˚ slants are most prevalent but project to the smallest area on the retina. To obtain the predicted prior for slants we can multiply the prevalence (in the natural environment) by projected surface area (in the image), i.e. p(θ on the retina) is proportional to cos(θ)sin(θ). The resulting curve, demonstrated in Figure 5.6 indicates that the most likely slant to be encountered is 45˚.

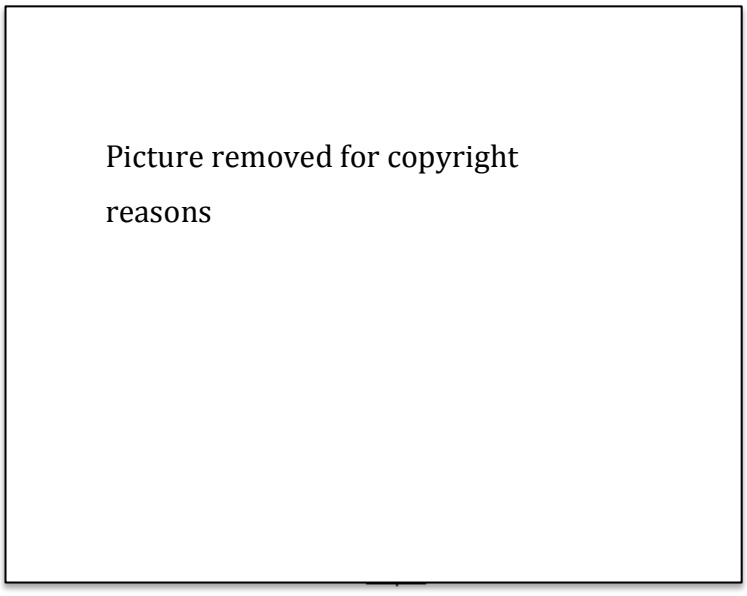


Figure 5.5 The tessellated sphere represents an approximation of all possible slant and tilt combinations in the world. Only the centre patch is frontoparallel. Slant prevalence increases as patches become farther from the centre.

The modelled distribution of slant does not match the results for frontoparallel bias found in previous psychophysical studies. There are three main reasons for this discrepancy. Firstly, the aforementioned calculations of surface prevalence being proportional to sin(θ) are dependent upon the assumption of uniform distribution of attitude. However, it is likely that there is a non-uniform distribution across surface attitude in the natural environment due to anisotropies present in natural scenes. Indeed, there are two major potential anisotropies in the distribution of surface attitude, one being a dominance of slants about a vertical tilt axis, and the other being about a horizontal axis. For instance, one might expect to find more slants at the vertical axis close to zero. This is because building design tends to involve walls and boundaries that go straight up as opposed to sloping. Surface attitudes in entirely natural environments might tend towards vertical or horizontal due to the effect of gravity. Similarly, more slants about a horizontal axis may be expected due to the ground plane itself.

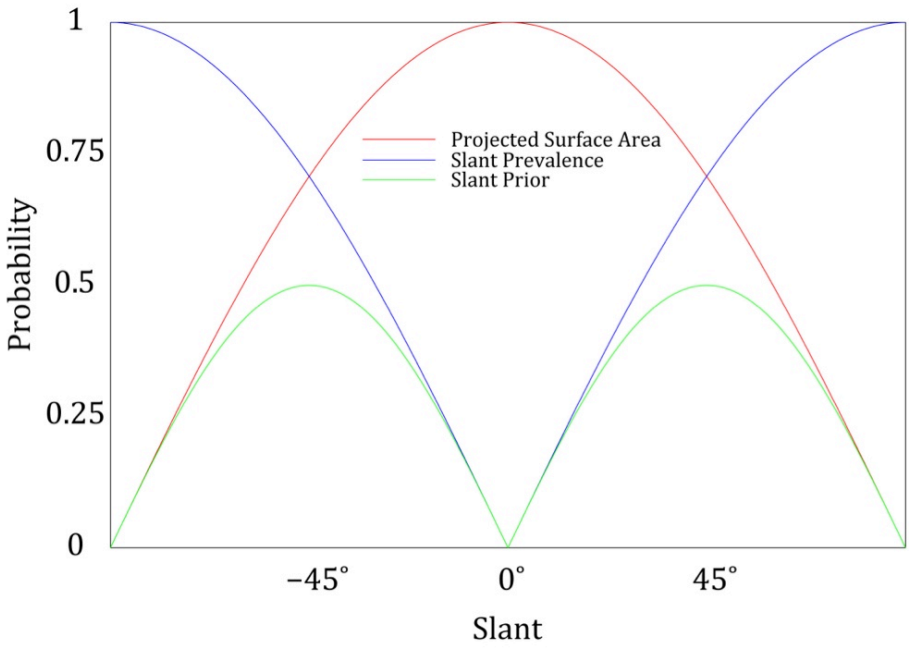


Figure 5.6 Projected surface area has a maximum value when slant is zero. Slant prevalence increases as values move away from frontoparallel. Here, the prior for slant peaks at ±45°.

Indeed, evidence shows there are some notable anisotropies in surface attitudes in the natural environment. Yang & Purves (2003) used a laser range scanner to measure properties of natural scenes and found anisotropies in the joint distribution of slant and tilt of planar surfaces (see Figure 5.7). It is very clear that neither slant nor tilt is isotropic (equally distributed). In particular, Figure 5.7b shows that surfaces are most often slanted about the horizontal and vertical axes. Furthermore, when calculated independently of tilt, slant peaks at 65˚. Potetz and Lee (2003) add that in most images surfaces are slanted due to the receding ground plane, and objects get progressively further away from the observer (in this case, the scanner) as the vertical distance increases. One implication of these findings is that observers should have more experience at judging slants about the cardinal axes and thus may be better at discriminating these. More importantly, when the tilt axis is more ambiguous (when slant is smaller for instance) prior knowledge might be used to bias perception towards more frequently seen attitudes (e.g. surfaces rotated about a horizontal or vertical axis).

Indeed, recent evidence demonstrates that observers exhibit biases when perceiving ambiguous 2D surface orientations and these closely match peaks in the distribution of surface orientations in natural images (Girshick, Landy & Simoncelli, 2011), thus one might also expect a tight link between biases in *3D depth perception* and the distribution of *surface attitudes* in the real environment.

A second reason for the poor correspondence between the modelled distribution of slant and reported frontoparallel bias is that participants are typically shown slants about a single tilt axis (e.g. Porrill et al., 2010; Todd, Thaler & Dijkstra, 2005). In such a situation, participants might rightly assume a prior conditional upon a single tilt, which is proportional to cos (θ), as modelled in Figure 5.6, in which case we would predict a bias toward frontoparallel (zero slant).

Figure 5.7.Figure showing the independent probability distribution of a) slant and b) tilt of planar surfaces in natural scenes (where zero tilt corresponds with slants about a horizontal axis) c) Shows the joint distribution of slant and tilt; contours are coloured according to the colour bar, with red representing high probability and blue representing low probability. Note the clear asymmetries in slant and tilt in the environment. Adapted from Yang & Purves (2003).

The third plausible reason for the documented underestimation of slant is that previous methodologies may have inadvertently contributed to, or caused the biases that have been reported. In particular, Watt and colleagues (Watt, Akeley, Ernst & Banks, 2005) provide compelling evidence that when slant stimuli are presented on displays that are frontoparallel to the observer, that depth and distance stimuli are confounded by uncontrolled focus cues to depth such as retinal blur. When viewing a scene under natural conditions blur occurs at distances nearer or farther than the focal point. Computer presented stimuli however are uniformly blurred, as the screen is set up frontoparallel to the observer. This distorts the perception of depth at all simulated points other than the true depth of the screen (Watt et al., 2005). By varying the physical distance of the monitor from the observer, Watt et al. were able to study the effect of eye vergence as a cue to depth in disparity displays. They concluded that vergence has a considerable impact on perceived depth. The implications are significant for studies using image disparity as a cue to depth since most procedures do not vary the focal distance to be equivalent to stimulus-defined depth. Focus cues are also an important issue to consider in studies of slant perception. The visual system is adept at combining all available cues in order to reduce variation in perceptual estimates (Hillis, Watt, Landy & Banks, 2004; Knill & Saunders, 2003). Without controlling for focus cues, the texture-defined slant may be combined with zero slant from the monitor used to display stimuli, thus procedures may not measure the true perceptual bias. The aim of the current study was, therefore to determine biases in slant and tilt perception using a method that would avoid, or control for confounding cues to zero slant.

5.2. Methodological Considerations

In the current study, participants were shown a stimulus that could be rotated about multiple tilt axes. They were asked to match the perceived slant using a haptic paddle and responses were recorded digitally. The equipment was designed and set up to facilitate different parts of the experiment, as discussed below.

5.2.1. Stimulus presentation

The simplest way to avoid confounding cues to zero slant from frontoparallel monitors is to present real stimuli that can be rotated about an axis. Many

researchers have tackled the problem of real stimulus presentation within the last decade or so, though each method has limitations that could be avoided. Example methods include: setting the slant of a physical stimulus using a protractor (Porrill et al., 2010), fixing surfaces to a mount that can be easily adjusted by hand (Durgin, Li & Hajnal, 2010) and presenting blocks of fixed slant that can be switched after each trial (see Figure 5.8; Taylor-Covill & Eves, 2013). There are three critical limitations to each of these methods: i) manual surface switching is time consuming, thus limits the number of possible trials ii) manual presentation makes it difficult to use more than one tilt axis and iii) they require the experimenter to be present at all times.

An ideal solution to these issues is to use equipment that can rapidly and accurately set the attitude of a surface, without the need for the researcher to be present. Thus, a pan-tilt unit was acquired for the current experiment that allowed i) rapid, high precision setting of slant, while ii) allowing surface attitudes across the full range of possible tilt axes (0-360) and iii) remote programming made it unnecessary for the experimenter to be present during trials.

5.2.2. Recording of attitude responses

Several methods have been used to record perceived surface attitude. These include verbal reports of slant (e.g. Durgin, Li & Hajnal, 2010; Proffitt et al., 1995), positioning a physical surface to match the stimulus (Durgin et al, 2010a, b; Taylor-Covill & Eves, 2013), and setting the angle of a line relative to a horizontal line on a computer screen (e.g. van Ee, Adams & Mamassian, 2003). While verbal reports of slant result in immediately accessible data and do not require calibrated equipment, the method does make the assumption that observers have a good internal (and conscious) representation of slant. The verbal report method was not suitable for the purpose of the current study, as one would have to assume that observers clearly understood the concepts of both slant *and* tilt, and that they would not be subject to cognitive strategies (e.g. rounding numbers to regular intervals).

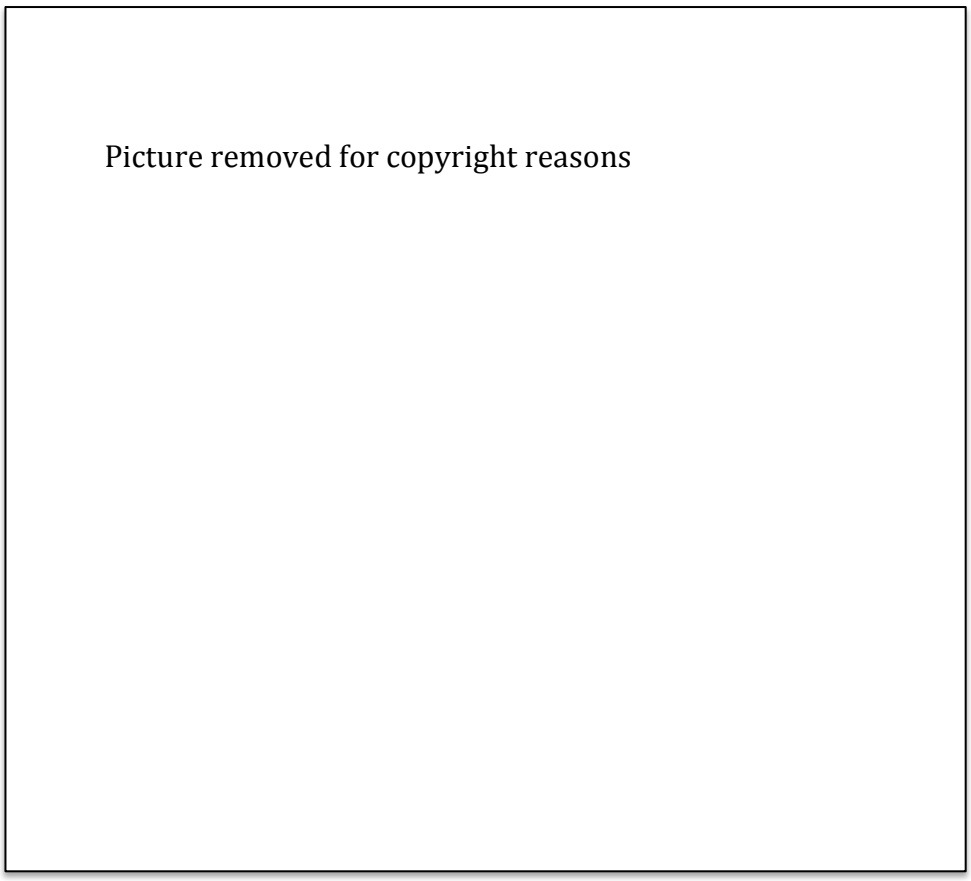


Figure 5.8 This image shows a participant setting a palm board to match the slant of the stimulus surface (right; wooden block). An inclinometer is attached to the palm board (positioned roughly at waist height, and rotated about the horizontal axis) to digitally record set slant. A head mount was worn to hide the hand from view. Adapted from Taylor-Corvill & Eves, 2013.

The method of recording perceived slant for the current study was to use a haptic paddle that could rotate about a horizontal and vertical axis, while two rotary encoders digitally recorded the set attitude of the haptic paddle. An example of a single-axis haptic paddle (or palm board) is shown in Figure 5.8. The use of digital equipment is a great improvement on manual recording methods as it means data acquisition is almost automatic and should record very accurate and reliable data. Nevertheless, caution must still be taken when designing haptic measures of slant perception, as evidence shows that there are large variations in results between different methods. Indeed, Durgin et al. (2010a, b) showed that slant is underestimated to a greater extent when a haptic paddle was used, relative to the free-hand method (where the hand itself is used to match perceived slant and responses are recorded using optical probes attached to the hand). The cause of this

discrepancy was found to be due to limit of proprioceptive movements when the paddle is at waist height, where the wrist is poorly calibrated to match slant surfaces accurately. The wrist has a limited degree of rotation about the horizontal plane, so perhaps when that limit is nearly reached then the hand feels more steeply sloped than it physically is. Since the haptic paddle in the current study will be positioned central to the observer's gaze it will i) allow participants to move both wrist and elbow, thus minimising any proprioceptive bias (see Durgin et al., 2010a) and ii) avoid unnecessary perceptual translations from the ground plane to frontoparallel plane.

Despite the improvements made to the experiment, a haptic paddle is not without its own difficulties. For instance, there may still be anisotropies in the error of haptic settings due to the limits of hand movements or wrist flexibility. The purpose of the first part of the experiment (Part 1) will be to calculate haptic response biases under full cue conditions so that they can be controlled for in the Part 2 of the experiment.

5.2.3. High- and low reliability stimuli

For the current experiment, two textured stimuli were designed: Texture 1 provided high reliability cues to slant, while Texture 2 provided low reliability cues to slant. Girshick, Landy and Simoncelli (2011) also used cross-noise trials to extrapolate a prior for orientation (also see Girshick et al, 2008, abstract). In their study, observers were shown two simultaneous displays of oriented Gabor patches and their task was to judge which of the two displays was rotated more clockwise on average. In low noise stimuli all orientations were the same (no noise), while for high noise stimuli the orientations were more variable. Perceptual judgements for both stimuli were subject to the same *internal* noise, yet differed in *stimulus* noise, thus the *relative* response bias between high and low noise stimuli should be informative about observers' perceptual bias. Indeed, in low noise versus low noise trials in Girshick's study, discrimination thresholds were lower at the cardinal axes and responses were more variable at oblique angles (such effects must be due to internal noise, since there was no stimulus noise). In contrast, the oblique effect was less noticeable in high noise versus high noise trials, due to the high level of stimulus noise. Nevertheless, upon examination of the *relative* bias in high versus low noise

trials at points of subjective equality, there was no relative bias at horizontal, vertical, or oblique orientations. There was, however, substantial relative bias (up to 12°) at all other orientations. Importantly, this implies a bimodal prior, whereby perception is biased towards horizontal and vertical orientations, yet neither bias is more influential at oblique orientations.

To allow similar comparisons in the current study, the stimulus in Part 2 will be interchanged between a 'low-reliability' and 'high-reliability' texture and the relative bias can then be calculated. If we were to predict a prior for zero slant, then we would expect no relative bias at zero and an increasing relative bias as slant increases. If no perceptual prior exists, then we should also find no evidence of relative biases, though the variability of settings would increase from the low to high reliability condition.

5.3. Outline of Current Experiment

In both parts of the study, participants viewed a textured surface that had been rotated away from frontoparallel by a slant of 0-60˚ about a tilt in the range of 0-360˚. Texture provided either high-reliability or low-reliability cues to surface attitude (Texture 1 and 2 respectively). Texture 1 consisted of high contrast, regular circles of the same size, while Texture 2 contained randomly oriented ellipses of different sizes and different aspect ratios, thus making changes in texture from increased slant more ambiguous.

Observers were asked to match the surface using a haptic paddle (see Figure 5.9) that was central to the observer's gaze (though not visible). A dual-axis rotary encoder attached to the haptic paddle accurately recorded participant settings and these could then be translated into perceived slant and tilt. A diagram of the room setup is shown in Figure 5.10.

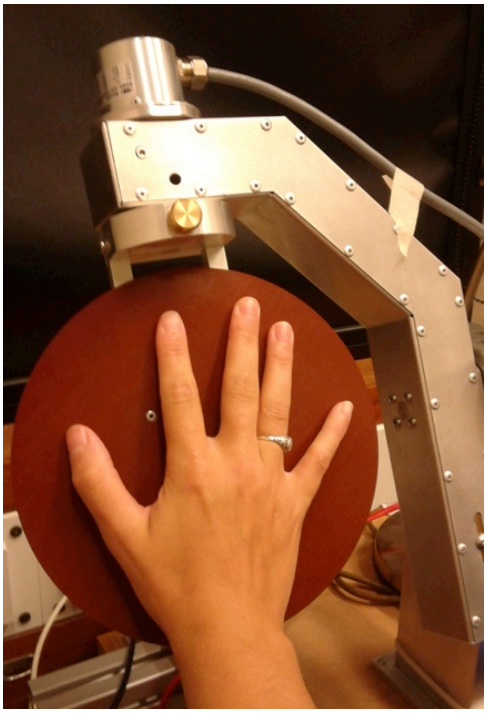


Figure 5.9 Photograph of haptic paddle in use. An unseen axis at the back of the paddle allows it to rotate left to right; the metal arm (right) rotates to allow paddle movements towards the ground or ceiling plane.

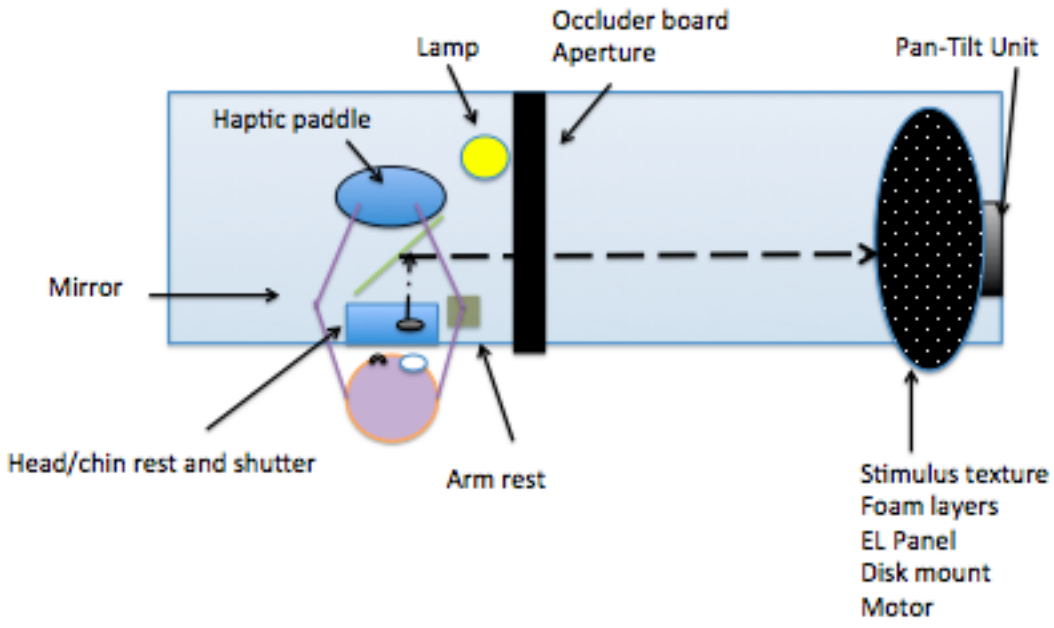


Figure 5.10 Layout of experiment room for Part 2 (limited cue condition). Participants could not see the haptic paddle as it was occluded by the mirror. In Part 1 the occluder was removed and a longer mirror was used for binocular viewing. Not to scale; see Appendix for dimensions.

Part 1. Full Cue Condition: In order to calculate the response bias associated with the haptic paddle, a full-cue condition was run in which observers' perception was assumed to be close to veridical (with the exception of internal error). Haptic paddle settings were obtained for a selection of surface attitudes. Participants viewed Texture 1 under binocular conditions and were free to observe the entire stimulus in motion between trials, thus cues to surface attitude were available from texture, binocular disparity, focus cues (blur and accommodation), motion and compression of the stimulus outline.

From participants' haptic paddle settings it was possible to quantify perceived attitude as a function of response attitude. This could then used to estimate perceived slant/tilt from the responses to the monocularly viewed stimuli from Part 2 of the experiment.

Part 2. Limited Cue Condition: For the second part of the study, participants viewed both Texture 1 and Texture 2 at the same surface attitudes as presented in Part 1. Cues to surface attitude were limited by i) viewing the stimulus through an aperture that a) concealed the stimulus outline and b) limited the size of the stimulus on the retina (reducing cues from texture), ii) monocular viewing, thus removing disparity cues, iii) using a shutter-operated aperture to prevent participants from viewing the stimulus in motion and iv) limiting viewing time to increase uncertainty.

5.4 Possible Outcomes

To our knowledge, no studies have thoroughly examined human slant perception across the multiple tilt axes (with the exception of Andersen et al., 1998, who used either a horizontal or vertical axis on separate blocks of the study). However, there are several outcomes that can be predicted based on previous research:

1. Haptic response bias: In the full cue condition (Part 1) it was assumed that perception would be veridical, and thus any error would originate from participants' haptic response bias. Based on the most comparable, 'free-hand' (Durgin et al., 2010a) method, it was predicted that haptic paddle settings would be close to veridical. Though there is evidence that the position of the haptic paddle relative to the observer can significantly affect the accuracy of paddle settings (Durgin et al., 2010a),

the cause of this was found to be limitations in the flexibility of the wrist; this should not be a factor in the current study, since the frontoparallel position of the paddle allows observers to move fingers, wrist, elbow and shoulder. Nevertheless, to the author's knowledge haptic paddles have only been used to examine perception of slant about a horizontal axis (relative to the frontoparallel or ground plane), thus care will be taken to examine haptic error.

2. Calculating biases in perceived surface attitude: As outlined in the main introduction, a frontoparallel prior only makes sense if observers assume a single tilt axis. Since observers will be shown surfaces that are slanted about multiple tilt axes in the current experiment, one would only predict a frontoparallel bias if a prior for slant was conditioned upon a single tilt.

The main prediction for this study is that observers will exhibit a prior towards 45° slant, reflecting the product of slant prevalence (about multiple tilt axes) and projected surface area in the image. Nevertheless, this prior largely relies upon the assumption of a uniform distribution of surface slant in the natural environment. Indeed, Potetz and Lee (2003) showed that the distribution of slant is i) anisotropic and ii) dependent on tilt, thus we may expect slant biases that are dependent on tilt. For this reason, another possible outcome is that observers might exhibit a bias towards ground plane slants in line with a ground-plane prior. In all cases, one would predict greater influence of a perceptual prior for Texture 2 relative to Texture 1, and also at low slants, where cues to surface attitude are more unreliable.

Relative to the wealth of research invested in understanding slant perception, little is known about observers' ability to match tilt. One notable exception is Norman et al., (2006) who found that participants are sensitive to differences in tilt but that tilt judgement is not independent of slant. Specifically, tilt discrimination is easier at larger slants. In order to explain this, they describe a Gauss sphere on which four locations are marked (see Figure 5.11). Though *ab* has the same tilt separation as *cd,* tilt discrimination is more difficult in the latter case. The authors argue that this is because observers use the arc length between points for discrimination. Nevertheless, observers were able to selectively discriminate along one dimension while ignoring changes in the other dimension. If observers do have a ground plane prior, then this should emerge particularly at low slants, where surfaces might be perceived to be ground plane slants more often than ceiling plane slants.

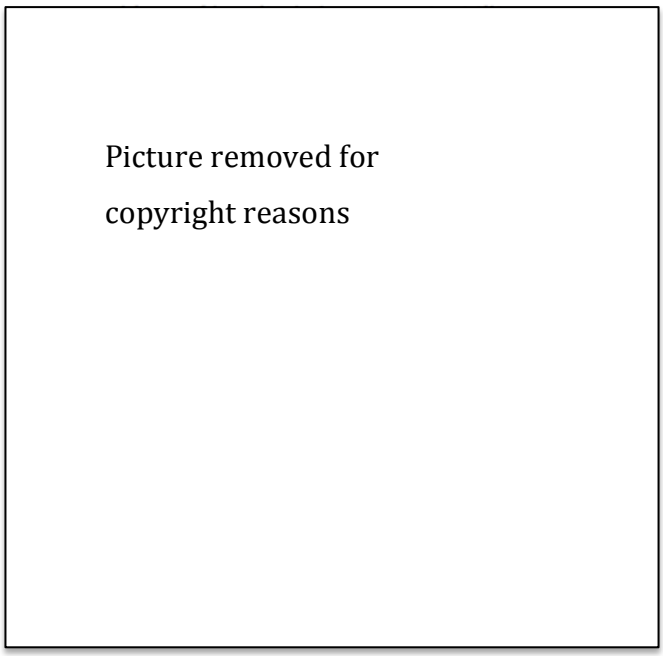


Figure 5.11 This is a Gauss map, showing slant and tilt with respect to the observer. The slant is the same at *cd* and *ab*, while tilt is equal for *ac* and *bd*, yet tilt is more difficult to discriminate at *cd* because the arc length is shorter. Adapted from Norman, Todd, Norman, Clayton & McBride (2006).

5.5. Method

5.5.1. Participants

Nine participants (5 female) from the University of Southampton took part (mean 28, SD 5.13). All had normal or corrected to normal vision and were right handed. The local ethics committee gave approval and participants signed written consent. Student participants received a small payment for their time.

5.5.2. Apparatus

5.5.2.1. Stimulus Presentation

A pan-tilt unit (PTU D46, PTU D48E; FLIR) was used to display textured surfaces at different attitudes (see Appendix for more details). The PTU was controlled remotely from a Macintosh computer and commands were sent from

Matlab using PsychToolbox (Brainard, 1997). The PTU unit was capable of rotating 360° around the azimuth and ~70° of slant (user limit set at 61°, as large slants limited the maximum aperture size, and thus viewing angle).

Stimuli were generated by printing textures onto 66cm diameter, transparent Perspex discs which were then attached to the PTU. The two stimulus textures were designed for 1) high-reliability and 2) low-reliability. There are several ways in which we can affect texture reliability (see *Figure 5.12*).

Texture 1 (high reliability) consisted of high contrast (luminance = 0 (min), .5, 1(max)), opaque circular discs with radius of 0.5cm. Texture 2 (low reliability) consisted of high contrast (seven luminance values from 0 to 1), randomly oriented opaque ellipses of varying lengths (~1cm – 12cm) and aspect ratios (see Figure 5.13). The second texture was less reliable for several reasons: i) randomly oriented texture elements disrupts the texture cue of compression/foreshortening ii) randomly sized elements means the cues of density and size gradient are less reliable iii) fewer texture elements per degree of visual angle means there is less information on which to base attitude judgements, iv) local contrast is reduced, making texture elements less distinct/salient.

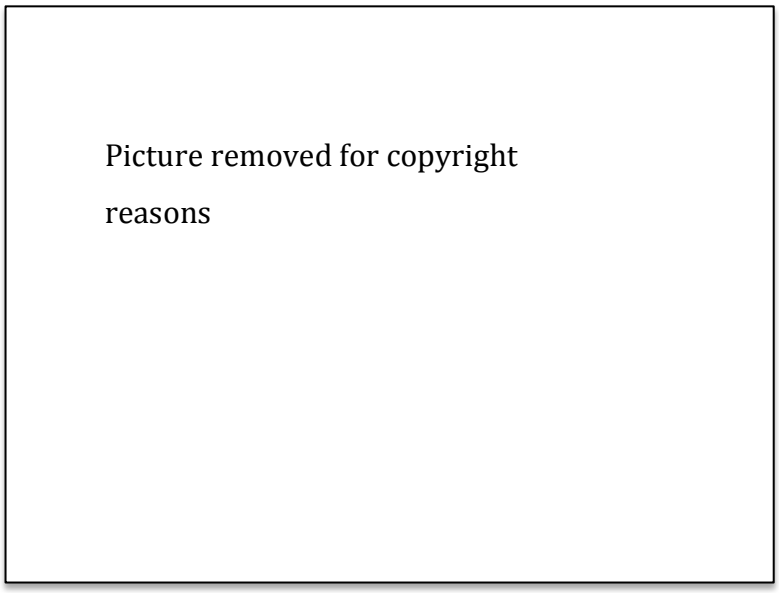


Figure 5.12 Outline and textural cues to surface attitude. a) The outline of the surface will become compressed when slanted; its aspect ratio also provides a cue to tilt, b) size gradient and c) density gradient can be a cue to surface attitude if elements have a uniform size, and d) compression gradient (or aspect ratio) can be informative if texture elements have a regular shape. Adapted from Saunders and Backus (2006).

Participants' heads were stabilised using a chin- and headrest, keeping their eyes aligned with a mirror, fixed at 45° from frontoparallel. The mirror allowed the haptic paddle to be central to the observer's gaze without obscuring the view to the stimulus. In Part 1 the mirror was positioned such that it was centred at 5.5cm in front of the nose; in Part 2 the mirror was centred in front of the right eye and a shutter occluded the mirror until opened by button press.

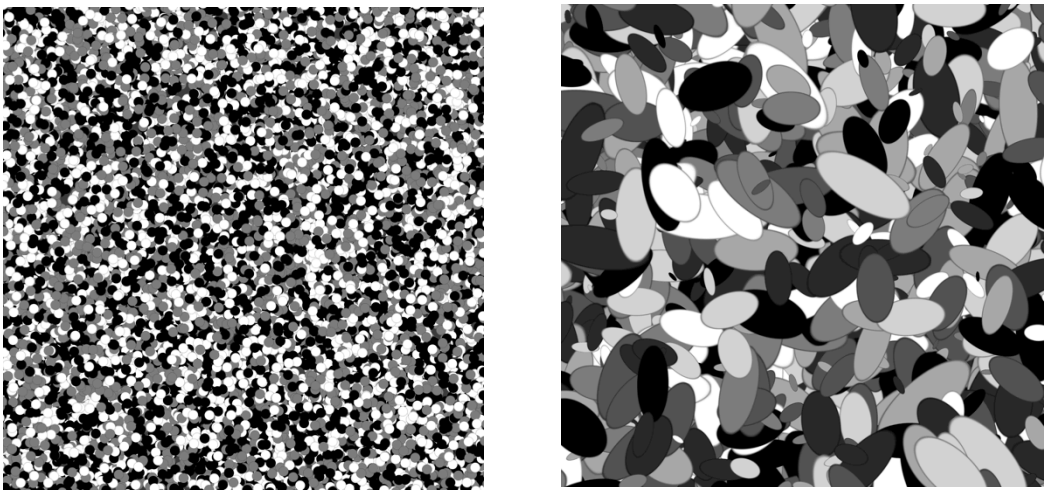


Figure 5.13 Textures used in the current study. Texture 1, left: reliable cues to slant (high contrast, small circles of the same radius). Texture 2, right: less reliable cues to slant (randomly oriented ellipses of varying size). Note: texture contrast decreased substantially when printed on Perspex.

In Part 2 a large occluder was placed at a distance of 54cm from the right eye (including distance from eye to mirror) in order to conceal the edges of the stimulus (thus removing attitude cues from the surface outline). A small (~3.5cm diameter), irregular aperture was cut in the occluder, directly aligned with the centre of the frontoparallel stimulus; the aperture limited the visible stimulus to ~3° visual angle. The aperture was set at a far distance from the stimulus, in order to increase blur and thus helping to avoid participants attributing the outline of the aperture to the boundary of the surface. An approximate demonstration of the reliable texture in the full cue condition, and unreliable texture in the reduced cue condition are given in *Figure 5.14* and *Figure 5.15*.

Figure 5.14 The two images represent example stimulus attitudes in the full cue condition (Part 1; Texture 1). Here, images show a slant of 60 degrees from frontoparallel (towards left wall and right wall slant respectively) about a vertical axis. Note that the compression of the stimulus outline (change in aspect ratio) and relative density cues to slant sign. The sense of slant is compelling in this demo, despite it being shown in 2D. The textures are not shown to scale.

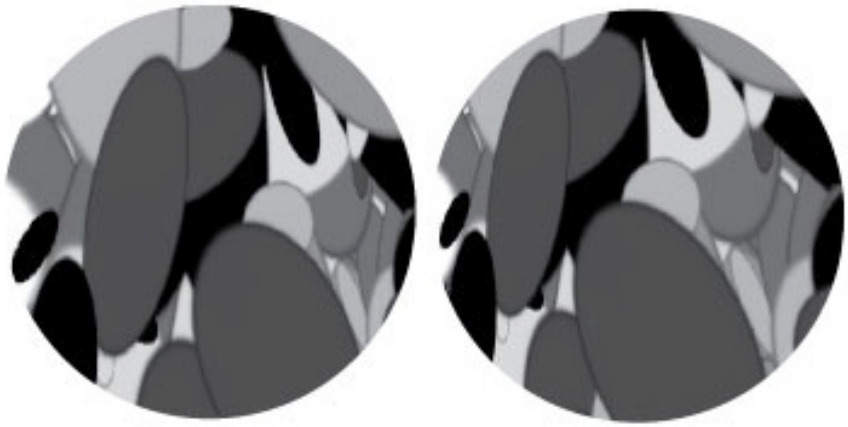


Figure 5.15. Here, images show a slant of 60 degrees from frontoparallel (towards left wall and right wall slant respectively) about a vertical axis, thus demonstrating how the stimuli might look in Part 2 (limited cue condition, Texture 2). The aperture occludes the edges of the stimulus and disparity cues are lost in monocular viewing.

An electroluminescent (uniformly lit; EL panel) light disc with a diameter of 66cm (SureLight) was installed between the texture stimulus and pan-tilt unit base in order to prevent luminance change as a function of surface attitude. If the light source came from above, then the brightness of a surface patch would be proportional to cos(θ) where θ is the angle between the surface normal vector and the primary direction of illumination, i.e. patch brightness would be a cue to surface attitude.

In the natural environment, the luminance of a patch within retinal image is correlated with distance to the observer, and lighter surfaces tend to have ground plane slants (Potetz & Lee, 2003). Observers might use these relationships to infer information about surface attitude, thus using a uniformly lit disc rather than a ceiling light for instance prevents the use of lighting changes as a cue to surface attitude.

A lamp was placed on the observer's side of the occluder, directly facing the aperture; its' purpose was to create a clear contrast between aperture and stimulus, as pilot testing revealed that, on occasion, observers interpreted the aperture edge as the bounding contour of the surface. All other lights were off during both parts of the experiment and the walls and tables of the lab were covered with a matte black cloth to prevent reflections from the EL panel.

Individual texture elements could become a cue to surface attitude in individual trials, as observers might use the relative orientation of ellipses to infer surface tilt. For this reason, a motor was positioned between the pan-tilt unit and the stimulus base that allowed the textures to be rotated slightly with each movement of the pan-tilt unit.

5.5.2.2. Data Acquisition.

A haptic paddle was used by participants to report perceived surface attitude. The paddle was a 19cm diameter wooden disc that was fixed to a mount at two independent axes (*Figure 5.9*). The paddle could be rotated ±170˚ around a vertical and horizontal axis. The paddle was securely mounted on the table in front of the observer, who had free movement of arm, wrist and shoulder to move the paddle; a padded armrest was used to prevent fatigue.

A two-axis absolute rotary encoder was attached to the paddle in order to digitally record haptic settings and relay these to the computer via a DAQ (data acquisition) device. The encoders work by continuously monitoring the angular

position of the shaft to the nearest degree, thus providing a rapid and accurate method of collecting data, which then simply needed to be transformed before analysis. Data from the haptic paddle were recorded every 10ms and the average of the final 10 judgements (prior to termination of the trial by button press) was used as the participant's haptic setting for that trial.

5.5.3. Procedure

In Part 1, participants viewed Texture 1 (high-reliability) binocularly, without the aperture or occluder. They then used their right hand to move the haptic paddle to match the perceived surface attitude and pressed a key with the left hand when they were happy with their setting. After each trial, the stimulus moved to its new attitude and the surface was rotated about the surface normal by a random amount (mean = ±15°, S.D. = 4°). Participants were free to watch the stimulus move.

In Part 2, the occluder and aperture were positioned in order to limit visual cues to the attitude of the stimulus surface. Participants wore an eye patch over the left eye and viewed the stimulus through the right eye. A shutter was positioned in front of the right eye that could be opened for 3 seconds upon button press. A beep sounded on each trial when the stimulus had stopped moving, thus signalling to the participant to press the shutter button and view the stationary stimulus. Participants were allowed to press the shutter button a maximum of twice per trial.

5.5.3.1. Trials

In order to gain a good measure of performance across attitude space, a program was used to sample 36 attitudes on the surface of a hemisphere (in radians). If the sum squared error of difference between the selected vector and previous vectors was less than .2 then the space was resampled. The same set of 36 stimulus attitudes was used for each participant in each condition (see Figure 5.16, top).

Treating Part 1 and each texture from Part 2 as a condition, each was split into two sessions, making a total of 6 sessions that were randomised for each participant. Each surface attitude was repeated 10 times and trial order was randomised, though no two consecutive trials had the same attitude.

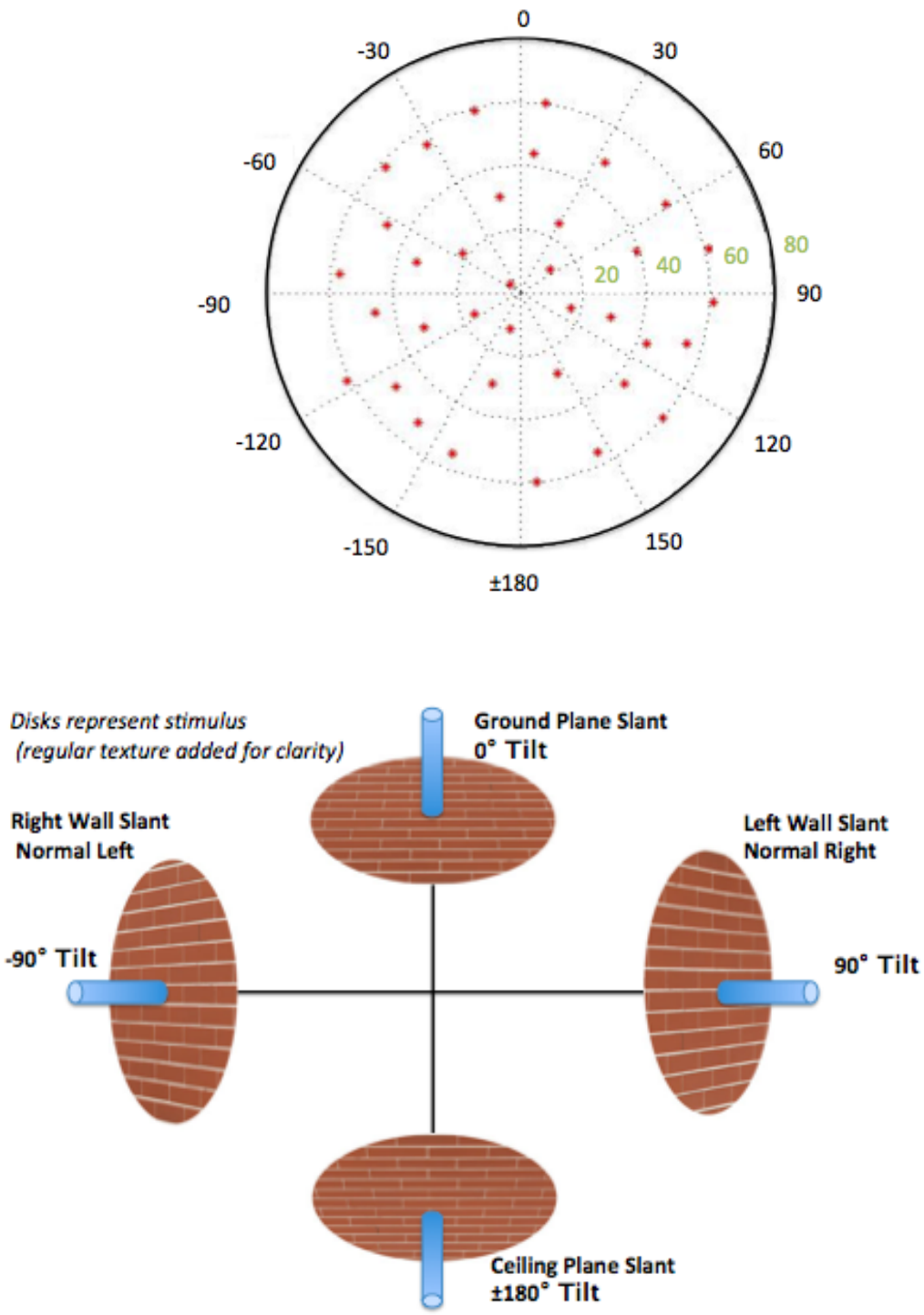


Figure 5.16. Top: The polar plot illustrates the 36 test attitudes used in the study. Tilt is defined along the circumference, slant across the radius. Bottom: Demonstration of the coordinate system for slant and tilt. Blue cylinders are added for visualisation of surface normal.

5.6. Results

5.6.1. Data Handling

Set attitude was recorded on each trial by taking readings from both axes of the rotary encoders that were attached to the haptic paddle. Since the haptic paddle operates around axes x (horizontal) and y (vertical), yet the PT Unit codes rotations from y (slant) and z (tilt, where z is perpendicular to stimulus; see Appendix) it was necessary to convert set attitudes into slant and tilt coordinates. In order to do this, the rotations of the haptic paddle from frontoparallel were calculated, about Axis 1 and Axis 2 to obtain the 3D surface normal after rotation (where the surface normal provides the vector coordinates at which a line is perpendicular to the surface). Set attitudes could then be converted into slant and tilt (see Appendix). Since participants viewed the stimulus through mirror (set at 45°from frontoparallel), all set and stimulus tilt values were subsequently mirrored about the vertical axis to correspond with the participants' view of the stimulus.

On some trials of the experiment we expected tilt to be perceptually reversed, since cues to surface tilt become more ambiguous at long distances. At short distances, an orthographic view of a slanted stimulus will generate a larger retinal image for closer parts of the surface than for distant parts. At longer distances, observers have a perspective view; the size of a retinal image is not directly dependent on the distance from the eye. Indeed, although the cue to slant from the compression gradient (change in aspect ratio) of texture elements is still available at long distances, opposite *tilts* will produce very similar images. This is because size and density gradients become more ambiguous at long distances and our stimuli do not contain a vanishing point, as might be informative in textures with parallel lines. A demonstration of a slanted surface with ambiguous and non-ambiguous tilt is given in Figure 5.17. To account for perceptual tilt reversals, in cases where set tilt was >±90 from the stimulus tilt, then the set tilt was mirrored about the axis orthogonal to set tilt, and saved for further analysis. The number of trials in which tilt was reversed was calculated in each condition for each participant.

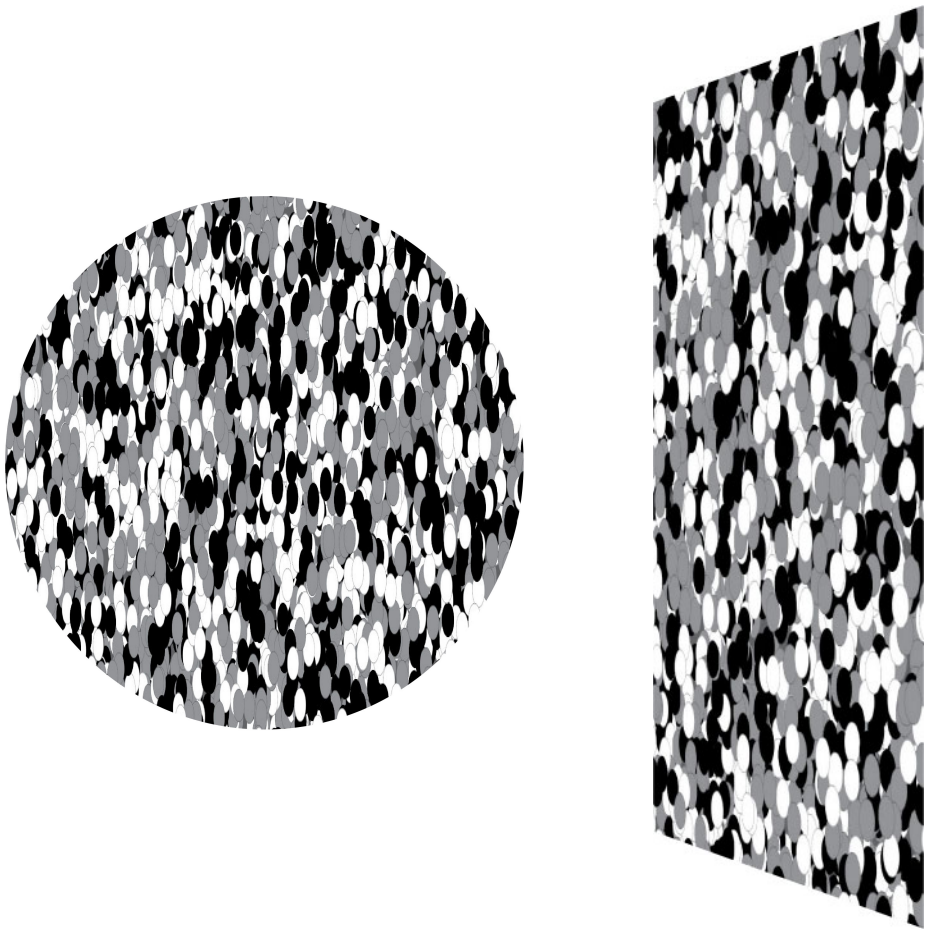


Figure 5.17 Left image shows a 70° right wall slant as it might appear through the circular aperture. Note that although it is clearly very slanted relative to frontoparallel, the sign of tilt is ambiguous. The right image shows the same texture, yet has the added cue of converging lines that aid interpretation of tilt.

As preliminary visualisation of the data revealed that response bias was largely independent of tilt the results were transformed into horizontal (V) and vertical (U) components of the surface normal vector, meaning that set and stimulus attitudes could be described as the degree of rotation towards the left or right, ceiling or ground (see Appendix for calculations and Figure 5.18 for coordinate system). Nevertheless, all subsequent data analyses included interaction terms, where V predicts U and vice versa; interactions are discussed in detail in the following sections. Tilt reversals are still referred to in their original coordinate system (±180) though

analysed as a function of ceiling/ground plane slants and left/right wall slants.

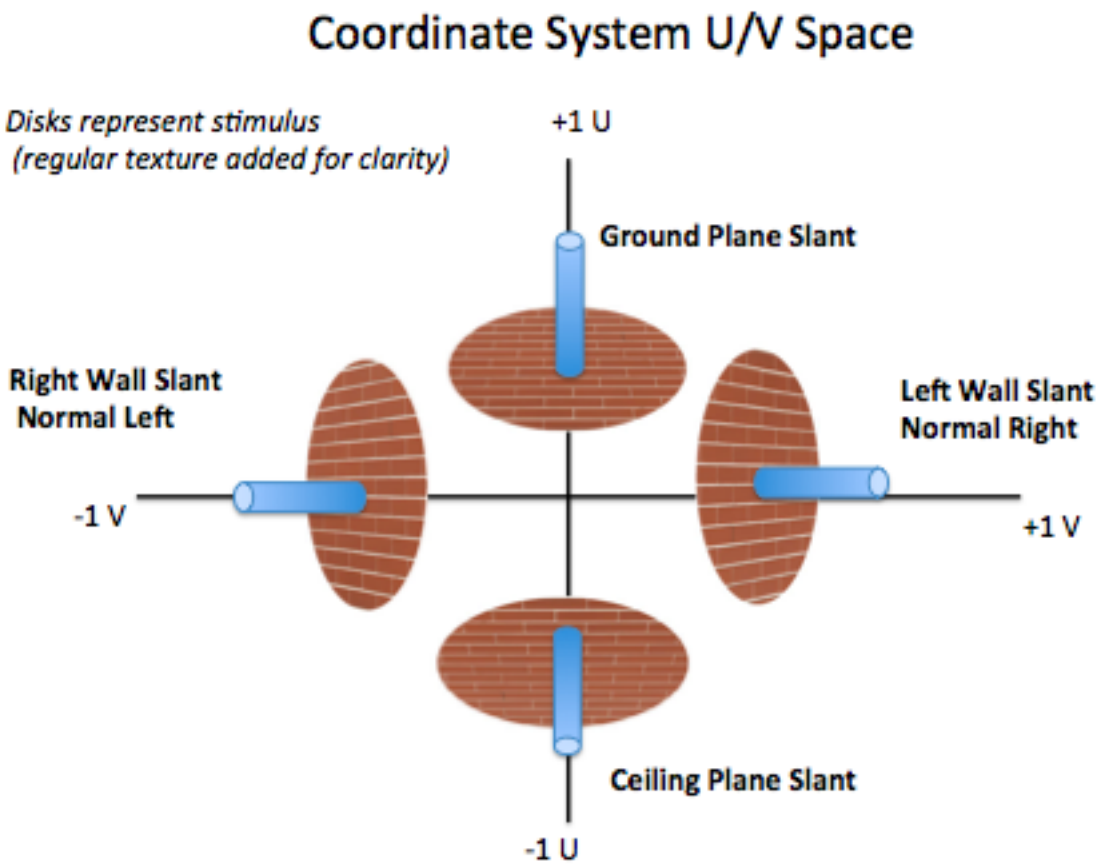


Figure 5.18. Coordinate system for UV space, as used in analysis of perceived surface attitude.

5.6.2. Calculation of Perceived Attitude from Set Attitude

In order to calculate perceived attitude in Part 2, it was first necessary to determine a function to relate perceived attitude to set attitude from Part 1. Since we assumed that perception was veridical in Part 1, then any deviations in attitude settings from stimulus attitude were assumed to be due to haptic response bias. Thus, the function determined using the data from Part 1 could be used to estimate perceived attitude in Part 2, by accounting for haptic response biases.

5.6.3. Haptic response bias in Part 1: Full-cue condition

Analysis of haptic response bias was conducted using stimulus U and/or stimulus V (independent variables; IV) as a predictor of set U or set V (dependent variables). Regression analyses up to 3rd order polynomial, with one IV (U or V) or

two IVs (U and V) were carried out on data from Part 1 for each participant. Leave-one-out cross-validation analysis was performed on each model to control for Type III error (over-fitting). Cross validation revealed that a two IV, cubic model resulted in the lowest standard mean error when averaged across participants, and thus provided the best fit for both U and V (Model 1, U: R^2 = .93; Model 2, V: R^2 = .94). Equation parameters for fitting the data were saved for each participant, in order to calculate perceived attitude in Part 2. Figure 5.19 shows set attitude as a function of stimulus attitude (perceived attitude) for one participant. In addition to plotting the data as a function of a single IV (U or V), set attitude is shown as surface plots to illustrate set attitude as a function of both stimulus U and V.

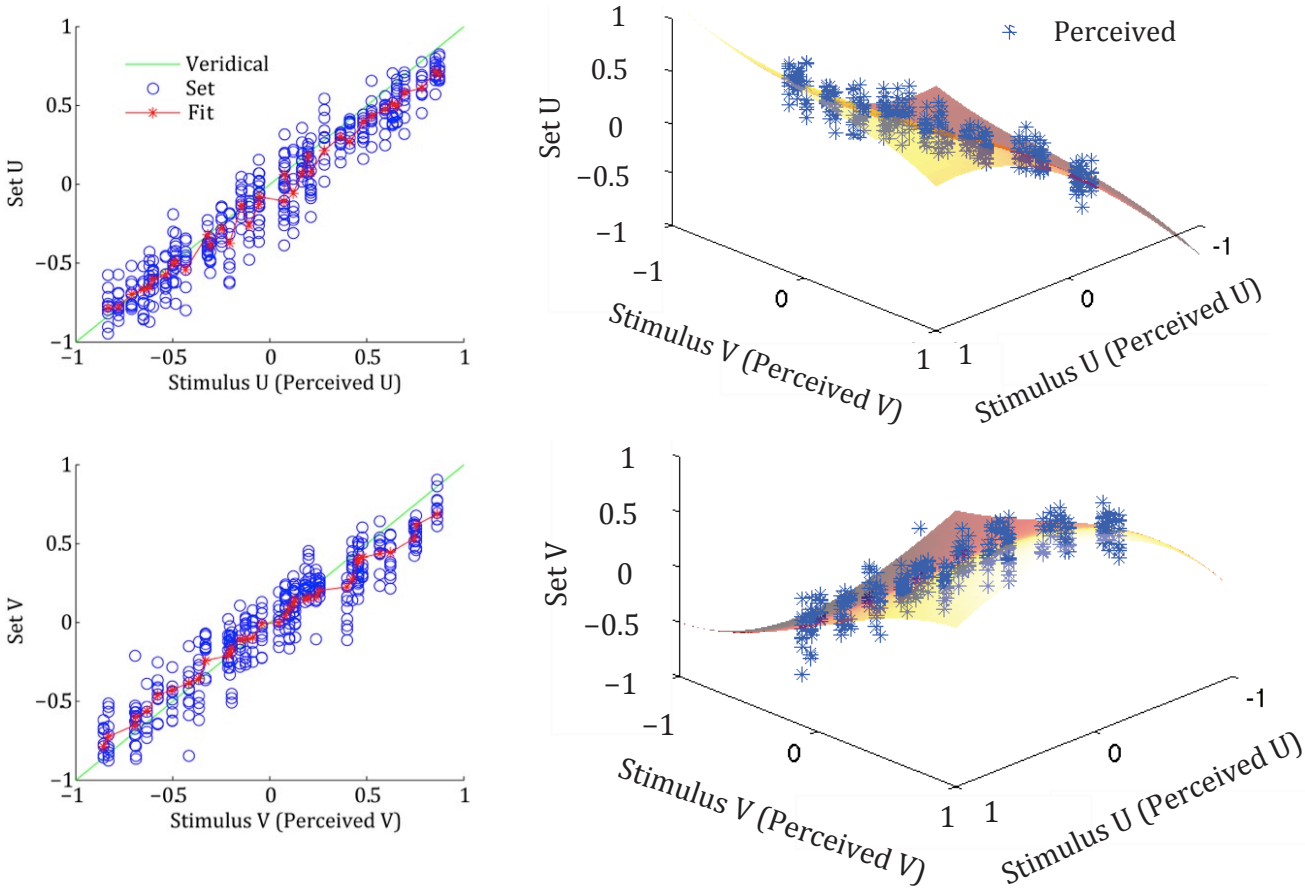


Figure 5.19 Figure showing data from Part 1 (full-cue condition) for participant CM. Set attitude is shown as a function of stimulus U or V (left), and as a function of both stimulus U and V (right). The 2 IV cubic polynomial regression fits are shown in red on left hand plots; the fits are represented by the 3D surface in right hand plots (changes in surface colour from light yellow to dark red represent changes in perceived attitude from positive to negative set U or set V).

The example plots in Figure 5.19 show that though there was a great deal of variance in settings, there appeared to be relatively little bias, as data fall roughly around a slope of +1. The curve of the fit itself deviates from a smooth line or curve in Figure 5.19 (left) as the data and fit are plotted as a function of a single variable (U or V), though set attitude is dependent upon both U and V. The surface plots in Figure 5.19 (right) show the regression fits as a smooth surface when plotted as a function of U and V, and this appears to be a good fit to the data for this participant.

The intercepts for both models were examined relative to zero (no response bias) in order to determine whether observers tended to rotate the paddle more to the left/right wall, ceiling/ground plane. Previous studies using haptic measures to record slant perception have done so about a single (horizontal) tilt axis and they report differences in response bias as a function of arm posture (Li & Durgin, 2010) and wrist flexion (Durgin, Li & Hajnal, 2010). Though response bias will be controlled for in the data from Part 2, it is still of interest to examine haptic response biases as they will expand on previous analysis of the accuracy of haptic settings to measure slant perception (e.g. Taylor-Covill & Eves, 2013).

Analysis of the intercepts for both U and V revealed no bias for participants to set attitudes closer to the ground/ceiling but observers did tend to rotate the paddle slightly to the right (i.e. towards a left wall slant) when perceived leftward/rightward slant was zero (U: $t(8) = .23$, $p = .82$; V: $t(8) = 2.32$, $p = .049$, from one sample t-tests). This is not surprising, as participants used the right hand to set the haptic paddle, and there is greater flexion in the wrist to rotate rightwards (towards a left wall slant) than leftwards; the results indicate that the hand was slightly rotated when observers thought the hand was not slanted about the vertical axis.

The slope of the data were analysed in order to determine whether there were any substantial biases in haptic responses. For instance, if participants just used the wrist to move the paddle they might not be able to reach the maximum and minimum rotations about Axes 1 or 2 and thus might scale all responses proportionately. In order to examine the slope of the data, it was first necessary to conduct a linear regression, in which set attitude (U or V) is predicted as a function of perceived attitude (U or V). For U, the slope was *significantly* shallower than 1 ($m = .88$) and for V the slope was *slightly* shallower than 1 ($m = .95$) (U: $t(8) = 2.80$, $p = .023$; V: $t(8) = 1.63$, $p = .14$, from one-sample t-tests). This indicates that observers tended to

overestimate the attitude of their haptic paddle settings, meaning their hand positions were under-rotated, particularly for ceiling/ground plane slants.

Finally, Figure 5.20 shows the percentage of total trials in which tilt was reversed in Part 1, as a function of stimulus slant and tilt. In this condition, we assumed perception would be veridical as observers had full cues to slant, and thus any tilt reversals are assumed to be due to response error. There are still some tilt reversals in this condition, but these occur at low stimulus slant, where only small changes in rotation about U or V are necessary to change the set tilt (see Figure 5.11; Norman et al., 2006).

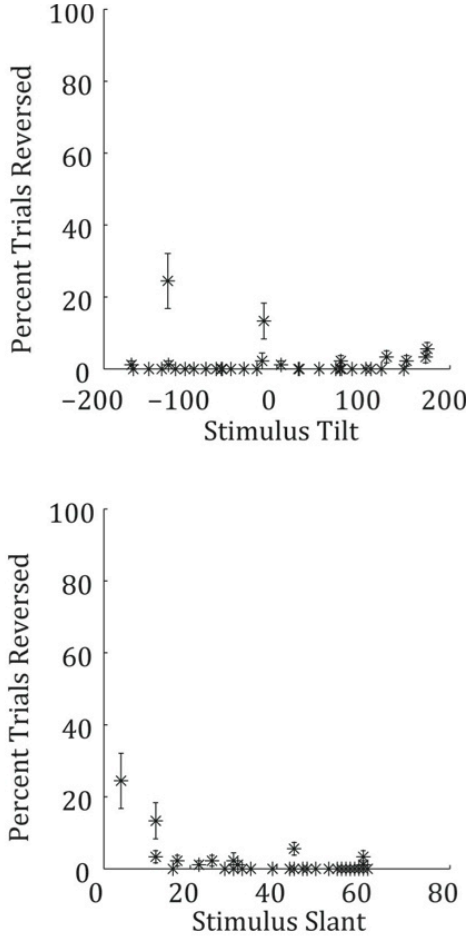


Figure 5.20. Part 1: Full Cue Condition. Percentage of trials in which tilt was reversed. Each data point represents the average for each stimulus attitude, across 10 repetitions and 9 participants. Data are plotted as a function of tilt (top row) and slant (bottom row). Error bars represent ±1 S.E.M.

5.6.4. Perceived attitude in Part 2: Limited cue condition

For each participant, two fine-grained look-up tables were generated in which set attitude was determined as a function of perceived attitude, using the 2IV cubic

regression fits to the data from Part 1 (Models 1 and 2). Matrix 1 contained values of set U, given perceived U and V, while Matrix 2 contained values of set V, given perceived U and V.

For each data point from Part 2, set U and set V were used to estimate perceived U perceived U and perceived V from matrices 1 and 2 respectively. This resulted in 360 perceived attitudes for each participant and each texture. An example data set for one observer is shown as a function of U and V (

Figure 5.21 and Figure 5.22, left), and as a function of U *and* V (

Figure 5.21 and Figure 5.22, right). Average data across all participants are shown in Figure 5.23 and

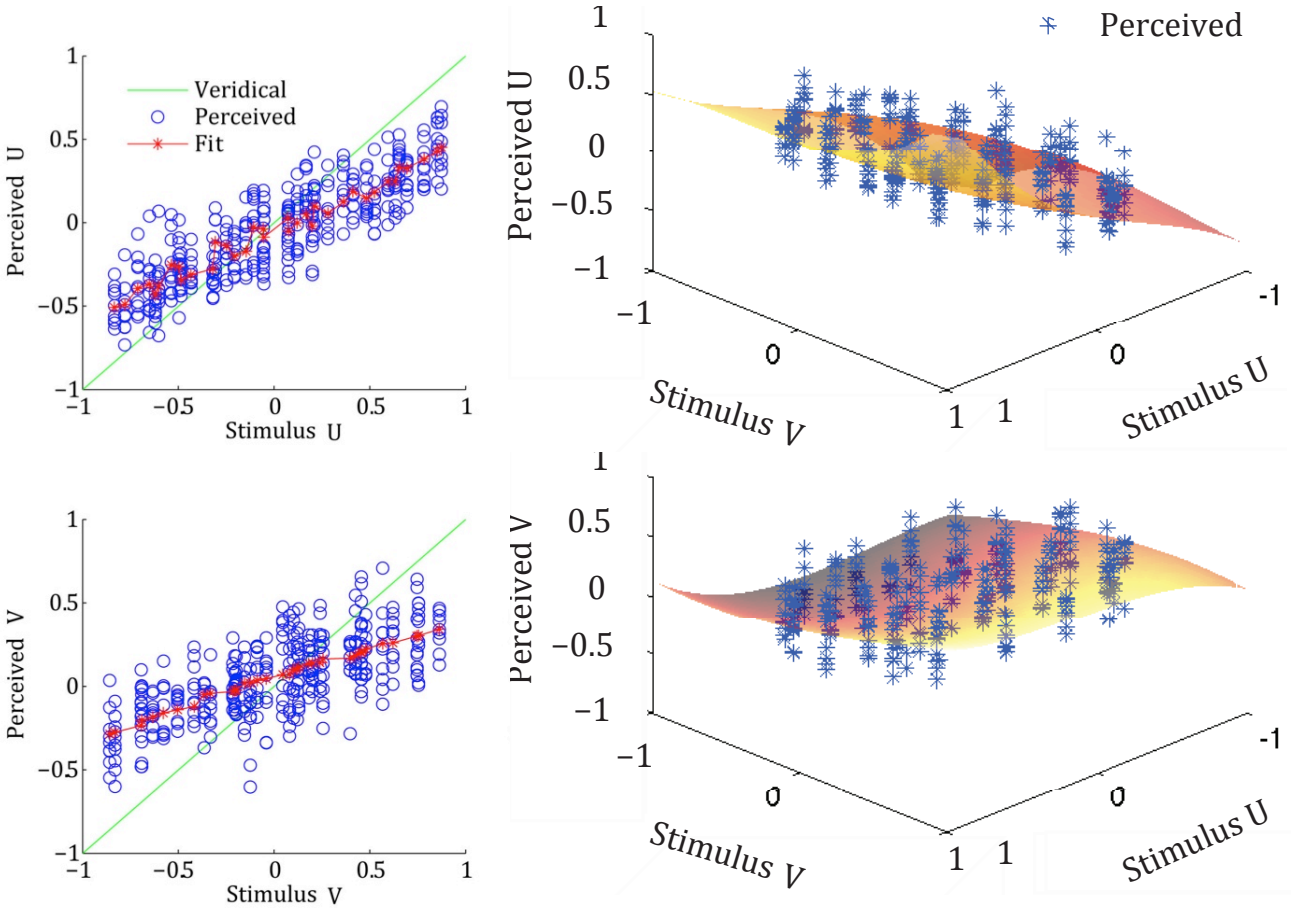


Figure 5.24. Fits to the data are explained below.

Figure 5.21. Part 2, Texture 1: Perceived surface attitudes and 2IV cubic polynomial fit for Participant CM. Left: perceived U as a function of stimulus U and perceived V as a function of stimulus V. Right: perceived U and V as a function of U and V with fits represented by the 3D surface plots (changes in surface colour from light yellow to

dark red represent changes in perceived attitude from positive to negative perceived U or V).

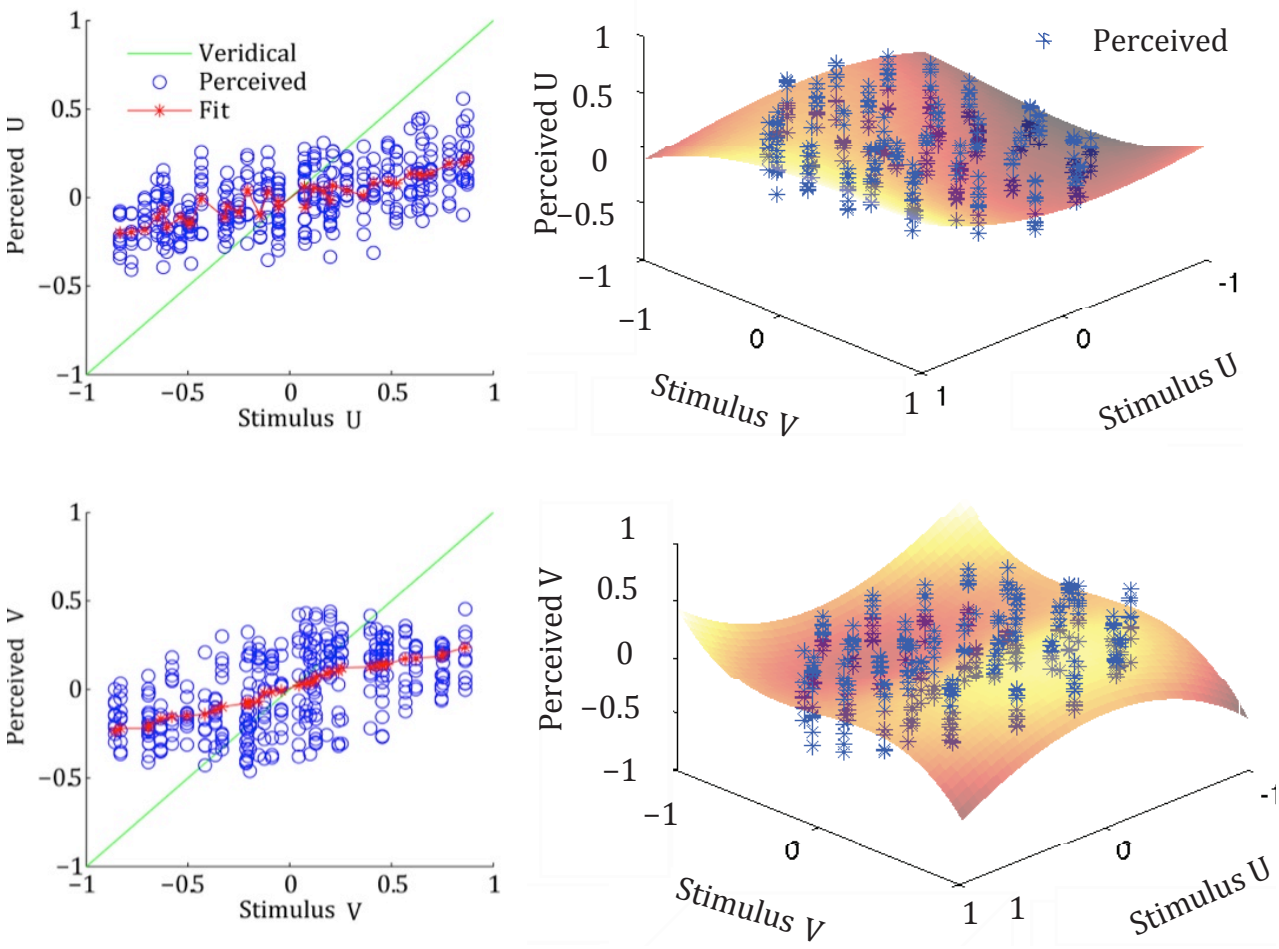


Figure 5.22. Part 2, Texture 2: Perceived surface attitudes and 2IV cubic polynomial fit for Participant CM. Left: Left: perceived U as a function of stimulus U and perceived V as a function of stimulus V. Right: perceived attitude as a function of U and V with fits represented by the 3D surface plots (changes in surface colour from light to dark represent changes in perceived attitude from positive to negative perceived U or V).

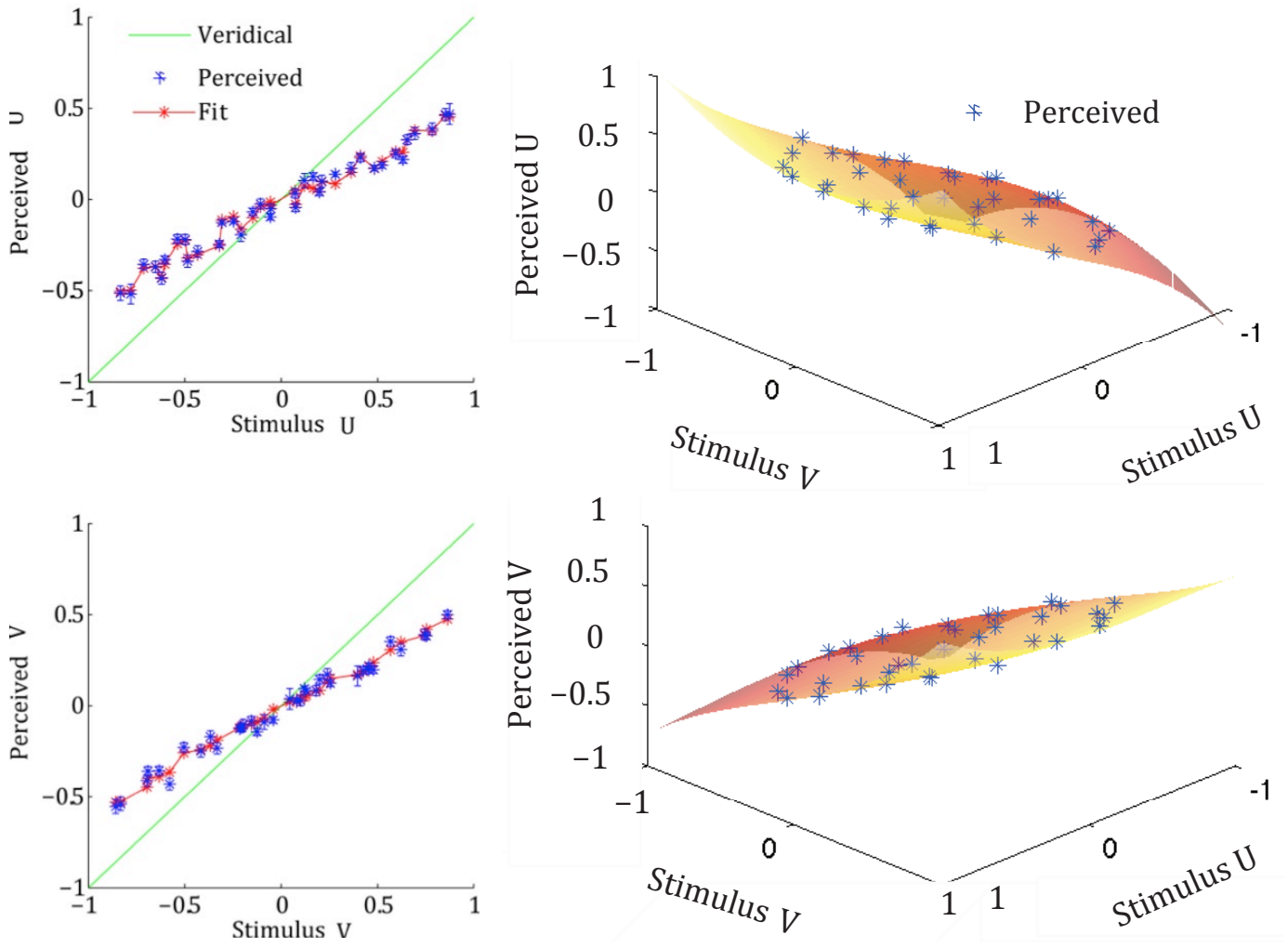


Figure 5.23. Part 2, Texture 1: perceived surface attitudes and 2IV cubic polynomial fit, plotted as an average across all participants (N = 9) and all reps (n=10). Left: perceived U as a function of stimulus U and perceived V as a function of stimulus V. Error bars represent ±1 SEM. Right: perceived attitude as a function of U and V with fits represented by the 3D surface plots (changes in surface colour from light yellow to dark red represent changes in perceived attitude from positive to negative perceived U or V).

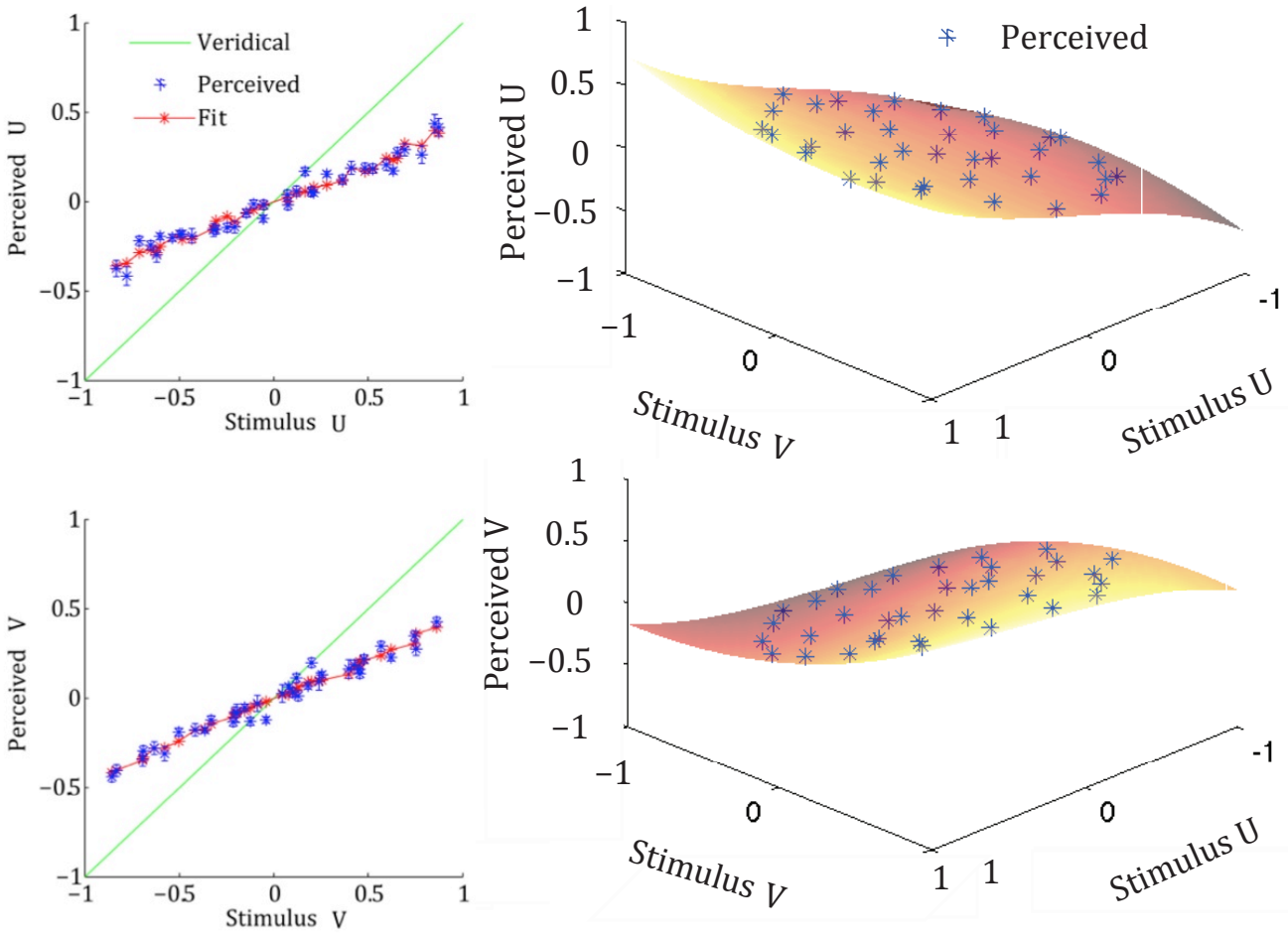


Figure 5.24. Part 2, Texture 2: perceived surface attitudes and 2IV cubic polynomial fit, plotted as an average across all participants (N = 9) and all reps (n=10). Left: perceived U as a function of stimulus U and perceived V as a function of stimulus V. Error bars represent ±1 SEM. Right: perceived attitude as a function of U and V with fits represented by the 3D surface plots (changes in surface colour from light yellow to dark red represent changes in perceived attitude from positive to negative perceived U or V).

5.6.5. Data Analysis

There were three main hypothesised biases in slant that I wished to analyse using this data: 1) a bias towards frontoparallel (zero slant), as has been suggested in previous studies (e.g. van Ee, Adams & Mamassian, 2003; Caudek, Fantoni & Domini, 2011), 2)a bias towards 45°, as might be expected if observers assumed an equal distribution of surface attitudes in the natural environment (see Figure 5.6), and 3) a

bias toward ground plane slants, since these are more prevalent than ceiling plane slants in the natural environment (Yang & Purves, 2003).

Perceptual biases (perceived slant minus stimulus slant) are plotted as a function of stimulus U and V for both Textures 1 and 2 in Figure 5.25 in order to visualise the differences between Texture 1 and Texture 2 models and data. The figures clearly demonstrate an increasing deviation from veridical perception from Texture 1 (high-reliability) to Texture 2 (low reliability). For Texture 2, perceived U is underestimated as a function of stimulus U. This means that when surfaces are slanted towards the ground or ceiling plane, they appear to be closer to frontoparallel. In contrast, for Texture 1 though perceived U is underestimated as a function of stimulus U, bias is also affected by stimulus V, such that bias decreases when the stimulus is highly slanted towards the left or right. This demonstrates that perceived slant about a horizontal axis is biased to a lesser extent when slant about the vertical axis is higher (when the texture is reliable), and perhaps this decrease in bias emerges due to the reliability of the texture being greater at high slants.

As can be seen in Figure 5.25, the interaction between stimulus U and stimulus V is also apparent for perceived V, particularly for Texture 1. For both textures, perceived V is underestimated as a function of stimulus V, though this is much more pronounced for Texture 2. The biases here demonstrate that observers underestimate rotations towards the left and right consistent with a frontoparallel bias.

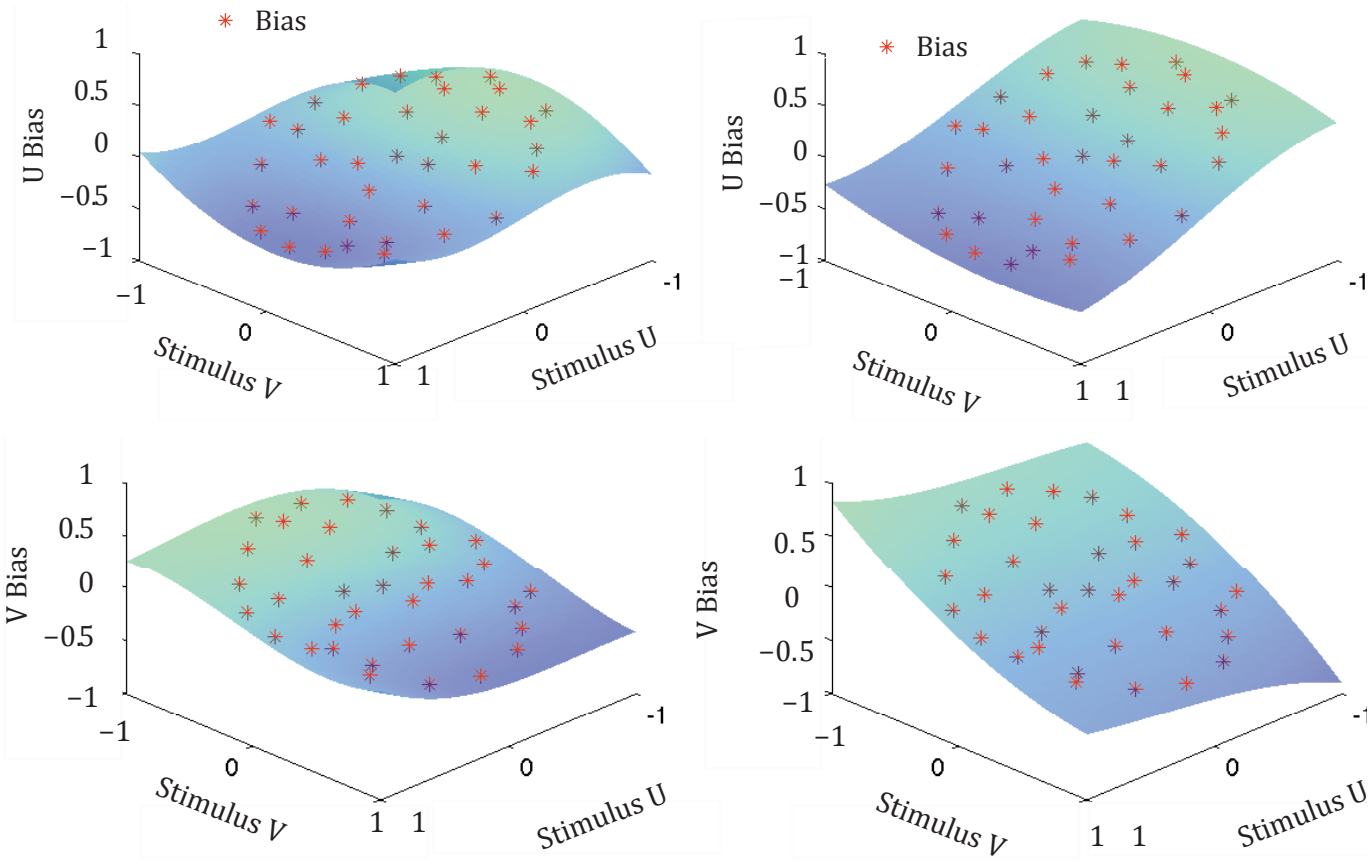


Figure 5.25. Part 2, Texture 1 (left), Texture 2 (right): bias in perceived attitude as a function of U and V with fits represented by the 3D surface plots, plotted as an average across all participants (N = 9) and all reps (n=10). Error bars represent ±1 SEM. Changes in surface colour green to blue represent changes in bias from positive to negative.

Several statistical analyses were conducted on data to address the three hypotheses. Firstly, the data were fit using regression models to determine how perceived U or V can be predicted as a function of stimulus U and/or stimulus V. This allowed examination of the coefficients of the model and analysis of how the influence of stimulus U and V change as stimulus texture becomes less reliable. Intercepts of the regression models were used to examine any influence of a ground/ceiling plane bias. The numbers of tilt reversals from ground/ceiling plane and left/right wall slants were also examined to investigate the potential ground plane bias. Finally, UV data were transformed back into slant (and tilt) coordinates in order to visualise and analyse data with respect to a 45° bias, as this would be difficult to interpret in the U/V coordinate frame.

In order to quantify biases in perception about U and V, the perceived attitude data were fit using regression analyses up to 3rd order, with one IV (stimulus U or V) or two IVs (stimulus U and V), as done for Part 1. Cross validation analysis revealed that a two IV, cubic model resulted in the lowest standard mean error for two of the four sets of data (Texture 1- perceived V, Texture 2- perceived U), and provided a good fit for the remaining two sets, and therefore were accepted as the best model to represent all data (Texture1, U: R^2 = .73; Texture2, U: R^2 = .58; Texture1, V: R^2 = .68; Texture2, V: R^2 = .51). See
Figure 5.21 and Figure 5.22 for example fits for one participant, or Figure 5.23 and Figure 5.24 for average data and fits across all participants; the standardised coefficients from the regressions performed on averaged data are presented in Table 1.

Table 1

Cubic model coefficients (standardised) for perceived U (top) and perceived V (bottom) as predicted by stimulus U and stimulus V for Textures 1 and 2, based on averaged data (N = 9).

	intercept	U	V	U^2	V^2	U*V	U^3	V^3	U*(V^)2	V*(U^2)
Texture 1	0.001	0.566	-0.105	-0.049	-0.049	-0.039	0.314	0.03	0.184	0.027
Texture 2	0.000	0.737	-0.021	0.004	-0.029	-0.051	0.175	0.002	0.075	-0.024

	intercept	V	U	U^2	V^2	U*V	V^3	U^3	U*(V^)2	V*(U^2)
Texture 1	-0.005	0.741	-0.082	-0.033	-0.023	0.004	0.176	0.117	0.051	0.094
Texture 2	-0.003	0.958	-0.118	-0.015	0.001	0.012	0.005	0.08	0.099	-0.043

The intercepts for all models are close to zero, suggesting that perceived attitude was unbiased for frontoparallel surfaces. Perceived U and V appear to be strongly predicted as a linear function of their main IV (e.g. stimulus U predicts perceived U) for Textures 1 and 2, but coefficients for the second main IV (e.g. stimulus V predicting perceived U) are far smaller, suggesting little influence of stimulus V linearly affecting perceived U and vice versa. In order to further investigate the relative contribution of the various predictors, a stepwise regression was carried out to examine which predictors significantly add to the model. An initial check on the data showed that although multicollinearity does exist (predictors are correlated), that the condition index was lower than 15 in each case, thus a stepwise regression was deemed suitable. Each prediction model considered 9 predictors, as outlined in Table 1 (without the constant). For Texture 1, U, the prediction model

contained 7 predictors (in order: U, U*V^2, U^3, V, U*V, U^2, V^2), $F(7,352) = 1128.76$, $p<0.001$. For Texture 2, U, the prediction model contained 5 predictors (U, U*V, U^3, U*V^2, V*U^2), $F(5,354) = 613.66$, $p<0.001$. For Texture 1, V, the prediction model contained 6 predictors (V, U^3, V^3, V*U^2, V^2), $F(6,353) = 1052.12$, $p<0.001$. For Texture 2, V, the prediction model contained three predictors (V, U*V^2 and V*U^2), $F(3,356) = 948.67$, $p<0.001$. This indicates that U and V were, as expected, the highest contributing coefficients when predicting perceived U and V respectively. It is also interesting that in each model the second highest component includes the secondary predictor to some degree (i.e. V has some contribution to perceived U and vice versa). Though this information is somewhat interesting, the method does not take into account any degree of collinearity, thus predictors noted here may be included predominantly due to shared variability (i.e. U and U^2). For this reason, a second method was used to further analyse the data.

As a second investigation into the relative contributions of the various predictors, the data were transformed to structure coefficients; when squared, the transformed coefficients measure how much of the variance of y (perceived U or V) is independently contributed by each coefficient (Kraha et al. 2012). This transformation is important when model coefficients are correlated (when there is a high level of multicollinearity) as may be the case for polynomial regressions. Importantly, it can indicate cases in which the contribution of coefficients has been suppressed, resulting in an artificially low beta. The transformed coefficients are presented in Table 2; significance levels relate to Pearson's correlations between the predictor and observed effects.

Table 2

Squared structure coefficients for perceived U (top) and perceived V (bottom) as predicted by stimulus U and stimulus V for Textures 1 and 2, based on averaged data (N = 9). See text for explanation of structure coefficients.

	Intercept	U	V	U^2	V^2	U*V	U^3	V^3	U*(V^)2	V*(U^2)
Texture 1		0.976**	0.000	0.007	0.010	0.004	0.853**	0.000	0.499**	0.00
Texture 2		0.990**	0.000	0.017	0.01*	0.002	0.867**	0.001	0.445**	0.00

	Intercept	V	U	U^2	V^2	U*V	V^3	U^3	U*(V^)2	V*(U^2)
Texture 1		0.990**	0.010	0.006	0.010	0.002	0.835**	0.006	0.028**	0.479**
Texture 2		0.994**	0.004	0.007	0.009	0.001	0.830**	0.001	0.027**	0.430**

** Pearson correlation is significant at 0.01 level

* Pearson correlation is significant at 0.05 level.

The structure coefficients for predicting perceived U that were significant were U, U^3 and U*(V^2), while V, V^3, U*(V^2) and V*(U^2) were significant components of the model for predicting perceived V. Where cubic terms are significant it reflects the fact that slant is underestimated more at low slants than high slants, and this is due to the fact that texture cues are unreliable at slants close to zero. It remains difficult to fully interpret the unique contribution of individual structure coefficients as although this method should prevent suppression of predictors by other predictors, it does not account for shared variance between coefficients, thus some of the contributions may be exaggerated.

A statistical analysis of the intercept for each (2IV, cubic) regression model was conducted in order to determine whether responses were significantly biased towards ceiling/ground or left/right wall slants (though there is no reason to predict perceptual bias towards the left or right). One would only predict a significantly positive or negative intercept for perceived U if observers had a prior for ground plane slants, as otherwise one would just predict veridical perception or a bias towards zero slant. One-sample t-tests revealed no significant differences from zero, meaning the intercepts show no indication of a bias towards either ground/ceiling plane or left/right wall slants (Texture1, U: $t(8) = .06$, $p = .96$; Texture2, U: $t(8) = .68$ $p = .52$; Texture1, V: $t(8) = .01$ $p = .99$; Texture2, V: $t(8) = .52$ $p = .62$).

Though it is possible to investigate individual components of the model fits between Textures 1 and 2, this does not allow a direct comparison of the slope of the data as a whole. Thus, in order to determine whether perceptual bias increased as cues to surface attitude became less reliable (as one would predict from the influence of a prior), it was necessary to examine the mean, linear slopes of data; if observers demonstrated a frontoparallel prior it would present as a decrease in slope from Texture 1 relative to Texture 2. Linear regression analysis revealed that the slope for Texture1 U was significantly steeper than for Texture 2 U ($m = .53$; $m = .42$; $t(8) = 3.65$, $p = .007$; adjusted alpha = .025) and the slope for Texture1 V was significantly steeper than for Texture2 V ($m = .56$; $m = .46$; $t(8) = 3.25$, $p = .012$, from paired samples t-tests; adjusted alpha = .025). This demonstrates that participants were utilising a perceptual bias that was more influential when slant cues were less reliable and this could be consistent with a frontoparallel prior.

In order to focus on investigating a prior for 45° slant, the perceived UV data were transformed back into perceived slant and tilt. Slant bias, averaged across all participants is plotted in Figure 5.26 as a function of stimulus slant. Slant bias around zero stimulus slant is artificially high due to an artefact of slant measurement (recorded slants were always positive, thus a slant of 5° to the left was considered the same as a slant of 5° to the right; here, the calculated difference between the two slants would be zero, though the absolute difference is 10°). Nevertheless, firstly, the figures show systematic underestimation of slant, which is consistent with a frontoparallel prior, and secondly there is greater underestimation at larger slants, consistent with a weighted average of the stimulus slant and frontoparallel prior. Linear and quadratic regressions were computed using stimulus slant as a predictor of perceived slant and cross-validation confirmed that a quadratic model provided the best fit to the data (as shown in Figure 5.26; Texture 1: R^2 = .54; Texture 2: R^2= .66), thus confirming the weighted influence of the prior as function of stimulus slant.

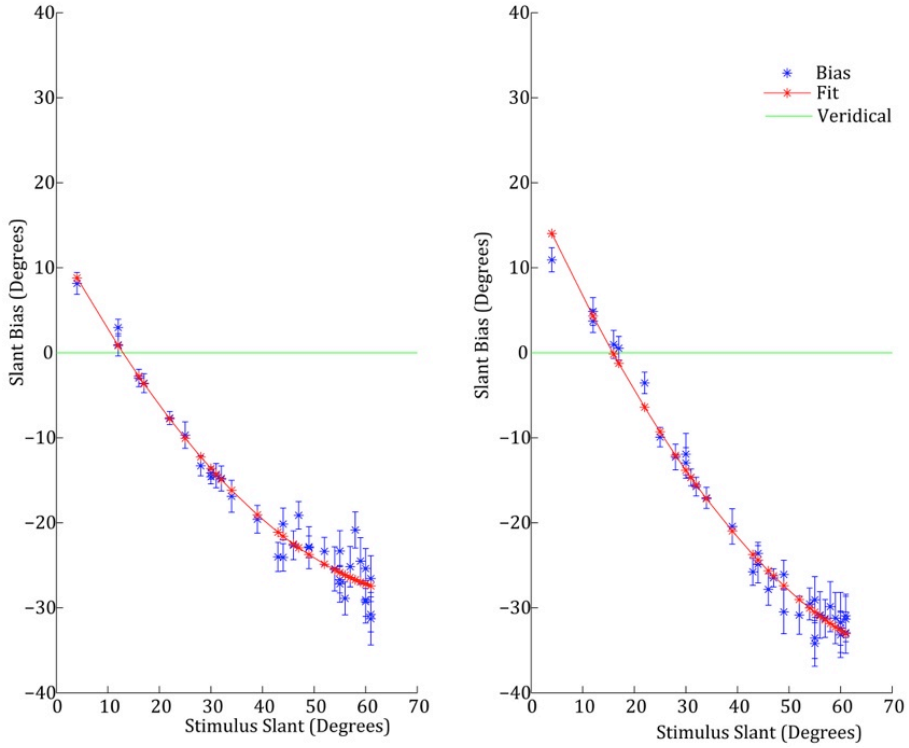


Figure 5.26. Average bias in perceived slant for Part 2 (limited cue condition; *N* = 9). Left: Texture 1, right: Texture 2. Quadratic fit to averaged data shown in red. Error bars represent ±1 SEM.

The analysis of ground versus ceiling plane bias can be elaborated on by examining the percentage of trials in which tilt was reversed, in the original, untransformed data; tilt reversals are presented as a function of stimulus tilt and stimulus slant in Figure 5.27. Contrary to expectations, it would appear that for Texture 1, there were more trials in which perceived tilt was reversed for ground plane stimulus slants (- 90 <tilt <90), consistent with a ceiling plane bias. In contrast, the spread of the number of tilt reversals across ceiling or ground plane for Texture 2 appears to be more isotropic. In addition, for both Texture 1 and 2, the percentage of trials in which tilt was reversed was lower for high slants, as texture becomes a more reliable cue to surface attitude.

Analysis of 95% confidence intervals and *p* values for the repeated measures differences between left/right and ceiling/ground tilt reversals were obtained via a standard bootstrapping procedure (using a method of random sampling with replacement). For Texture 1 there were more trials in which the tilt of ground-plane surfaces were reversed (surfaces were perceived to be slanted from the ceiling plane)

than ceiling-plane reversals (M=15.72, 95% CI [7.36, 24.24], *p*=.0004). In addition, there were more trials in which rightward slants were reversed (i.e. trials in which a left wall slant were shown but were perceived as right wall slants) than leftward slants (M=-10.06, 95 CI [-19.22 -0.833], *p*=.031). For Texture 2, the number of trials in which tilt was reversed was equal for ceiling and ground plane slants (*M*=-3.5, 95% CI [-9.36 2.39], *p*=.241), though there were more trials in which rightward slants were reversed, relative to leftward slants (*M*=-8.174, 95 CI [-13.00 -3.22], *p*=.002).

These results were surprising, as there is no obvious reason to predict a bias towards ceiling slants, if the distribution of surfaces in the natural environment tends to be dominated by the ground plane (Yang & Purves, 2003). Indeed, we also predicted no perceptual bias towards left or rightward slant (towards right wall or left wall percepts), though this was found for both textures.

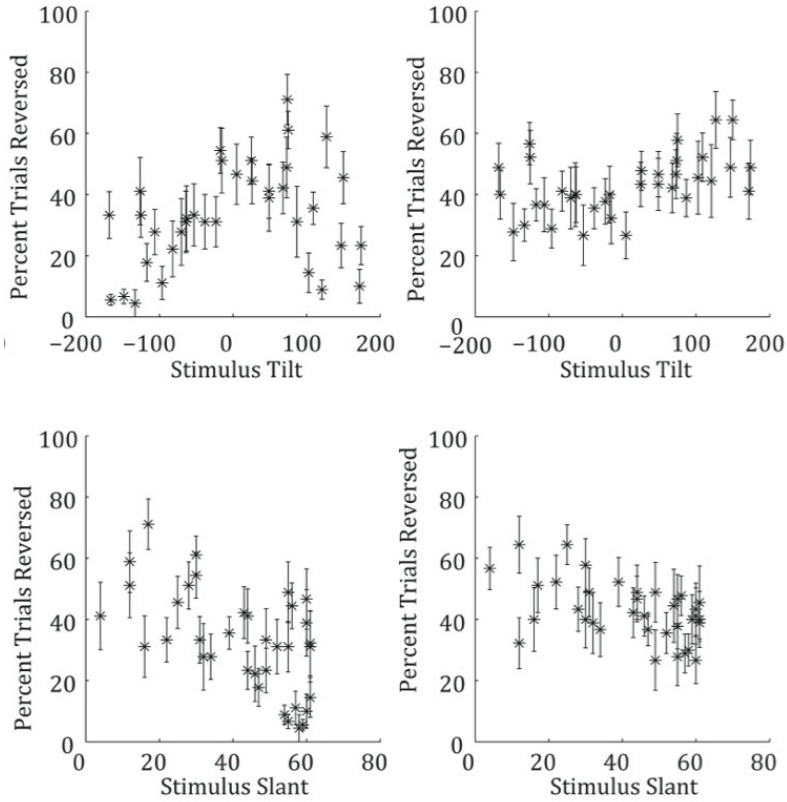


Figure 5.27. Percentage of tilt reversals for Part 2: Limited Cue Condition; Texture 1 (left) and Texture 2 (right). Each data point represents the average for each stimulus attitude, across 10 repetitions and 9 participants. Data are plotted as a function of tilt (top row) and slant (bottom row). Error bars represent ±1 S.E.M.

5.7. Discussion

The aim of the current study was to examine three potential biases in slant perception: 1) a frontoparallel bias 2) a bias towards 45° slant, and 3) a bias towards ground plane slants. Participants were asked to use a haptic paddle to match slanted stimuli about multiple tilt axes, thus it was possible to analyse perception at a wide range of stimulus attitudes. In addition, haptic response error was taken into account in Part 1 of the experiment (full-cue condition), where perception was assumed to be veridical, thus allowing biases in slant perception to be measured without being confounded by haptic biases.

5.7.1. Do observers have a prior for slant?

Regression analysis using data from Part 2 revealed that a 2IV cubic model provided the best fit for predicting perceived U as a function of stimulus U and V. This demonstrates that perceived slant about a ceiling/ground plane is influenced by stimulus slant about a left/right wall, thus the two axes are not independent. Statistical analysis also revealed that for stimulus U as a predictor of perceived U, and stimulus V as a predictor of perceived V, the slope of a linear regression was significantly shallower for Texture 2 (low-reliability) than Texture 1. This demonstrates that a prior for slant had more influence when the stimulus slant was more ambiguous. Indeed, the result was consistent with slant underestimation for both U (ceiling/ground) and V (left/right) components of surface attitude.

Analysis of intercepts for the cubic regression models found no significant difference from zero for U or V, suggesting that observers had no bias towards ceiling/ground or left/right wall when the stimulus was not slanted about those axes. This finding is consistent with a frontoparallel prior, or prior for 45° slant, but *not* a ground plane bias. Furthermore, after converting U and V back into slant coordinates, plots of data revealed a systematic underestimation of slant and this bias increased as a function of stimulus slant, again, consistent with a frontoparallel prior. Crucially, this effect contradicts the predicted prior for 45° slant, as the results would show no bias at 0° *and* 45° if this prior were true. Furthermore, the frontoparallel prior is confirmed by the fact that observers exhibited negative biases (slant underestimation) for both stimulus U and V in the transformed data.

The results for tilt reversals contradicted the prediction that observers should have a prior towards ground plane slants when tilt is ambiguous. In fact, observers showed a significant bias towards perceiving ceiling plane slants than ground plane slants for Texture 1, despite slants about the ground plane being more prevalent in natural scenes (Yang & Purves, 2003). Since there were also a significant bias for observers to perceive right wall slants relative to left wall slants (as seen by the tilt reversal data) then it might suggest that these biases are an artefact of the task.

One potential reason for these unexpected biases could be due to the way that the stimulus was lit. In the current study the stimulus was uniformly lit in order to prevent cues to surface attitude from luminance changes; the brightness of a surface patch is proportional to the angle between the surface normal and the primary direction of illumination. In the natural environment surfaces are often not uniformly lit, and observers use a prior consistent with knowledge that there is a single light source overhead (e.g. Adams, 2007), thus participants might have assumed zero slant since the uniform brightness of the surface implied that it was frontoparallel (this would occur if the light source was centred in front of the surface, rather than above). It is, however, unclear why participants would exhibit a bias towards perceiving surfaces rotated towards the left, rather than the right, unless they assumed a light source that was on the left of the surface; indeed, a prior for a light source above and to the left has been argued by researchers investigating the light from above prior (see Mamassian & Goutcher, 2001).

In sum, the collection of results outlined here describe a frontoparallel prior, but show little or no evidence for a ground plane bias or prior for 45° slant. In the introduction I explained how a frontoparallel surface takes up the greatest proportion of the visual field compared to the same surface when slanted, yet in an environment where there are multiple tilt axes (and surface slant is equally distributed) then actually a 45° slanted surface is the most probable. Crucially, the fact that observers exhibit a frontoparallel prior implies that observers are assuming a single tilt axis, in which case a frontoparallel prior is logical. Thus, the demonstrated frontoparallel prior implies that slant is coded separately for different tilts.

5.7.2. Haptic response biases

Analysis of data from Part 1 of the experiment revealed that participants had a slight bias to rotate the haptic paddle more towards the right, and also to under-

rotate the paddle. Though such biases were controlled for in this study, this clearly demonstrates that it is necessary to control or account for haptic response biases in measurements of slant perception, in order to prevent artificially low estimates of perceived slant.

Furthermore, care should be taken when interpreting results about slants close to zero (as also noted in Taylor-Covill & Eves, 2013), particularly since the direction of slant of the haptic paddle changes with only slight rotations of the paddle close to frontoparallel. Researchers wishing to use haptic devices to measure true biases in slant perception would be advised to use a condition similar to Part 1 of the current study, in which haptic response biases were isolated due to veridical perception.

5.7.3. Benefits and Limitations of the Method

Though previous studies have also investigated slant perception using real surfaces (thus avoiding confounding cues to zero slant from frontoparallel monitors; Watts, Akeley, Ernst & Banks, 2005), though as they only investigated slant perception about a single tilt axis, then a frontoparallel bias might still have been a confounding factor (e.g. Porrill, Duke, Taroyan, Frisby & Buckley, 2010; Todd, Thaler & Dijkstra, 2005). The current study avoided this bias by using multiple tilt axes, though observers still demonstrated bias consistent with a frontoparallel prior.

One of the benefits of our method was that the Pan-Tilt unit used to rotate the stimuli was remotely programmed and allowed rapid changes in surface attitude that would not be possible using manual methods (e.g. Porrill et al., 2010; Taylor-Covill & Eves, 2013). Nevertheless, it was necessary to set up every aspect of the experiment with extreme precision in order to avoid introducing bias, such as offsets in the angle of the 45°mirror, the slant of the wall upon which the Pan-Tilt unit was mounted, and the table upon which the haptic paddle rested.

5.7.4. Further research

As mentioned above, a bias towards perceiving surfaces to be slanted towards the ceiling plane and towards a right wall slant was unexpected and further investigation could be conducted to examine this further, for instance, there may be a correlation between the observer's light from above prior and the number of trials in which tilt was reversed from left to right.

Furthermore, future research will be able determine the discriminability of surface attitudes for Textures 1 and 2 in order to separate the effects of internal noise from perceptual bias (see Stocker & Simoncelli, 2006; Girshick, Landy & Simoncelli, 2011) and thus determine the shape of the prior.

Chapter 6

General Discussion

The objective of this thesis was to examine the hypothesis that vision is informed by the statistics of the natural environment. Since the topic allows for a wide variety of experimental and observational approaches, the empirical studies focused on psychophysical methods in order to study human perceptual biases, though the experiments were influenced by knowledge and theoretical insights from image and scene statistics, and neural encoding in human vision.

Chapters were varied in their focus, but all aimed to investigate known, or hypothesised perceptual biases and examine how these might be linked to the natural environment. There were several key outcomes: 1) I found no evidence to support the theory that image amplitude spectra can be diagnostic of scene category; 2) I demonstrated that human observers can infer 3D surface shape from 2D contour curvature, and provide the first evidence that perceived metric depth is affected by curvature magnitude; 3) I developed a novel stimulus and response set-up that allowed accurate recording of observers' perception of surface slant about multiple tilt axes; and 4) I confirmed that observers demonstrate a frontoparallel prior.

The studies presented in Chapters 2 and 3 demonstrate that although the visual system might be sensitive to the statistics of natural images, this sensitivity is task-dependent. In the case of amplitude information, amplitude-phase interaction can increase the salience of naturalistic and manmade contours, but amplitude alone cannot perceptually or conceptually facilitate dominance of amplitude-matched images in subsequent binocular rivalry. The findings relating to contour curvature and perceived metric depth (Chapter 4) in particular highlight the need for further examination of scene statistics, both in order to ecologically validate the findings that contour curvature can provide a cue metric depth, and also to investigate potential perceptual biases that have yet to be examined. Furthermore, the findings from Chapter 5 highlight the need to study perceptual biases with respect to our knowledge about the natural environment and, in this case, geometrical theory. For instance, measuring slant perception about multiple axes revealed an interaction, whereby perception of stimulus slants about a horizontal axis were affected by the

stimulus slants about the vertical axis, and vice versa. This implies that examination of only one axis does not provide enough information to measure bias accurately.

6.1. Findings, Implications and Limitations

6.1.1. Chapters 2 and 3

The aim of Chapter 2 was to examine the hypothesis that unlocalised amplitude information could be diagnostic of scene category (e.g. Guyader et al., 2005; Oliva & Torralba, 2001). This theory stemmed from previous research showing that there are characteristic differences in amplitude spectra between image categories (Oliva & Torralba, 2001, 2003), meaning that these image statistics might have been internalised by human observers. Thus, in Chapter 2 I used a method of conceptual priming to examine whether exposure to a natural or manmade amplitude caricature (containing the averaged amplitude across multiple images) would bias subsequent categorisation of a hybrid test image (where amplitude was equalised and phase morphed across a single natural and manmade image). Though I predicted a priming effect (based on previous research, Guyader et al., 2005), all prime test durations resulted in suppression, rather than facilitation and thus it was difficult to isolate any unique effects of prime amplitude without confounding effects of amplitude-phase interaction. Indeed, follow-up experiments demonstrated that image-specific amplitude spectra did bias scene categorisation of hybrid (phase-morphed) images, more so than scene averaged amplitude spectra, indicating that amplitude-phase interaction is important for recognition of diagnostic naturalistic and manmade contours.

Chapter 3 extended the findings from Chapter 2 by using an alternative design, with which facilitation and suppression in binocular rivalry has been shown previously using simple grating stimuli (Pearson, Clifford & Tong, 2008; Brascamp et al., 2007). The study aimed to examine conceptual priming in rivalry in order to question whether the visual system associates characteristic amplitude spectra with specific scene categories (e.g. forest or city). Since the result showed no pattern of facilitation to suppression when corresponding test and prime were different exemplars from the same image category, it was concluded that the visual system does not associate image amplitude spectra with specific scene categories. Indeed, further examination revealed that perceptual facilitation was possible when test and

prime were identical amplitude images, but there was no conceptual facilitation when test and prime contained the same amplitude spectra but different phase information.

Though the studies from Chapters 2 and 3 demonstrate that unlocalised amplitude information is not diagnostic of scene category, there is still evidence that the visual system is sensitive to the amplitude spectrum of natural images. For instance, natural ($1/f$) amplitudes dominate in binocular rivalry (Baker & Graf, 2009) and $1/f$ surrounds maximally suppress central images relative to surrounds with other spectral slopes (MacDonald & Tadmor, 2006), while the visual system itself has peak contrast sensitivity for low spatial frequencies (e.g. Banks & Salapatek, 1978), which is perhaps related to the predominance of low spatial frequencies in natural images (e.g. Field, 1987). Nevertheless, these results imply that a higher level of processing is necessary in order to recognise scenes by their component parts, as amplitude must be localised in order to convey scene category (e.g. Loschky & Larson, 2008; Joubert et al., 2009). Perhaps it is diagnostic information from localised amplitude (phase) that drives rapid scene categorisation, as complex cells within the visual system have been shown to respond specifically to features within phase-intact natural scenes, but they were insensitive to the same features presented within phase randomised scenes (Felsen, Touryan, Han & Dan, 2005).

The results from Chapter 2, experiments 3a and b indicated that diagnostic naturalistic or manmade contours (made more salient by amplitude-phase interaction) might play an important role in rapid scene categorisation. Perhaps further study of local phase and diagnostic structures within images might aid our understanding of rapid scene categorisation, though this would require a higher-level focus that digresses from the study of the more low-level information that is investigated in relation to prior knowledge and the natural statistics approach.

6.1.2. Chapter 4

Chapter 4 investigated the hypothesis that 2D contour curvature could provide a cue to metric depth near to the boundary of ambiguous shapes. Previous behavioural research had shown that observers could infer 3D shape from object outlines or silhouettes (e.g. Wagemans et al., 2008), and that the proximity from the bounding contour can influence judgements of metric depth (Norman & Raines, 2002). In the experiment reported in Chapter 4, participants were shown

stereoscopic 'potato' stimuli with limited cues to depth and asked to adjust the depth of a disparity-defined probe dot until it rested on the surface of the potato, close to a convex, concave or flat contour.

The results showed a significant main effect of the magnitude of contour curvature on perceived depth, thus demonstrating that the shape of the 2D bounding contour did affect perceived 3D shape. The implication of this finding is that contour curvature might still be a cue to metric depth when other cues such as lighting or texture provide ambiguous cues to 3D shape. The results could be used to aid models of depth propagation within an object or to expand on knowledge about why some 2D shapes look flat, and some 3D (see Tse, 2002).

One of the main limitations of the study was that the task itself was exceptionally hard, as noted by participants, and the fact that several observers could not fuse the stereoscopic depth probe. Another limitation of this study is that there is huge scope for expansion of the design to include a wider range of contour curvatures, partly to determine whether there are subtle differences in perceived depth between different contour magnitudes, and to further examine the effect of contour sign on perceived depth. Furthermore, behavioural studies would ideally be informed by our knowledge about the environment itself. Although the sign of contour curvature has been used when investigating the relationship between range and 2D images (thus explaining figure-ground relationships, Burge 2010), analysis of range images and depth has not yet been done.

6.1.3. Chapter 5

Here, I investigated a known bias that observers tend to underestimate slant (e.g. Andersen, Braunstein & Saidpour, 1998; Proffitt, Bhalla, Gossweiler & Midgett, 1995). There were three main hypotheses: 1) observers might underestimate slant, consistent with a frontoparallel prior, 2) observers might demonstrate a bias towards 45 degrees slant and 3) observers might demonstrate a bias towards ground plane slants as the stimulus becomes more ambiguous (e.g. at low slants).

I wanted to frame the known bias in terms of known and theorised anisotropies in the natural environment, in order to guide experimentation and interpret results. For this reason, it was necessary to design a study in which we could examine observer biases in slant about multiple tilt axes, since 1) observers

should have a frontoparallel bias if there is only one tilt axes; though 2) surfaces in the real environment can be slanted about different tilt axes since the distribution of surface attitudes in the environment is anisotropic (e.g. Yang & Purves, 2003). The results showed perceptual biases consistent with a frontoparallel prior. This implies that observers used a prior that was separately coded for a single tilt axis, rather than one that takes into account the distribution of all potential surface attitudes. An unexpected finding was that observers demonstrated a bias for perceiving slants about the ceiling plane, and right wall slants. Further investigation would be necessary to examine this discrepancy.

A follow-up of this experiment will be conducted using the same stimuli; the results from the limited cue condition demonstrated that observers exhibit a frontoparallel prior. In order to determine the shape of the prior it would be necessary to measure perceptual discrimination of surface attitude in order to separate the effects of internal noise from perceptual bias (see Stocker & Simoncelli, 2006; Girshick, Landy & Simoncelli, 2011).

6.2. Future Directions

6.2.1. Perceptual biases

Though one way of looking at perceptual biases is to examine those that have already been observed, for the natural statistics approach it makes more sense to let the distribution of image and scene statistics drive our search for perceptual biases. Perhaps if researchers further examined existing databases of scene statistics it might highlight potential biases that have yet to be explored (e.g. Potetz & Lee, 2003; Yang & Purves, 2003). Such a database will soon be compiled by the Adams and Graf lab.

Alternatively, if one considered that perceptual biases are a result of our interaction with the environment and the tasks that we carry out, then perhaps examination of regular behaviours might guide us in finding perceptual priors (Purves et al., 2001). Indeed, Geisler and Ringach (2009) argue that it is important to use naturalistic tasks when studying perceptual biases. A study that examined biases in orientation perception supports this argument, as the researchers found that observers were more sensitive to detecting oblique stimuli than horizontal or vertical stimuli within naturalistic images, and this contradicts knowledge that humans are

less sensitive to oblique orientations when the stimuli are simple gratings (Essock, DeFord, Hansen & Sinai, 2003).

6.2.2. Image statistics

Since researchers use scene photographs to measure properties of natural images, then it is important to make sure that the photographs used are representative of our experience of the environment. Some researchers claim that rapid animal detection tasks often use images that frame the animal centrally in the picture (Wichmann, Drewes, Rosas & Gegenfurtner, 2010) as photographers tend to focus on salient objects. In contrast, other researchers have asked observers to take photographs of their environment at regular intervals as they travel from one location to another (Coppola, Purves, McCoy & Purves, 1998). More recently however, Lei et al. (2004) recorded photographs from the perspective of a monkey as it explored a field environment, whilst also recording neuronal activity. Perhaps a similar method could be used with human observers, such as using a head-mounted eye tracker to enable recordings of image and fixations. This method would be beneficial, as observers would have no control over when images were taken, and thus the viewpoint would be more accidental.

6.2.3. Scene statistics

In the scene statistics approach, future research will aim to include a broader variety of landscapes for analysis. Most current studies are located in America, close to university campuses (e.g. Yang & Purves, 2003; Potetz & Lee, 2003; Howe & Purves, 2002). Though these will be representative of the environments that most of the university population will have experienced, it is hard to argue that the scenes used are universally appropriate. It would be ideal if scene statistics could be analysed in different countries and environments. Indeed, image studies already show that there are *qualitative* differences between cities in different countries (Japan and America; Miyamoto, Nisbett & Masuda, 2006) so it would be interesting to quantify this and search for any differences in perceptual biases. In addition, differences have been shown in the way that some cultures perceive depth and size in 3D scenes. Turnbull (1961) described how an adult living in the Congo forest was taken into the mountains, where he saw a herd of buffalo and was confused about what they were.

Since the buffalo were at a great distance they would appear small on the retina and the man did not believe that he was seeing large animals, believing them instead to be insects. Turnbull explained that the BaMbuti Pygmies' experience of distance was limited to a relatively small range within the forest, thus he believed that the Pygmies had not learned the same size-distance relationships that we have much experience with in urbanised landscapes. It seems that by analysing scene statistics and examining perception in different cultures we would gain a true test of the natural statistics approach. This would benefit the understanding of whether priors are innate or learned.

6.3. Applications and Collaborations

It became apparent from the beginning of this project that there was great variety in the approaches and methods to study the links between human vision and the natural environment. Though the ultimate goal of the research was purely to better understand how the visual system works and why we are subject to perceptual biases, it is also clear that other researchers take a much more direct and applied route to understanding both image statistics and scene statistics. Indeed, researchers whose backgrounds lie in computer graphics (e.g. Pouli, Cunningham & Reinhard, 2010) and robotics (e.g. Vandapel, Donamukkala & Hebert, 2006) study image and scene statistics (respectively) in order to solve practical problems (e.g. how to use algorithms to detect objects within natural images) and then apply this knowledge (e.g. preventing a robot from falling into a hole). Importantly, additional research exists examining range statistics (e.g. Huang, Lee & Mumford, 2000), and it is possible to relate this information to biases in human vision (Yang & Purves 2003), so it would be beneficial for researchers across different fields to collaborate. Similarly, there already exist databases of co-registered intensity and range images that would benefit researchers of low-level vision (e.g. Potetz & Lee, 2003) and more applied sciences alike. For instance, animators of computer-generated imagery (CGI) would benefit from understanding the typical statistics of natural images and objects, in order to create the most realistic images (see Pouli et al., 2010, for review). Developers of computer games might utilise research about perception of 3D shape from 2D images in order to provide gamers with a sense of depth, and to *avoid* situations in which the images might be ambiguous. Designers of virtual reality

environments might wish to limit the information that they have to provide, and thus focus on information that will be directly informative for the viewer.

6.4. Conclusions

The natural statistics approach directs research towards perceptual biases that can be ecologically validated by measuring the statistics of the natural environment or how natural scenes project to 2D retinal images. Conversely, it drives research towards investigating the physical properties of the environment such as the distribution of oriented- or curved surfaces, and how these project to the retinal image or relate to known perceptual biases. Furthermore, investigation into phenomena such as observers' ability to perceived metric depth dependent on the curve of a 2D bounding contour could lead the generation of computational models to describe shape perception.

REFERENCES

Adams, W.J. (2007). A common light-prior for visual search, shape, and reflectance judgments. *Journal of Vision, 7,* (11), 1-7.

Adams, W.J. (2008). Frames of reference for the light-from-above prior in visual search and shape judgements. *Cognition, 107*, 137-150.

Adams, W.J., Graf, E.W., & Ernst, M.O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience, 7*, 1057-1058.

Adams, W.J., & Mamassian, P. (2004). Bayesian combination of ambiguous shape cues. *Journal of Vision, 4* (10), 921-929.

Andersen, G.J., Braunstein, M.L., & Saidpour, A. (1998). The perception of depth and slant from texture in three-dimensional scenes. *Perception, 27*, 1087-1106.

Atick, J.J. & Redlich, A. N. (1992). What does t he retina know about natural scenes? *Neural Computation, 4*, 196-210.

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

Bacon-Mace, N., Mace, M.J.M., Fabre-Thorpe, M. & Thorpe, S. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research, 45*, 1459-1469.

Baker, D.H. & Graf, E.W. (2009). Natural images dominate in binocular rivalry. *Proceedings of the National Academy of Sciences, 106 (*13), 5436-5441.

Banks, M.S. (1982). The development of spatial and temporal contrast sensitivity. *Current Eye Research, 2*, 191-198.

Banks, M.S., & Salapatek, P. (1978). Acuity and contrast-sensitivity in 1-, 2- and 3-month old human infants. *Invest. Ophthalmol. Visual* Science, 17, (4) 361-365.

Bertamini, M., & Wagemans, J. (2013). Processing convexity and concavity along a 2-D contour: Figure-ground, structural shape, and attention. *Psychonomic Bulletin & Review, 20*, 191-207.

Bex, P.J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of the Optical Society of America, 19* (6), 1096-1106.

Bex, P.J., Solomon, S.G., & Dakin, S. C. (2009). Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure. *Journal of Vision, 9* (10), 1-19.

Blake, R. (1989). A neural theory of binocular rivalry. *Psychological Review, 96* (1), 145-167.

Blake, R. (2001). A primer on binocular rivalry, including current controversies. *Brain and Mind, 2*, 5-38.

Brascamp, J.W., Knapen, T.H.J., Kanai, R., van Ee, R. & van den Berg, A.V. (2007). Flash suppression and flash facilitation in binocular rivalry. *Journal of Vision, 7*(12), 1-12.

Breitmeyer, B.G., & Ogmen, H. (2006). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception and Psychophysics, 62* (8), 1572-1595.

Brunswick, E., & Kamiya, J. (1953). Ecological cue-validity of "proximity" and other Gestalt factors. *American Journal of Psychology, 66*, 20-32.

Bulthoff, H.H. & Mallot, H.A., (1988). Integration of depth modules: stereo and shading. *Journal of the Optical Society of America A., 5* (10), 1749-1758.

Burge, J., Fowlkes, C. C., & Banks, M. S. (2010). Natural-scene statistics predict how the figure-ground cue of convexity affects human depth perception. *The Journal of Neuroscience, 30*, 7269–7280.

Campbell, W.F & Green, D.G. (1965). Optical and retinal factors affecting visual resolution *Journal of Physiology, 181*, 576-593.

Campbell, W.F. & Robson, J.G. (1967). Applications of Fourier analysis to the visibility of gratings*. Journal of Physiology*, 197, 551-566.

Caudek, C., Fantoni, C. & Domini, F. (2011). Bayesian modelling of perceived surface slant from actively-generated and passively-observed optic flow. *PloS One, 6*(4), e18731.

Cave, C.B., Blake, R., & McNamara, T.P. (1998). Binocular rivalry disrupts visual priming. *Psychological Science, 9 (4),* 299-302.

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

Coppola, D.M., Purves, H.R., McCoy, A.N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences, 95* (7), 4002-4006.

Cumming, G. (2001-2011). ESCI, Exploratory Software for Confidence Intervals.

Craven, B.J. (1993). Orientation dependence of human line-length judgements matches statistical structure in real-world scenes. *Proceedings of the Royal Society of London B, 253*, 101-106.

De Winter, J., & Wagemans, J. (2008). Perceptual saliency of points along the contour of everyday objects: A large-scale study. *Perception & Psychophysics, 70,* 50–64.

Durgin, F.H., Li, Z., & Hajnal, A. (2010). Slant perception in near space is categorically biased: Evidence for a vertical tendency. *Attention, Perception & Psychophysics, 72*, (7), 1875-1889.

Durgin, F.H., Hajnal, A., Li, Z., Tonge, N., Stigliani, A. (2010a). Palm boards are not action measures: An alternative to the two-systems theory of geographical slant. *Acta Psychologica, 134*, 182-197.

Elder, J.H., & Goldberg, R.M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*, 2, 324 -353.

Essock, E. A., DeFord, J.K., Hansen, B.C., & Sinai, M.J. (2003). Oblique stimuli are seen best (not worst!) in naturalistic broad-band stimuli: A horizontal effect. *Vision Research, 43*, 1329-1335.

Fantoni, C., Bertamini, M. & Gerbino, W. (2005). Contour curvature polarity and surface interpolation. *Vision Research, 45*(8), 1047-1062.

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7*(1), 1-29.

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America, 99*(14), 9596-9601.

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

Felsen, G., Touryan, J., Han, F. & Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. *PLoS Biology, 3*(10), 1819-1828.

Field, D.J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Optical Society of America, 4* (12), 2379-2394.

Field, D.J. (1989). What the statistics of natural images tell us about visual coding. *SPIE, 1077,* 269-276.

Furmanski, C.S., & Engel, S.A. (2000). An oblique effect in human primary visual cortex. *Nature Neuroscience, 3,* 535-536.

Gallant, J.L. (2003) Neural mechanisms of natural scene perception. In: The visual neurosciences (Chalupa LM, Werner JS, eds). Boston: MIT.

Gaspar, C.M., & Rousselet, G. A. (2009). How do amplitude spectra influence rapid animal detection? *Vision Research, 49*, 3001-3012.

Gauthier, I. (2000). Visual priming: The ups and downs of familiarity*. Current Biology, 10*, R753-R756.

Geisler, W.S. (2007). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology, 59*, 167-192.

Geisler, W.S. & Kersten, D. (2002). Illusions, perception and Bayes. *Nature Neuroscience, 5*, (6), 508-510.

Geisler, W.S., Perry, J.S., Super, B.J., & Gallogly, D.P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research, 41*, 711-724.

Geisler, W.S., & Ringach, D. (2009). Natural systems analysis. *Visual Neuroscience, 26*, 1-3.

Gibson, J.J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.

Gibson, J.J., & Cornsweet, J. (1951). The perceived slant of visual surfaces – optical and geographical. *Journal of Experimental Psychology, 44*, 11-15.

Girshick, A., Burge, J., Ehrlickman, G. & Banks, M. (2008, abstract). Prior expectations in slant perception: Has the visual system internalized natural scene geometry? *Journal of Vision, 8* (6), article 77.

Girshick, A.R., Landy, M.S., & Simoncelli, E.P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*, 926-932.

Gogel, W.C. & Tietz, J.D. (1979). A comparison of oculomotor and motion parallax cues of egocentric distance. *Vision Research, 19* (10), 1161-1170.

Goldstein, E.B. (2007). Sensation and Perception (6th ed.), England: Wadsorth.

Greene, M.R., & Oliva, A. (2009). The briefest of glances: the time course of natural scene understanding. *Psychological Science, 20* (4), 464-472.

Greenlee, M.W., Georgeson, M.A., Magnussen, S., & Harris, J.P. (1991). The time course of adaptation to spatial contrast. *Vision Research, 31*(2), 223-236.

Gregory, R.L. & Wallace, J.G. (1963). Recovery from early blindness: A case study. *Quarterly Journal of Experimental Psychology*, Monograph No.2.

Gruber, H.E. & Clark, W.C. (1956). Perception of slanted surfaces. *Perceptual and Motor Skills, 6*, 97-106.

Guyader, N., Chauvin, A., Peyrin, C., Herault, J. & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *C.R. Biologies, 327*, 313-318.

Von Helmholtz, H. (1867) Handbuch der physiologischen Optik, volume 9 of Allgemeine Encyk- lopaedie der Physik. Leipzig: Voss

Held, R. T., Cooper, E.A. & Banks, M.S. (2012). Blur and disparity are complementary cues to depth. *Current Biology, 22*, 426-431.

Hillis, J.M., Watt, S.J., Landy, M.S. & Banks, M.S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*, 967-992.

Howe, C, Q., & Purves, D. (2002). Range image statistics can explain the anomalous perception of length. *PNAS, 99* (20), 13184-13188.

Howe, C, Q., & Purves, D. (2005). Natural-scene geometry predicts the perception of angles and line orientation. *PNAS, 102* (4), 1228-1233.

Huang, J., Lee, A.B., & Mumford, D. (2000). Statistics of range images. *Computer Vision and Pattern Recognition, 1*, 324-331.

Joubert, O. R., Rousselet, G. A., Fize, D. & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object inference. *Vision Research, 47*, 3286-3297.

Joubert, O.R., Rousselet, G.A., Fabre-Thorpe, M. & Fize, D. (2009). Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. *Journal of Vision, 9* (1), 1-16.

Kaping, D., Tzvetanov, T. & Treue, S. (2007). Adaptation to statistical properties of visual scenes biases rapid categorization. *Visual Cognition, 15* (1), 12-19.

Kelley, T. A., Chun, M. M., & Chua, K-P. (2003). Effects of scene inversion on change detection of targets matched for visual salience. *Journal of Vision, 2*, 1-5.

Kleffner, D.A., & Ramachandran, V.S. (1992). On the perception of shape from shading. *Perception & Psychophysics, 52* (1), 18-36.

Knill, D.C., (1992). Perception of surface contours and surface shape: From computation to psychophysics. *Optical Society of America, 9* (9), 1449-1464.

Knill, D. & Richards. W. (1996). *Perception as Bayesian Inference*. Cambridge University Press

Knill, D.C., & Saunders, J.A. (2003). Do humans optimally integrate stereo and texture information for judgements of surface slant? *Vision Research, 43*, 2539-2558.

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

Koenderink, J.J., van Doorn, A.J., Kappers, A.M., & Todd, J.T. (1997). The visual contour in depth. *Perception & Psychophysics, 59*, 828-838.

Koutstaal, W., Wagner, A.D., Rotte, M., Maril, A., Buckner, R.L. & Schacter, D.L. (2001). Perceptual specificity in visual object priming: functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. *Neuropsychologica, 39*, 184-199.

Kovács, I, Papathomas, T.V., Yang, M. & Fehér, A. (1996). When the brain changes its mind: Interocular grouping during binocular rivalry. *PNAS, 93*, 15508-15511.

Kraha, A., Turner, H., Nimon, K., Reichwein Zientek, L. & Henson, R.K. (2012). Tools to support interpreting multiple regression in the face of multicollinearity. *Frontiers in Psychology, 3* (44).

Kristjánsson, Á. (2006). Simultaneous priming along multiple feature dimensions in a visual search task. *Vision Research, 46*, 2554-2570.

Langer, M.S., & Bülthoff, H.H. (2001). A prior for global convexity in local shape-from-shading. *Perception*, 30(4), 403-410.

Larson, A.M., & Loschky, L.C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision.* 9 (10): 6, 1-16.

Lewicki, M.S. (2002). Efficient coding of natural sounds. *Nature Neuroscience, 5*(4) 356-363.

Li, Z., & Durgin, F.H. (2010). Perceived slant of binocularly viewed large-scale surfaces: A common model from explicit and implicit measures. *Journal of Vision, 10*, 1-16.

Logothetis, N.K., Leopold, D.A. & Sheinberg, D.L. (1996). What is rivalling during binocular rivalry? *Nature, 380*, 621-624.

Loschky, L.C., Hansen, B.C., Sethi, A. & Pyimarri, T.N. (2010). The role of higher order image statistics in masking scene gist recognition. *Attention, Perception & Psychophysics, 72* (2), 427-444.

Loschky, L.C. & Larson, A. M. (2008). Localized information is necessary for scene categorization, including the Natural/Man-made distinction*. Journal of Vision.* 8(1), 1-9,

Loschky, L.C., & Larson, A.M. (2010). The natural/man-made distinction is made prior to basic-level distinctions in scene gist processing. *Visual Cognition, 18*(4), 513-536.

Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D. & Corbeille, J.L. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance, 33* (6), 1431 -1450.

Maloney, M.T. (1986). Evaluation of linear models of surface spectral reflectance with small numbers and parameters. *Optical Society of America, A, Optics, Image & Science. 3*(10), 1673-1683.

Mante, V., Frazor, R.A., Bonin, V., Geisler, W., & Carandini, M. (2005). Independence of luminance and contrast in natural scenes and in the early visual system. *Nature Neuroscience, 8* (12), 1690-1697.

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition, 81*, B1-B9.

Mamassian, P. & Goutcher, R. (2005). Temporal dynamics in bistable perception. *Journal of Vision, 5,* 361-375.

Mamassian, P., Landy, M., & Maloney, L. T. (2002). Bayesian modelling of visual perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of the brain: Perception and neural function* (pp. 13–36). Cambridge, MA: MIT Press.

Marr, D.C. (1982). Vision: A computational investigation into the human representation and processing of visual information. New York: Freeman.

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and

measuring ecological statistics. In *Proc. 8th Int'l. Conf. Computer Vision*, 2, 416–423.

McDonald, J.S. & Tadmor, Y. (2006). The perceived contrast of texture patches embedded in natural images. *Vision Research, 46*, 3098-3104.

Miyamoton, Y., Nisbett, R.E., & Masuda, T. (2006). Culture and the physical environment: Holistic versus analytic perceptual affordances. *Psychological Sciences, 17,* 113-119.

Norman, F.N., Phillips, F, & Ross, H.E. (2001). Information concentration along the boundary contours of naturally shaped solid objects. *Perception, 30*, 1285-1294

Norman, J.F., & Raines, S.R. (2002). The perception and discrimination of local 3-D surface structure from deforming and disparate boundary contours. Defining Configural Processing

Norman, J., Todd, J.T., Norman, H.F., Clayton, A.M. & McBride, T.R. (2006). Visual discrimination of local surface structure: Slant, tilt and curvedness. Vision Research, 46 (6-7), 057-1069.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145-175.

Olshausen, A. B. & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*, 607-609.

Palmer, S.E. (2002). Vision Science: Photons to Phenomenology (3rd ed). MIT Press, England.

Parraga C.A, Troscianko T., & Tolhurst D.J. (2002). Spatiochromatic properties of natural images and human vision. *Current Biology, 12,* 483–87.

Pearson, J. & Brascamp, J. (2008). Sensory memory for ambiguous vision. *Trends in Cognitive Sciences, 12* (9), 334-341.

Pearson, J., Clifford, C., & Tong, F. (2008). The functional impact of mental imagery on conscious perception. *Current Biology, 18*(13), 982-986.

Perlin, K. (2002). Improving noise. *ACM Transactions on Graphics, 21* (3), 681–682.

Porrill, J., Duke, P.A., Taroyan, N.A., Frisby, J.P. & Buckley, D. (2010). The accuracy of metric judgements: Perception of surface normal. *Vision Research, 50*, 1140-1157.

Potetz, B., & Lee, T.S. (2003). Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes. *Optical Society of America, 20* (7), 1292 – 1303.

Pouli, T., Cunningham D.W., & Reinhard, E. (2010). Image statistics and their applications in computer graphics. *Eurographics State of the Art Report, STAR.*

Proffitt, D.R., Bhalla, M., Gossweiler, R. & Midgett., J. (1995). Perceiving geographical slant. *Psychonomic Bulletin & Review, 2* (4), 409-428.

Purves, D., Lotto, R.B., Williams, S.M., Nundy, S., & Yang, Z., (2001). Why we see things the way we do: evidence for a wholly empirical strategy of vision. *Philosophical Transaction of the Royal Society, 356,* 285-297.

Rajimehr, R. (2004). Unconscious orientation processing. *Neuron, 41*, 663-673.

Ramachandran, V.S. (1988). Perception of shape from shading. *Nature,* 331, 163-166.

Richards, W.A., Koenderink, J.J. & Hoffman, D.D. (1987). Inferring three-dimensional shapes from two-dimensional silhouettes. *Optical Society of America, 4* (7), 1168-1175.

Rousselet, G., Joubert, O. & Fabre-Thorpe, M. (2005). How long to get the 'gist' of real-world natural scenes? *Visual Cognition.* 12(6), 852-877.

Russel, B., Torralba, A., Murphy, K. & Freeman, W.T. (2008). LabelMe: a database and web-based tool for image annotation. *International Journal of Computer Vision, 77*, 157-173.

Saunders, J.A. & Backus, B.T. (2006). Perception of surface slant from oriented textures. *Journal of Vision, 6*, 882-897.

Schacter, D.L., & Buckner, R.L. (1998). Priming and the brain. *Neuron, 20*, 185-195.

Schyns, P.G. & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recogition. *Psychological Science.* 5(4), 195-200.

Sekuler, R. & Blake, R. (2002) Perception, 4th Edition, McGraw-Hill.

O'Shea, J., Banks, M., & Agrawala, M. (2008). The preferred angle of illumination in shape-from-shading.. *In Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization (Los Angeles, California, August 09-10, 2008).*(pp. 135-142. New York, NY: ACM.

Shore, D., & Klein, R. (2000). The effects of scene inversion on change blindness. *The Journal of General Psychology, 127* (1), 27-43.

Simoncelli, E.P., & Olshausen, B.A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24*, 1193-1216

Simons, J.S., Koutstaal, W., Prince, S., Wagner, A.D., & Schacter, D.L. (2003). Neural mechanisms of visual object priming: Evidence for perceptual and semantic distinctions in fusiform cortex. *NeuroImage, 19*, 613-626.

Snowden, R.J., Stimpson, N., & Ruddle, R.A. (1998). Speed perception fogs up as visibility drops. *Nature, 392*, 450.

Stocker, A.A., & Simoncelli, E.P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience, 9*, 578-585.

Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience 1,* 3, 183–184.

Tanaka, Y., & Sagi, D. (1998). A perceptual memory for low contrast visual signals. *PNAS, 95*, 12729-12733.

Taylor-Covill, G.A. & Eves, F.F. (2013). The accuracy of 'haptically' measured geographical slant perception. *Acta Psychologica, 144* (2), 444-450.

Todd, J.T. & Reichel, F.D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychological Review, 4*, 643-657.

Todd, J.T., Thaler, L. & Dijkstra, T.M.H. (2005). The effects of field of view on the perception of 3D slant from texture. *Vision Research, 45*, 1501-1517.

Tolhurst, D.J., & Tadmor, Y. (2000). Discrimination of spectrally blended natural images: Optimisation of the human visual system for encoding natural images. *Perception, 29*, 1087-1100.

Torralba, A. & Oliva, A. (2003). Statistics of natural image categories. *Network: computational and neural systems*. 14, 391-412.

Tse, P.E. (2002). A contour propagation approach to surface filling-in and volume formation. *Psychological Review, 109* (1), 91-115.

Tsuchiya, N. & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neurosciene, 8*, 1096-1101.

Vandapel, N., Donamukkala, R.R., & Hebert, M. (2006). Unmanned ground vehicle navigation using aerial ladar data. *The International Journal of Robotics Research, 25 (*1), 31-51.

van Ee, R., Adams, W. J., Mamassian, P. (2003). Bayesian modelling of cue interaction: bistability in stereoscopic slant perception. *JOSA A, 20*, 1398-1406.

Vinje, W.E., & Gallant, J.L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science, 287,* 1273-1276.

Vuong, Q. C., Domini, F., & Caudek, C. (2006). Disparity and shading cue cooperate for surface interpolation. *Perception, 35*, 145-155.

Wagemans, J., De Winter, J., Op de Beeck, H., Ploeger, A., Beckers, T., & Vanroose, P. (2008). Identification of everyday objects on the basis of silhouette and outline versions. *Perception, 37*(2), 207– 244.

Warren, P.A., & Mamassian, P. (2010). Recovery of surface pose from texture orientation statistics under perspective projection. *Biological Cybernetics, 103*, 199.

Watt, S.J., Akeley, K., Ernst, M.O., Banks, M.S. (2005). Focus cues affect perceived depth. *Journal of Vision, 5*, 834-862.

Webster, M., & Miyahara, E. (1997). Contrast adaptation and the spatial structure of natural images. *Journal of the Optical Society of America 14* (9), 2355-2366.

Weiss, Y., Simoncelli, E.P., & Adelson, E.H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5* (6), 598-604.

Wichmann, F.A., Drewes, J., Rosas, P., & Gegenfurtner, K.R. (2010). Animal detection in natural scenes: Critical features revisited. *Journal of Vision, 10*, 1-27.

Wiggs, C., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology, 8* (2), 227-233.

Yang, Z., & Purves, D. (2003a). Image/source statistics of surfaces in natural scenes. *Network: Computation in Neural Systems, 14*, 371-390.

Yang, Z., & Purves, D. (2003b). A statistical explanation of visual space. *Nature Neuroscience, 6 (6),* 632-640.

Yao, A.Y.J., & Einhauser, W. (2008). Color aids late but not early stages of rapid natural scene recognition. *Journal of Vision, 8* (16), 1-13.

APPENDIX

Slant-Tilt Project Manual

Contents:

Definition of Terms

Outline of Experiment and Setup

Experiment Room Setup

My Coordinate Systems and Conversions

Analysis

Equipment Details

Definition of Terms

Slant: The degree to which a surface is rotated in depth from frontoparallel (FP; zero slant).

Tilt: The axis of rotation (e.g. horizontal/vertical).

PT Unit /PTU/Pan-Tilt Unit: The main electronic component used in the Slant-Tilt project. The PT unit produces movement about two perpendicular axes, so that a surface attached to the device can be slanted at different tilts.

Disk mount: The circular mount placed between the PT Unit and the stimulus. The disk acts as the base for rotation of the stimulus and also provides stability.

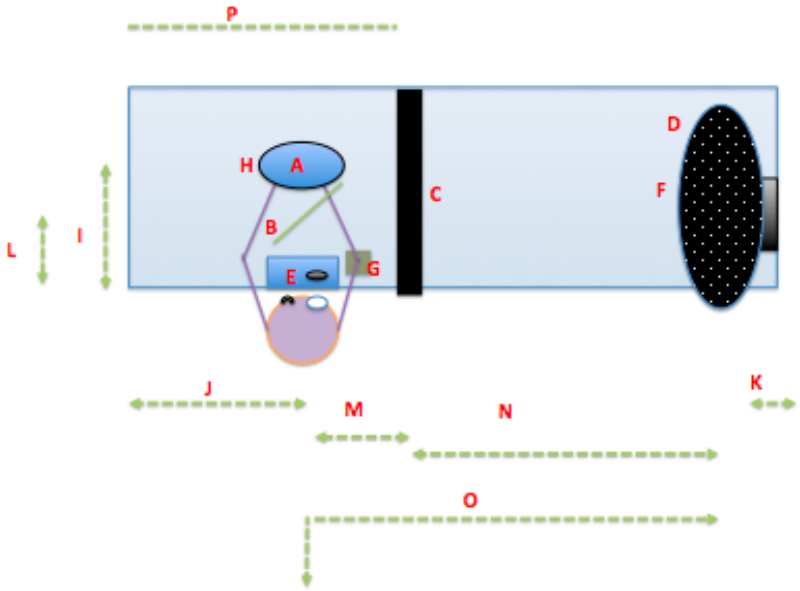
Outline of Experiment and Setup

The purpose of the experiment is to measure human slant perception about different tilt axes in the frontoparallel plane. The Pan-Tilt unit is used to rotate the attitude (slant and tilt) of a textured stimulus, and is controlled remotely using Matlab (Mathworks) on a Mac computer.

Participants view the stimulus through a 45° mirror that is connected to a chin and headrest. Participants then move a haptic paddle to match the perceived attitude of the stimulus; 2 axis rotary encoders in the haptic paddle measure the new attitude of the paddle and then send them via serial port to the computer and the data is then saved.

This guide includes details on each component of the experimental setup, as well as notes on how to convert between different coordinate systems.

Experiment Room Setup



A. Haptic paddle: diameter: 19cm B. Monocular mirror length: 10cm; binocular mirror length: 20cm C. Aperture: 5cm diameter D. Stimulus disk: 66cm diameter E. Eye/shutter/ height: 41.5cm F. Height at centre of aperture/frontoparallel stimulus: 41.5cm G. Height of arm rest: 22cm H. Height of Haptic Paddle, centre at frontoparallel: 41.5cm	I. Distance to Haptic Paddle from edge of table: 42cm J. Distance from left wall to centre of monocular mirror (77.5); binocular mirror (82.5), K. Distance from right wall to stimulus texture at frontoparallel: 43.5cm L. Distance from eye to mirror: 5.5 cm M. Distance from mirror to aperture: 54cm N. Distance from aperture to stimulus at FP: 121 O. Total viewing distance ~180cm P. Distance to occluder board from left wall: 132cm

Figure 0.1. Layout of experiment room with detailed measurements.

Haptic Paddle Coordinate System

The haptic paddle has two axes of rotation: vertical and horizontal. Rotary encoders from both axes feed back to the computer via serial port.

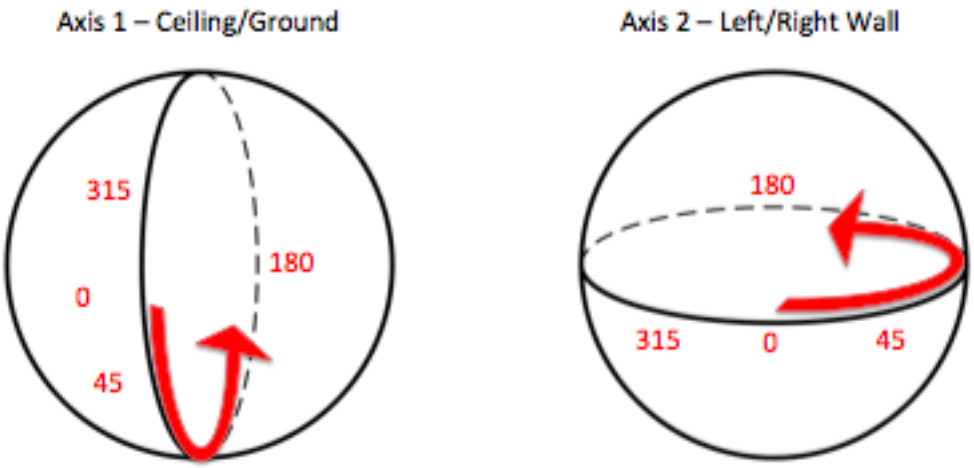


Figure 0.2 Diagram showing coordinate system for haptic paddle.

Analysis

Converting Rotation Values to Slant and Tilt

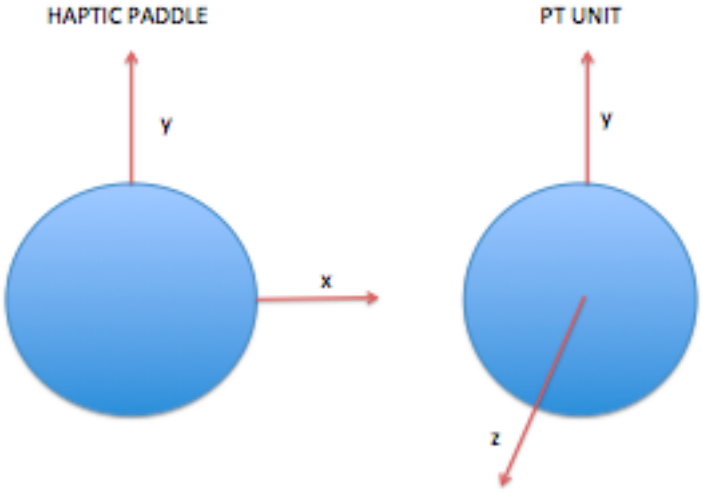


Figure 0.3. Rotary encoders attached to the haptic paddle code rotations about x (horizontal axis) and y (vertical axis). The PT Unit rotates about y (slant) and z (tilt). Participant settings of the haptic paddle are recorded as rotations around Axis 1 and 2 and thus must be converted in order for comparison with stimulus slant and tilt.[12]

To convert:

Let n = z represent the normal of the paddle at frontoparallel.

Let n' represent the normal after rotation by angles θ_1 and θ_2 respectively.

Now, using rotation matrices (linear algebra),

$$R_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_1 & -\sin\theta_1 \\ 0 & \sin\theta_1 & \cos\theta_1 \end{pmatrix}, \text{ and thus } R_1 n = \begin{pmatrix} 0 \\ -\sin\theta_1 \\ \cos\theta_1 \end{pmatrix}$$

[12] Thanks to James Elder for help with coordinate system conversion.

The axis of rotation for R_2 in world coordinates is $\mathrm{R}_1\mathbf{y} = \begin{pmatrix} 0 \\ \cos\theta_1 \\ \sin\theta_1 \end{pmatrix}$.

R_2 thus has the form

$$\mathrm{R}_2 = \begin{pmatrix} \cos\theta_2 & -\sin\theta_1\sin\theta_2 & \cos\theta_1\sin\theta_2 \\ \sin\theta_1\sin\theta_2 & \cos\theta_2\sin^2\theta_1 + \cos^2\theta_1 & \cos\theta_1\sin\theta_1\left(1-\cos\theta_2\right) \\ -\cos\theta_1\sin\theta_2 & \cos\theta_1\sin\theta_1\left(1-\cos\theta_2\right) & \cos\theta_2\cos^2\theta_1 + \sin^2\theta_1 \end{pmatrix}$$

Multiplying through and simplifying, we obtain:

$$\mathbf{n}' = \mathrm{R}_2\mathrm{R}_1\mathbf{n} = \begin{pmatrix} \sin\theta_2 \\ -\sin\theta_1\cos\theta_2 \\ \cos\theta_1\cos\theta_2 \end{pmatrix}.$$

This is the new surface normal, after rotation.

Let (ϕ,ψ) be the slant, tilt angles, respectively.

Then: $\cos\phi = \mathbf{z}^t\mathbf{n}' = \cos\theta_1\cos\theta_2 \rightarrow \phi = \arccos\left(\cos\theta_1\cos\theta_2\right)$

And $\tan\psi = -\sin\theta_1\cot\theta_2 \rightarrow \psi = -\arctan\left(\sin\theta_1\cot\theta_2\right)$.

Substitute angles with rotations of haptic paddle axis 1 and 2 (in radians) to obtain slant and tilt.

Converting Slant and Tilt to Vector Form

Though it is easy to represent slant and tilt visually using a polar plot, it is useful for analysis to convert into vector space. In vector space, a stimulus attitude or set attitude can be described in terms of surface normal towards ceiling/ground plane and left/right wall (see Figure 5.18).

First, convert slant and tilt to radians using pi/180 * θ.
Then:
Sin slant is the length of the projection of the stimulus on the retina.[13]
To calculate U, multiply sin slant by sin tilt (because my tilt values increase clockwise from vertical).
To calculate V, multiply sin slant by cos tilt (axis is rotated 90deg relative to U)

Calculate sin slant, sin tilt, and cos tilt.
U = sin (slant θ) * sin (tilt θ)
V = sin (slant θ) *cos (tilt θ)

[13] Thanks to James Elder for help with conversion.

Figure 0.4. PTU mounted to wall (shown rotated from home by 90). Mounting bracket allows entire unit to be shifted left/right or up/down if desired. Stimulus surface (shown from behind on the left) is 40cm from the wall at frontoparallel. Cable boxes are stuck to the wall to avoid collision with the surface as it rotates.

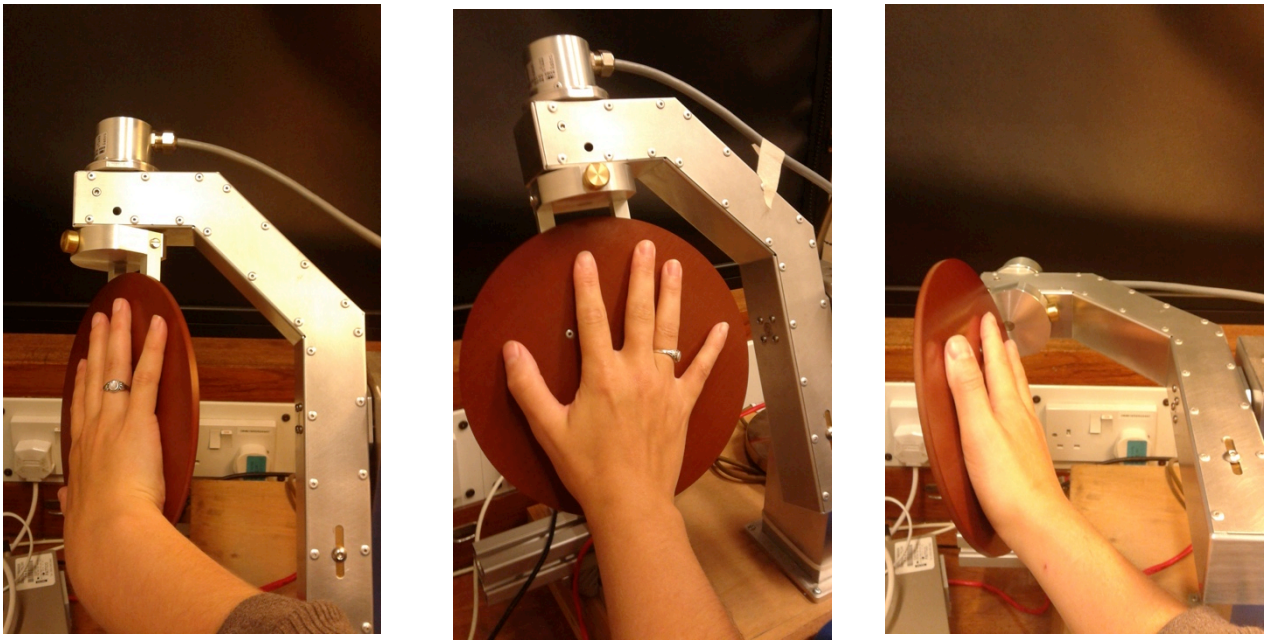


Figure 0.5. Example of haptic paddle in use. Left: steep right wall slant; centre: slight right wall slant, close to frontoparallel, right: steep ground-plane/left wall slant.

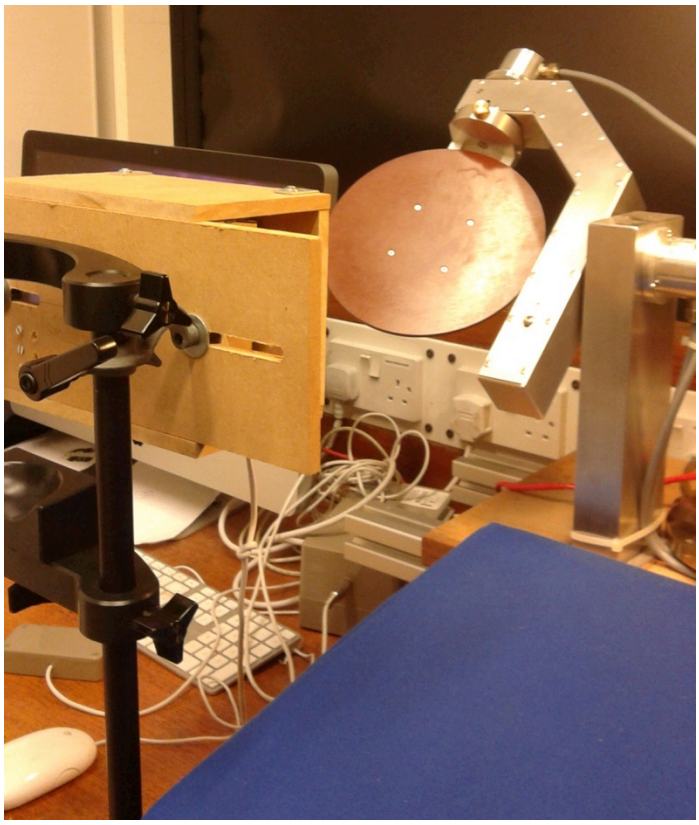


Figure 0.6. View of haptic paddle and armrest. Paddle is positioned in line with the mirror, central to the eyes.

Stimulus and EL Panel

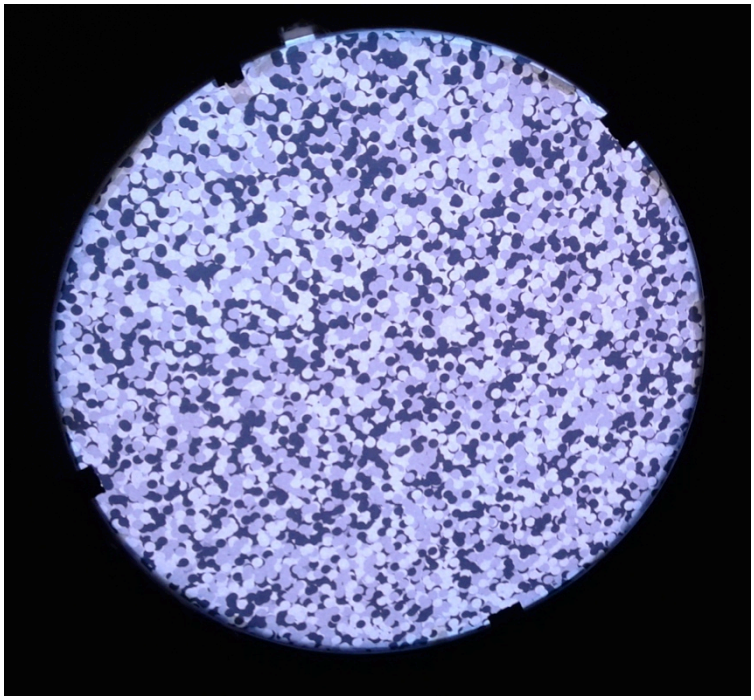
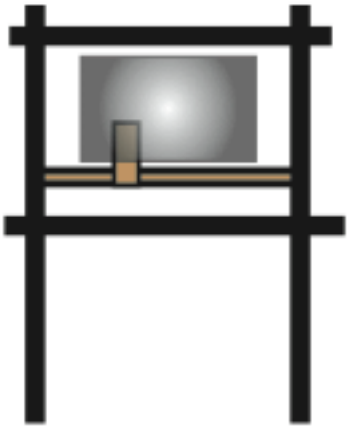


Figure 0.7. Full stimulus (Texture 1) in darkened room.

Mirrors and Head Rest

A



B

C

D

Figure 0.8. A:C show mirror set up for Part 1; D shows mirror set up for Part 2. A) Front diagram of mirror and chin rest, B) top diagram of mirror and chin rest, C) photograph of binocular mirror set up, D) photograph of monocular mirror set up, showing 45° mirror and button operated shutter (left). Mirror 1 length 20cm, mirror 2 length 10cm (designed to accommodate binocular and monocular viewing respectively).