Depth cue contributions in complex natural scenes

Rebecca E. Ranson

A thesis submitted for the degree of

Doctor of Philosophy

Department of Psychology

University of Essex

April 2023

“Remember that ‘seeing is believing’ puts the cart before the horse. Art is the concrete artifact of faith and expectation, the realization of a world that would otherwise be little more than a veil of pointless consciousness stretched over a gulf of mystery.”

- Stephen King, Duma Key

Acknowledgements

I will forever be grateful to my supervisor Paul Hibbard for being the best teacher and mentor. I have been so grateful for your endless generosity and patience. Thank you for enduring my many lists, and for creating such a warm and welcoming research environment. I am a better researcher for having worked with you.

Thank you also to my friends, family and work colleagues, who have given me the encouragement to continue, as well as patience and understanding while I disappeared to write this thesis. I particularly want to thank my husband Tom, my mother Sue and my brother Ben, I could not have done this without your unwavering love and support. Huge thanks also to my wonderful children, Oscar and Luna, who have been very understanding and patient while “Mummy work”. You have both inspired me to do this more than you know.

Thank you to my examiners for their interest in my work and their brilliant ideas and feedback. Thanks also to them, and my wonderful viva chair, for creating such an enjoyable viva experience.

Finally, thank you to everyone else I have met and worked with during my time at the university. To my second supervisor Loes van Dam, I am so grateful for your insightful feedback and helpful suggestions. To the admin and technical staff who have always been more than happy to help, thank you for all your hard work and quick replies. To my fellow PhD students, those who have completed and those who have yet to, thank you for the motivation and encouragement, as well as the fun and friendly environment.

Thesis abstract

Perception of depth in natural viewing is based on information reported by many cues. These cues have been studied extensively individually, but many experiments in this area isolate a single cue, or a pair, in order to measure their relative reliabilities, and therefore weighting within the depth estimate under Bayesian cue combination. However, isolating cues does not simulate natural viewing, and studies show that simplifying viewing can make decoding a scene challenging for the visual system. Additionally, much work in this area presents overly-simplified stimuli, such as lines or dots, although these do not represent the complexities of natural objects, and are a poor fit for simulating the challenge posed to the visual system in decoding natural scenes. A branch of research has sought to address these issues of simplification of viewing and stimuli alike, by presenting more complex stimuli under naturalistic viewing conditions. However, the stimuli chosen, often photographs of smooth real or simulated shapes, often do not represent complex, naturally-occurring scenes, and report relative interactions rather than discrete weightings. The work presented in this thesis sought to address these issues by using advanced technological equipment and techniques to produce complex, naturalistic scenes with which to measure the contribution of depth cues to the weighted Bayesian estimate. The experiments contained within explore both retinal and extraretinal cues, binocular and pictorial cues, and a range of global and local viewing conditions. Overall, results show modest weightings for binocular cues such as disparity and vergence, little benefit to the overall weighting for the occluding contour and surface luminance pictorial cues, and a clear result of pictorial shape from shading in a real-world facial makeup study, with evidence throughout of benefits for the inclusion of complex naturalistic stimuli in future work of this kind.

Table of Contents

Acknowledgements 3
Thesis abstract 4
List of figures and tables 10
1 Introduction 17
1.1 Perceiving the world 17
1.2 Cues to depth 21
1.2.1 Binocular cues 22
1.2.2 Monocular cues 24
1.2.3 Retinal and extraretinal cues 29
1.3 Combining cues 29
1.3.1 The ‘ideal Bayesian observer’ 36
1.4 Limitations 37
1.5 Perception of surfaces 40
1.6 Measuring the contributions of depth cues in complex natural scenes 41
2 Methods 45
2.1 Limits of traditional methods 45
2.2 Use of advanced technology 45
2.3 Creating naturalistic 3D models 49
2.3.1 Scanning technology and procedures 49
2.4 Producing 3D scenes 52
2.4.1 Creating 3D meshes 52
2.4.2 Rendering the scenes 53
2.4.3 Face stimuli 54
2.5 Introduction of experimental methods 54
2.5.1 Gauge figure task 55
2.5.2 Recreating a 3D mesh 59
2.6 Analyses 63
2.6.1 Statistical power 64
2.6.2 Regression 65
2.6.3 Linear mixed effects models 67
3 Vergence scales binocular depth estimates, but does not account for shape constancy 69
3.1 Abstract 69
3.2 Introduction 70
3.2.1 Calculating the vergence angle 71
3.2.2 Vergence, disparity and depth 76
3.2.3 Shape constancy 77
3.2.4 Vergence uncertainty 81
3.2.5 Measuring certainty of vergence and shape constancy 84

3.3 Methods ... 90
3.3.1 Participants ... 90
3.3.2 Apparatus ... 91
3.3.3 Stimuli ... 92
3.3.4 Procedure ... 95
3.4 Results ... 97
3.4.1 Data treatment and psychometric functions ... 97
3.4.2 Effect of distance of PSE and JND ... 102
3.4.3 Shape constancy ... 107
3.4.4 Shape constancy and certainty of vergence ... 109
3.5 Discussion ... 111
3.5.1 Findings ... 111
3.5.2 Other future work ... 116
4 Evidence that 'Dark is Deep', but only at far distances ... 120
4.1 Abstract ... 120
4.2 Introduction ... 121
4.2.1 Lighting and luminance ... 121
4.2.2 Previous work ... 124
4.2.3 The present work ... 127
4.3 Methods ... 130
4.3.1 Participants ... 130
4.3.2 Design ... 132
4.3.3 Apparatus ... 132
4.3.4 Stimuli ... 133
4.3.5 Procedure ... 135
4.4 Results ... 136
4.4.1 Data treatment ... 136
4.4.2 Accuracy of judgements ... 136
4.4.3 Object close ... 138
4.4.4 Object far ... 139
4.4.5 Scene close ... 140
4.4.6 Scene far ... 141
4.5 Discussion ... 142
4.5.1 Findings ... 142
4.5.2 Future work ... 144
5 Size and sampling scale both affect the texture and shape of objects captured with a gauge figure task ... 146
5.1 Abstract ... 146
5.2 Introduction ... 147
5.2.1 Importance of scale ... 150
5.2.2 Global versus local viewing ... 154

5.2.3 Measuring the change in captured shape 157
5.3 Methods 160
5.3.1 Participants 160
5.3.2 Apparatus and materials 161
5.3.3 Stimuli 163
5.3.4 Design 167
5.3.5 Procedure 168
5.4 Results 169
5.4.1 Data treatment 169
5.4.2 Experiment 1 172
5.4.3 Experiment 2 180
5.5 Discussion 183
5.5.1 Experiment 1 183
5.5.2 Experiment 2 187
5.5.3 Future work 189
6 Binocular cues provide a modest contribution to depth judgements at close distances in naturalistic scenes 191
6.1 Abstract 191
6.2 Introduction 192
6.2.1 Calculating binocular disparity 193
6.2.2 Gain on binocular disparity 202
6.2.3 Binocular disparity in cue combination 205
6.2.4 Measuring the contribution of individual cues 207
6.2.5 Measuring surface orientation 209
6.2.6 Simulating IOD gain 210
6.3 Methods 217
6.3.1 Participants 217
6.3.2 Apparatus 218
6.3.3 Stimuli 220
6.3.4 Procedure 222
6.4 Results 223
6.4.1 Data treatment 223
6.4.2 Experiment 1 224
6.4.3 Experiment 2 235
6.5 Discussion 242
6.5.1 Experiment 1 242
6.5.2 Experiment 2 246
6.5.3 Future work 248
7 Occluding contour and shape from shading alone do not convey 3D metric depth 250
7.1 Abstract 250

7.2 Introduction 250
7.2.1 Pictorial cues 251
7.2.2 Measuring depth estimates for pictorial cues 255
7.2.3 Creating a gain on pictorial cues 255
7.2.4 Cue combination considerations 257
7.3 Methods 260
7.3.1 Participants 260
7.3.2 Apparatus and materials 261
7.3.3 Stimuli 262
7.3.4 Procedure 266
7.4 Results 267
7.4.1 Data treatment 267
7.4.2 Experiment 1 271
7.4.3 Experiment 2 276
7.5 Discussion 284
7.5.1 Findings from Experiment 1 and 2 284
7.5.2 Future studies 288
8 Makeup filters narrow and elongate the appearance of facial features 292
8.1 Abstract 292
8.2 Introduction 293
8.2.1 Makeup and perception 293
8.2.2 Measuring depth estimates for pictorial cues 297
8.3 Methods 299
8.3.1 Participants 299
8.3.2 Apparatus and stimuli 299
8.3.3 Procedure 301
8.4 Results 302
8.4.1 Data treatment 302
8.4.2 Global versus localised feature analysis 304
8.4.3 Depth range 304
8.4.4 Affine transformations 308
8.4.5 Surface roughness 313
8.5 Discussion 315
8.5.1 Findings 315
8.5.2 Future studies 319
9 General discussion 321
9.1 Review of thesis aims 321
9.2 Summary of main findings 321
9.3 Laboratory versus online data collection 328
9.3.1 Pros of online data collection 329

9.3.2 Cons of online data collection 330
9.3.3 Conclusions 332
9.4 Future directions 333
References 335
10 Appendix 366
10.1 Chapter 3 366
10.2 Chapter 4 367
4.1.1 Model comparisons 367
10.3 Chapter 5 369
10.3.1 Experiment 1 369
10.3.2 Experiment 2 370
10.4 Chapter 6 370
10.4.1 Experiment 1 370
10.4.2 Experiment 2 372
10.5 Chapter 7 373
10.5.1 Experiment 1 373
10.5.2 Experiment 2 376
10.6 Chapter 8 378
10.6.1 Depth range 378
10.6.2 Affine transformations 379
10.6.3 Surface roughness 379

List of figures and tables

Figures

Figure 1.1: Distance versus depth **22**

Figure 1.2: Binocular disparity **23**

Figure 1.3: Vergence angle **24**

Figure 1.4: Accommodation **25**

Figure 1.5: Light and shading **26**

Figure 1.6: Occlusion **26**

Figure 1.7: Relative size **27**

Figure 1.8: Texture gradient **28**

Figure 1.9: Bayesian inference **33**

Figure 1.10: Cue combination **35**

Figure 2.1: 3DPixx system **47**

Figure 2.2: Results of scanning **52**

Figure 2.3: Binocular versus cyclopean viewpoint rendering **53**

Figure 2.4: Gauge rotations **56**

Figure 2.5: Gauge slant and tilt **57**

Figure 2.6: Barycentres **60**

Figure 2.7: Exponential gradients **61**

Figure 2.8: Affine transformations **62**

Figure 3.1: Triangulation **72**

Figure 3.2: Predicted vergence angles **74**

Figure 3.3: Difference between vergence angles **75**

Figure 3.4: Visual angle **78**

Figure 3.5: Change in visual angle **79**

Figure 3.6: Probability of vergence angle by distance (Scarfe & Hibbard, 2017) **82**

Figure 3.7: Nonius lines stimuli **92**

Figure 3.8: Differing nonius orientations **93**

Figure 3.9: Triangles stimuli **94**

Figure 3.10: View of the experiment **96**

Figure 3.11: Psychometric fit **98**

Figure 3.12: PSE and JND **99**

Figure 3.13: Psychometric curve **101**

Figure 3.14: Vergence and depth task results **102**

Figure 3.15: Fixation error **105**

Figure 3.16: Fixation distance **106**

Figure 3.17: Fixation disparity **107**

Figure 3.18: Shape constancy and certainty of vergence **110**

Figure 4.1: Shading by light source **122**

Figure 4.2: Angle of reflectance **123**

Figure 4.3: Charcoal drawing of a skull **124**

Figure 4.4: Binocular and pictorial cue conflict **127**

Figure 4.5: Stimuli scenes **133**

Figure 4.6: Luminance to distance correlation **134**

Figure 4.7: Relative distance errors **138**

Figure 5.1: Gauge size consideration **152**

Figure 5.2: Face Constriction Process **158**

Figure 5.3: Stimuli **163**

Figure 5.4: Gauge locations **164**

Figure 5.5: Letterbox conditions **165**

Figure 5.6: Sweetcorn **165**

Figure 5.7: Sample locations **166**

Figure 5.8: Reconstructed meshes Experiment 1 ..... **171**

Figure 5.9: Depth range Experiment 1 172

Figure 5.10: Surface roughness 174

Figure 5.11: Local versus global viewing 177

Figure 5.12: Presentation order 179

Figure 5.13: Reconstructed meshes Experiment 2 180

Figure 5.14: Surface roughness Experiment 2 181

Figure 5.15: Depth range Experiment 2 182

Figure 6.1: Geometry of triangulation 193

Figure 6.2: Vergence angle comparison 195

Figure 6.3: The horopter and disparity 196

Figure 6.4: Crossed and uncrossed disparity 197

Figure 6.5: Geometric relationship between the eyes and the point of fixation 198

Figure 6.6: Geometric relationship between the eyes and a target point 200

Figure 6.7: Change in disparity with viewing distance 202

Figure 6.8: Predicted change in disparity with varying IOD 203

Figure 6.9: Simulating IOD gain 211

Figure 6.10: Predicted effect of gain for binocular cues only 213

Figure 6.11: Predicted effect of gain for binocular disparity and other cues 215

Figure 6.12: Pomegranate with gauge locations 220

Figure 6.13: Cluttered scene with gauge locations 221

Figure 6.14: Reconstructed meshes Experiment 1 224

Figure 6.15: Gradient point comparison 225

Figure 6.16: Sampling points Experiment 1 226

Figure 6.17: Setting gain against IOD gain 227

Figure 6.18: Relative depth range Experiment 1 230

Figure 6.19: Mean affine transformations Experiment 1 232

Figure 6.20: Reconstructed meshes Experiment 2 235

Figure 6.21: Mean relative depth Experiment 2 237

Figure 6.22: Squashed and stretched gain comparison Experiment 2 238

Figure 6.23: Affine transformations Experiment 2 239

Figure 6.24: Gauge size consideration 243

Figure 7.1: Showing shape from shading and occlusion 251

Figure 7.2: Self occlusion 253

Figure 7.3: Pictorial cues, Bayesian prior and binocular cues slope comparison 256

Figure 7.4: The cyclopean eye 261

Figure 7.5: Stimuli Experiment 1 263

Figure 7.6: Sample points Experiment 1 264

Figure 7.7: Depth of the 3D model presented in Experiment 1 264

Figure 7.8: Point distances 265

Figure 7.9: Stimuli Experiment 2 266

Figure 7.10: Reconstructed meshes Experiment 1 268

Figure 7.11: Maximum angle constraint 270

Figure 7.12: Large gradient point 271

Figure 7.13: Affine transformations Experiment 1 273

Figure 7.14: Relative depth Experiment 1 276

Figure 7.15: Reconstructed meshes Experiment 2 277

Figure 7.16: Affine transformation Experiment 2 single object 278

Figure 7.17: Affine transformation Experiment 2 cluttered scene 280

Figure 7.18: Depth gain 282

Figure 7.19: Relative depth Experiment 2 283

Figure 7.20: Cyclopean versus binocular viewing 287

Figure 8.1: The golden ratio 294

Figure 8.2: Highlight and contour 298

Figure 8.3: Face stimuli 300

Figure 8.4: Gauge locations 301

Figure 8.5: Reconstructed meshes 303

Figure 8.6: Localised feature analysis 304

Figure 8.7: Relative depth range comparison 305

Figure 8.8: Noses in profile 307

Figure 8.9: Affine transformations full face 309

Figure 8.10: Affine transformations 311

Figure 8.11: Noses from above 312

Figure 8.12: Surface roughness full face analysis 313

Figure 8.13: Surface roughness localised features 315

Figure 8.14: The change in depth and shape of the nose 318

Figure 10.1: Participant 27 373

Tables

Table 3-1: LME results for PSE and JND of both tasks 103

Table 3-2: Correlation coefficients for PSE and JND in the vergence task 105

Table 3-3: Correlation coefficients for JND of vergence and PSE of depth..... 109

Table 3-4: Correlation coefficients for JND in the depth and vergence tasks..... 110

Table 4-1: LME results for predicting relative error in distance estimations object close 139

Table 4-2: LME results for predicting relative error in distance estimations object far 139

Table 4-3: LME results for predicting relative error in distance estimations scene close 140

Table 4-4: LME results for predicting relative error in distance estimations scene far 141

Table 5-1: Depth range ANOVA results Experiment 1..... 173

Table 5-2: Surface roughness ANOVA results Experiment 1..... 175

Table 5-3: Gauge size preference 175

Table 5-4: LME results for the impact of gauge scaling on surface roughness.....182

Table 6-1: LME results for predicting set gradient Experiment 1 228

Table 6-2: LME results for affine transformation Experiment 1 233

Table 6-3: LME results for squashed and stretched IOD affine transformation Experiment 1. 234

Table 6-4: LME results for predicting reported depth Experiment 2 236

Table 6-5: LME results for squashed and stretched gains Experiment 2 238

Table 6-6: LME results for affine transformations Experiment 2 240

Table 6-7: LME results for squashed and stretched gains Experiment 2 241

Table 7-1: LME results for affine transformations Experiment 1 274

Table 7-2: LME results for affine transformations for the single object Experiment 2 279

Table 7-3: LME results for affine transformations for cluttered scene object Experiment 2 . ..281

Table 7-4: LME results for relative depth range object and scene comparison 284

Table 8-1: LME results for depth range full face analysis 305

Table 8-2: LME results for depth range of localised features 306

Table 8-3: Depth range ANOVA results affine transformations full face analysis 310

Table 8-4: Depth range ANOVA results affine transformations localised feature analysis 311

Table 8-5: LME results for surface roughness full face analysis 314

Table 10-1: Goodness of fit comparison for PSE and JND of both tasks 366

Table 10-2: Goodness of fit comparison for fixation disparity against distance 367

Table 10-3: LME results ground truth comparison 367

Table 10-4: LME results relative distance error for the four groupings 368

Table 10-5: Goodness of fit comparisons for depth range Experiment 1 369

Table 10-6 Goodness of fit comparisons for surface roughness Experiment 1.....369

Table 10-7: Goodness of fit comparisons for surface roughness Experiment 2 370

Table 10-8: Goodness of fit comparisons for depth range Experiment 2 370

Table 10-9: Goodness of fit comparison between LME models for set depth..... 371

Table 10-10: Goodness of fit comparison for affine transformations Experiment 1..... 372

Table 10-11: Goodness of fit comparison for reported depth Experiment 2 372

Table 10-12: Goodness of fit comparison for affine transformations Experiment 2..... 373

Table 10-13: Slope values from regression analysis including participant 27..... 374

Table 10-14: Affine transformations Experiment 1 including participant 27 375

Table 10-15: LME results including vertex producing extreme depth gradients 375

Table 10-16: Goodness of fit comparison for affine transformations Experiment 1..... 376

Table 10-17: Goodness of fit comparison for affine transformations single object 377

Table 10-18: Goodness of fit comparison affine transformations scene object 377

Table 10-19: Goodness of fit comparison relative depth Experiment 2 378

Table 10-20: Goodness of fit comparisons for depth range 378

Table 10-21: Goodness of fit comparison for affine transformations 379

Table 10-22: Goodness of fit comparisons for surface roughness full face 379

1 Introduction

1.1 Perceiving the world

In order to safely navigate the world around it, an animal needs to be able to perceive obstacles in its path (Chen, McNamara, Kelly, & Wolbers, 2017). To perform this task well, the perception of the distance to, and shape of, objects needs to be as accurate as possible. However, there are competing models of how the brain interprets the world around us, which Linton et al. (2022) summarise in their recent work on approaches to three-dimensional (3D) vision. These fall into three main groups: approaches and models that recover the 3D metric properties of the scene, 3D models that recover non-metric properties of the scene, and approaches that do not include a 3D model of any kind.

The dominant approach to human vision consists of 3D models that recover metric properties from scenes (Linton, et al., 2022). These include Bayesian models that are based around likelihood functions that express the probability density function or probability distribution of the most likely scene properties to have created a given image on the retina. Bayesian models and related frameworks have remained the most dominant approaches in vision research for the last 25 years. There are three variations of the Bayesian approach, two of which are highly relevant to this work and will be outlined here.

The first Bayesian approach is the linear cue combination, or weak fusion, model. This model defines perception as a process of combining a series of sensory cues, assuming that cues are unbiased and accurate, although noisy and uncertain. By taking a weighted average, where cues are weighted by their relative reliabilities

derived from how uncertain they are, the effect of noise on the overall depth estimate is reduced.

This approach makes two key assumptions. Firstly, the model assumes that cues are unbiased, such that the most likely scene property specified by the likelihood function corresponds with the true value in the physical world. Secondly, the weak fusion model predicts that more reliable cues should be given more weight in the weighted average. Many studies have provided evidence consistent with the weighted-averaging model for motion (Glennerster, Tcheang, Gilson, Fitzgibbon, & Parker, 2006; Johnston, Cumming, & Landy, 1994; Scarfe & Hibbard, 2011; Svarverud, Gilson, & Glennerster, 2010), texture (Hillis, Watt, Landy, & Banks, 2004; Johnston, Cumming, & Parker, 1993; Knill & Saunders, 2003; Saunders & Chen, 2015), shading (Lovell, Bloj, & Harris, 2012), stereopsis (Chopin, Levi, & Bavelier, 2017) and focus cues (Watt, Akeley, Ernst, & Banks, 2005), as well as for other aspects of depth judgements (Baird, 1970; Brenner & van Damme, 1998), visual and haptic combinations (Burge, Girshick, & Banks, 2010; Ernst & Banks, 2002; Gepshtein, Burge, Ernst, & Banks, 2005; Helbig & Ernst, 2007), and other sensory modalities (Ernst, 2006; Hillis, Ernst, Banks, & Landy, 2002). However, biases in the estimation of 3D scene properties (Koenderink, van Doorn, Kappers, & Todd, 2002; Koenderink, van Doorn, & Lappin, 2000; Wagner, 1985; Foster, Fantoni, Caudek, & Domini, 2011), derived from several cues (Bradshaw, Parton, & Glennerster, 2000) or integration of many cues (Tyler, 2019; Scarfe & Hibbard, 2011; Domini, 2023; Koenderink, van Doorn, Kappers, & Lappin, 2002) and deviations from optimal cue-weighting (Landy & Kojima, 2001; Oruç, Maloney, & Landy, 2003; Rosas, Wichmann, & Wagemans, 2007; Chen & Tyler, 2015; Scarfe, 2022; Rahnev & Denison, 2018; Rosas & Wichmann, 2011) have both been reported.

Another Bayesian approach is the nonlinear cue combination, or strong fusion, model. This model handles some of the potential limitations of the weak fusion model by accepting that individual cues inherently contain some level of bias, and the process is focused rather on constraints around the most likely scene to create the given retinal image, rather than reducing sensory noise in the weighted average. It is based on the Gregorian theory of perceptual hypothesis testing that uses a probabilistic match between the sensory input data and prior knowledge (Gregory, 1963). Given that this model suggests a full 3D representation of any scene is available (Tyler, 2020), it still concerns itself with metric scene recovery (Linton, et al., 2022) and therefore cue combination rules around the inverse mapping from images to world remain for this approach. The strong model addresses another important criticism of the weak fusion model in that the latter does not account for the increased perception of depth from weak depth cues observed empirically (Tyler, 2020). For instance, Tyler (2020) discusses how weak fusion as an averaging model would not sum multiple weak cues to approach veridical depth, instead averaging to a flattened percept, and instead suggests the addition of cues to address this.

Another consideration is how the visual system identifies which estimates come from the same surface or object and therefore should be combined in the weighted average. Girshick and Banks (2009) explored the role of conflict in cue combination by introducing bias to cues. They presented participants with surfaces with differing levels of conflict between the available cues of disparity and texture, and asked them to report the slant of the surface. They found that small levels of conflict led to the weighted averaging. However, large conflict resulted in robust averaging behaviour, in that variation was not overly skewed by outlier estimates. This meant that where estimates were so far removed from the central tendency they became outliers, i.e.

with longer tails than a Gaussian distribution, they were no longer included in the weighted average and the perceived slant was instead determined by a single cue. Rideaux and Welchman (2018) purport that this robust averaging of cues is due to proscription of unrealistic combinations of neuron excitation.

Other work on the strong fusion model proposes interdependencies between cues to combat the idea from weak fusion that posterior distributions, that is the likelihood of a certain scene given the observed image, are independent, whereas it is argued that this form of prior information may cause them to be dependent and even contradictory at times (Yuille & Bülthoff, 2008). Indeed, some work in this area has shown integrating non-independent cues such as shading and texture under strong fusion results in better depth accuracy than if the cues were used independently (Bülthoff & Mallot, 1988).

Another approach to vision outlined by Linton et al. (2022) concerns 3D models that do not recover metric properties from scenes. Given the documented failures of the visual system to accurately retrieve the metric properties of the scene, this group of approaches questions whether the visual system is trying to retrieve this metric information at all, instead focusing on qualitative models of scene geometry, for example the ability to identify flowing topographical landscapes of hills and dales (Koenderink, van Doorn, & Wagemans, 2015; Koenderink, 2012). This method, however, would work differently between simply observing a scene, and interacting with it, for example reaching to grasp an object, which would require metric estimates (Goodale & Milner, 1992).

Finally, there are theoretical approaches that do not include a 3D model, referred to as direct perception and sensorimotor perception. Under Gibson's direct

perception account, vision operates through a process in which we directly perceive the structure of the world through the recognition of invariants in the retinal image (Gibson, 1950; Gibson, 1966; Gibson, 1979) which, as the name suggests, derive directly from the constantly changing image on the retina, rather than this process being mediated by the construction of an explicit three-dimensional model (Marr, 1982). Wagner (1985) argued that this model implies that visual space should be strictly Euclidean in nature, conflicting with empirical findings that suggest non-Euclidean aspects of visual space which includes recovery of 3D structure from motion (Domini & Braunstein, 1998; Domini & Caudek, 2003; Domini, Caudek, & Richman, 1998; Tittle, Todd, Perotti, & Norman, 1995; Todd & Bressan, 1990; Todd & Norman, 1991), as well as other cues such as texture, shading and disparity (Norman & Todd, 1992; 1993; Todd & Reichel, 1989; Fernandez & Farell, 2009; Glennerster, Rogers, & Bradshaw, 1996). Sensorimotor approaches to understanding perception, in which the focus is on how sensory information is used to directly control behaviour, can also be classified as direct theories that do not make use of an explicit three-dimensional representation (O'Regan & Noë, 2001).

1.2 Cues to depth

Possessing two forward-facing eyes, humans have a number of sources of information on which to base visual judgements of the structure of the environment, some involving both eyes which creates binocular cues, and some stemming from the information of each eye individually, giving monocular cues. The cues relevant to this body of work will briefly be outlined here through illustrative examples, with further detail provided in the subsequent chapters. Cutting and Vishton (1995)

provide an in-depth look at the differing cues to depth, including their reliabilities and limitations. This work is used as a reference guide to cues throughout this body of work to discuss depth and distance judgements.

We define distance as the egocentric distance between an observer and a location in 3D space, and depth as the difference in distance between two points (Tresilian & Mon-Williams, 2000). For clarity, this description shall be used throughout this work, as defined in Figure 1.1, where distance (D) is the length along the line of sight between the observer and the object, and depth (z) is the measure across the object itself.

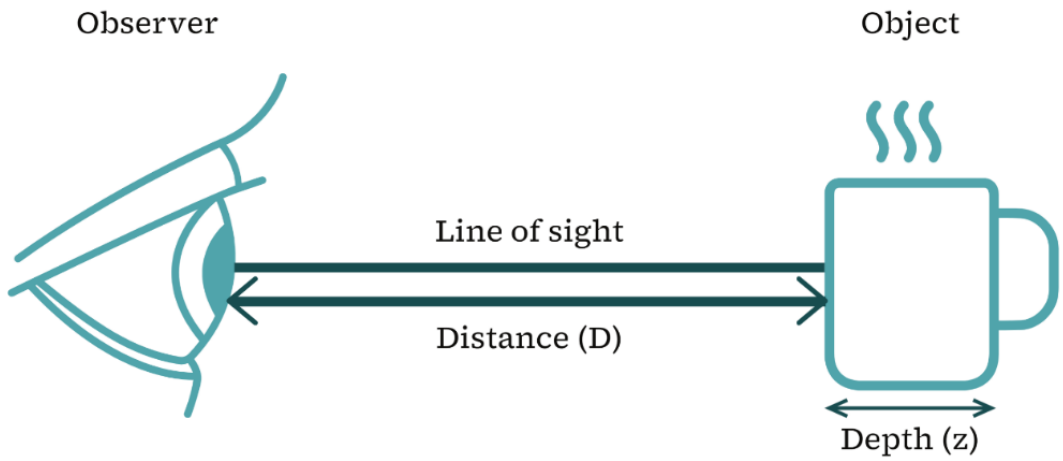


Figure 1.1: Distance versus depth. Image illustrating the difference between distance and depth in this work, showing the distance (D) from the observer to the object, and the depth (z) across the object.

1.2.1 Binocular cues

1.2.1.1 Binocular disparity

Being separated by, on average, 6.3cm (Dodgson, 2004), the two eyes see a slight variation in their image of the world, the differences between which can provide information about the location and shape of objects. These differences are referred to as binocular disparities. The brain combines these two images in a process called

fusion, and the differences provide information about depth. Binocular disparities are illustrated in Figure 1.2 where the apple and the tree fall on differing points on the retina. When fixating the tree, the image falls on the same location on each retina, resulting in no disparity (Stidwell & Fletcher, 2011). However, the apple falls on different, non-corresponding points on each retinal image, giving it a non-zero disparity, which in this case creates a crossed disparity and makes the apple appear closer than the tree. Binocular disparity is often cited as the most reliable cue to depth (Harris, 2004; Keefe, Hibbard, & Watt, 2011), and it has been shown that depth can be perceived with only this cue available (Julesz, 1971). Related to this cue is stereopsis, defined as the awareness of 3D space (Koenderink, van Doorn, & Wagemans, 2011).

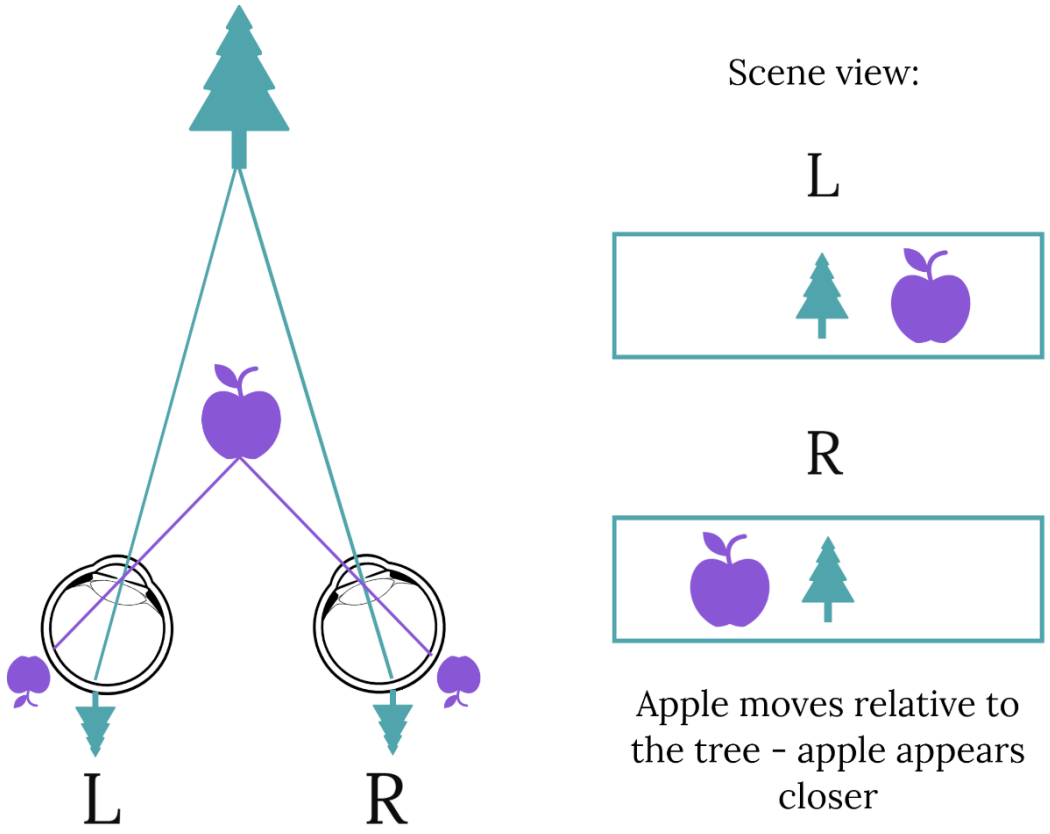


Figure 1.2: Binocular disparity. Image to illustrate how objects at different distances correspond to differing points on the retina, the difference between which creates the cue of disparity.

1.2.1.2 Vergence

The angle at which the two eyes' views meet, or converge, on an object provides information for the cue of vergence (Mon-Williams, Tresilian, & Roberts, 2000), as illustrated in Figure 1.3. When the two eyes meet on an object close to the observer, the angle they subtend is wider than the angle at which they would join on an object further away in the visual field. When viewing an object on the horizon, the lines of sight of the eyes are almost parallel.

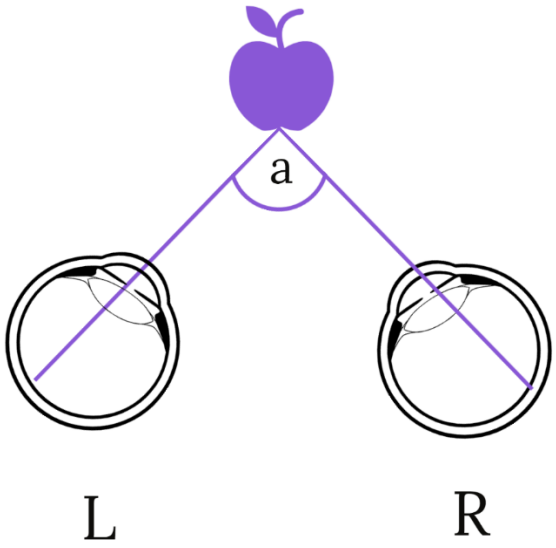


Figure 1.3: Vergence angle. Image illustrating how the vergence angle (a) is created when the two eyes converge on an object.

1.2.2 Monocular cues

1.2.2.1 Accommodation

Although the majority of cues in the monocular category are pictorial in nature, meaning deriving from the visual information in the environment, accommodation is the only monocular cue to derive from information within the person. For accommodation, the ciliary muscles in the eye flex the lens to change its shape and focus the image of an object on the retina in order to provide a clear image. This can

be seen in Figure 1.4, where the shape of the lens changes to accommodate between a far point where the lens is expanded and near point where the lens is contracted. Accommodation can act as a reflex, such as in conjunction with vergence in the accommodation-vergence reflex, but it can also be consciously controlled (Stidwell & Fletcher, 2011). The signal from the ciliary muscles provides an estimate of depth, along with vergence and pupil reflex in the 'near triad' system of mechanisms of eye control (Takeda, Hashimoto, Hiruma, & Fukui, 1999).

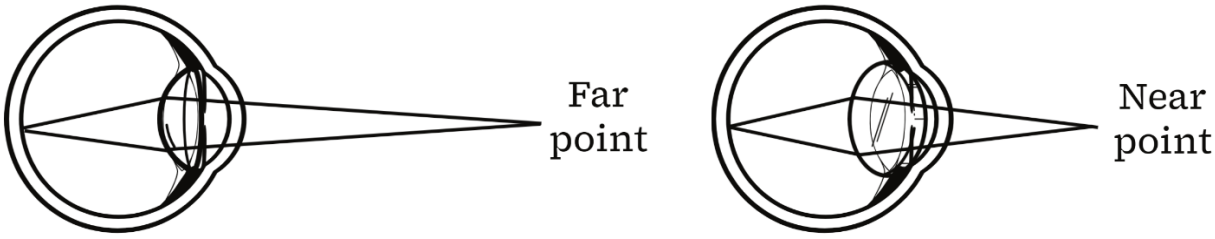


Figure 1.4: Accommodation. Diagram showing how the shape of the lens changes by way of the ciliary muscles to accommodate far and near points.

1.2.2.2 Light and shading

Variations in light and shadow within the visual image provide a cue to depth in the form of shading (Ramachandran, 1988). The visual system assumes the light source to be above the visual field, as is the case for the sun (Mamassian & Goutcher, 2001). This allows us to use shading as a cue to orientation. Figure 1.5a demonstrates how the shadows in the scene infer the position of the sun. In addition, with diffuse light, arriving at each point in the scene from many directions, points closer to the light source in a local region are hit with more of this light, and therefore are lighter than points further away, where less light can reach, such as a hole or crevice. This is known as the 'Dark is Deep' rule (Likova & Tyler, 2003). In Figure

1.5b, the centre of the ball is closest to the observer, and therefore appears brightest.

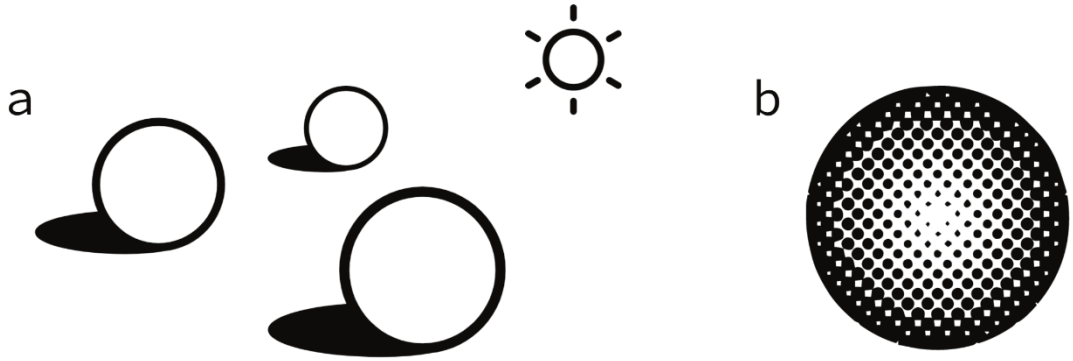


Figure 1.5: Light and shading. Image demonstrating the two shape from shading effects in this work, where (a) shows directional lighting, and (b) shows the 'Dark is Deep' rule.

1.2.2.3 Occlusion or interposition

When viewing a scene, objects that overlap with others in the 2D projection, and whose edges intersect those of the other objects are perceived as in front of these and therefore closer to the observer, through the cue of occlusion or interposition (Koenderink, 1984). Figure 1.6 shows how the pear is partly occluded by the apple, and is therefore seen as further away from the observer.

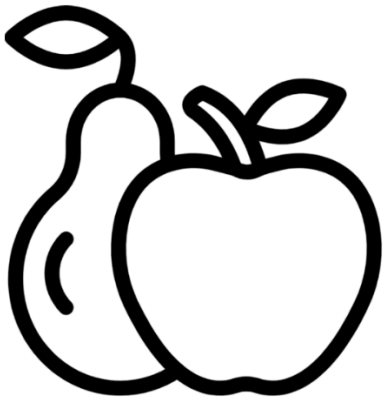


Figure 1.6: Occlusion. Illustrating an example of occlusion, where the pear is partly hidden from view by the apple, suggesting that the apple is closer.

1.2.2.4 Relative size and linear perspective

The cue of relative size, or relative density, describes the idea that similar objects will be viewed as being at different distances if they differ in their retinal size (Knill, 1998). Some objects are known to be a familiar or predictable size, for example humans. When one is viewed as smaller than the other in the retinal image, it is interpreted to be further away. In addition, if two people were to be stood next to each other, and one walks away from the observer, they would extend a smaller size in the visual field but this would be interpreted as them getting further away, not getting smaller, which is known as size or shape constancy (Johnston, 1991). In Figure 1.7:, the road appears to recede into the distance due to the markers in the middle of the road reducing in retinal image size, with the trees also appearing to recede in depth as they get smaller in the image size.

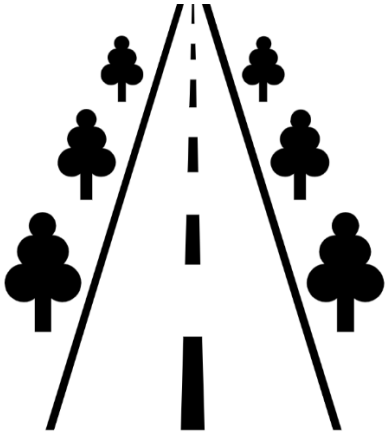


Figure 1.7: Relative size. Image showing an example of relative size, or density, as well as linear perspective.

Linear perspective is also observed in Figure 1.7:, where the sides of the road are perceived as parallel as they recede to the horizon, despite the gap between the lines on the retinal image being wider at the bottom than at the top. Cutting and

Vishton (1995) describe this cue, often cited in the literature, as a combination of other pictorial cues such as relative size and density of texture, which can be seen as the lines in the middle of the road decrease in size and increase in density with distance.

1.2.2.5 Texture gradient

Related to the idea of relative size, but here considering the texture of the surface of an object rather than objects themselves, the cue of texture gradient concerns the change in texture across the object as a measure of depth (Saunders & Backus, 2006; Warren & Mamassian, 2010; Witkin, 1981). When looking at surfaces close to the observer, it creates a 'texture unit' whereby the shape and size of the pattern is measured against this scale. As the observer views surfaces further away, the texture becomes finer in detail, as the size and shape subtended on the retina changes, and the surface appears smoother, in comparison to the texture unit observed on closer surfaces. This scaling of texture across an object gives a measure of depth, as well as of overall object size, as shown in Figure 1.8 (Gibson, 1950).

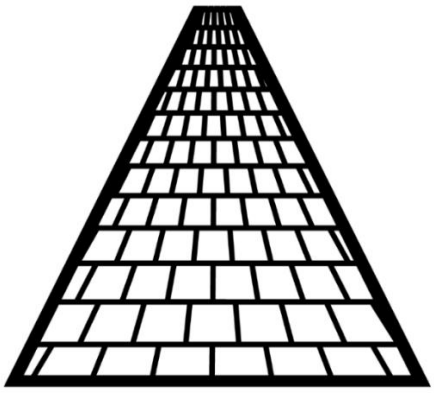


Figure 1.8: Texture gradient. Image showing how the 'texture unit' changes across the image to give the impression of the surface, here a brick path, receding in distance away from the observer.

1.2.3 Retinal and extraretinal cues

Retinal cues are, as the name suggests, cues to depth using information that originates from the retina of the eye. Pictorial cues are all retinal cues as they stem from the visual image. From the above, binocular disparity, light and shading, occlusion/interposition, relative size and linear perspective, and texture gradient are all retinal cues. Extraretinal cues are a source of visual information coming from sources other than the picture on the retina, meaning they come from the person viewing and are biological in nature. For instance, the vergence angle at which the two eyes meet on an object is provided through extraretinal signals originating from the extraocular muscles which control movement of the eyes, and the accommodation is provided by information about the flexing of the ciliary muscles in the eye.

1.3 Combining cues

As highlighted above, judgements of visual depth perception are made up of information provided by many cues. In fact, no other sensory modality contains so many sources of information (Cutting & Vishton, 1995), but when considering the complexity of natural scenes, this makes sense; as Wagner (1985) states, as the quality and quantity of perceptual information increases, the perception of the visual field becomes more veridical to the Euclidean ideal. This means that the visual system can perceive the underlying geometry of scenes and make more accurate judgements by utilising many of these cues to depth at once. Indeed, experiments have shown that increased information generally results in more consistent and

accurate judgements (Kunnapas, 1968; Mather & Smith, 2004; Landy, Maloney, Johnston, & Young, 1995).

When viewing an object, each individual cue reports an estimate to the brain of the depth of that object. For instance, if an observer were to estimate the depth of a bottle of water on a table, the pictorial shape from shading cue may report the depth as 6.5cm. The Minimum Principle, based on Gestalt theory (Wagemans, et al., 2012), suggests that the visual system selects the simplest explanation for a given scene. An example of this is perceptual filling-in, whereby the visual system compensates for gaps in sensory input by perceptually filling in the scene, such as perceiving an object as whole when viewed through a picket fence, despite no corresponding retinal input (Revina & Maus, 2020). Another well-known example is the ‘blind spot’, which we generally do not perceive despite there being no photoreceptors relating to this area (Raman & Sarkar, 2016). However, some criticisms of this theory include a lack of motivation or justification for this principle (Hatfield & Epstein, 1985).

Alternatively, the likelihood theory purports that this is achieved by the cue working out the most likely depth to have produced the corresponding signal in the brain, which from an evolutionary point of view is advantageous to ensure safe interpretation of the world (Feldman, 2009). However, these maximum likelihood estimates are subject to bias and signal noise (Cutting & Vishton, 1995; Hillis, Watt, Landy, & Banks, 2004; Keefe, Hibbard, & Watt, 2011). If the cue is particularly noisy, the reported depth may not be a reliable one. Knill, Kersten and Yuille (2008) present an example of the Bayesian formulation of visual perception using the analogy of a communications system. They present the idea that the signal picked up by the

'receiver' is noisy and bandlimited, meaning an estimation must be made by the 'decoder' on the most likely stimuli to have produced such a signal. They attribute this noise and uncertainty from bandwidth limitation to several things. Firstly, there are physical issues such as improper light diffraction and optical aberrations, where imperfections in the surface of the eye cause the light to be focused incorrectly on the retina. This creates a blurry image that is usually corrected with glasses or lenses. A further physical issue is the high variability in photon emission and absorption, giving rise to additional noise and uncertainty in the visual system. Secondly, additional noise and uncertainty is introduced during the process of receptors in the retina converting light to electrochemical energy. The authors point out that even without this additional noise and uncertainty, the very task of mapping a 3D representation of space to a 2D retinal image results in some loss of specificity.

This noise can cause uncertainty in the maximum likelihood estimates, and therefore different estimates of depth. For example, as before, for an observer viewing a bottle of water on a table, the pictorial shape from shading cue may report a depth of 6.5cm, and the binocular disparity cue may report a depth of 6.2cm. This noise can be reduced or partially cancelled out by utilising multiple cue inputs (Chen, McNamara, Kelly, & Wolbers, 2017). While combining information between senses often requires a recalibration of cues (Ernst, Banks, & Bülthoff, 2000; Burge, Girshick, & Banks, 2010), cues within senses are often highly correlated due to estimating based on the same source of information, such as that from the retina (Hillis, Ernst, Banks, & Landy, 2002).

One possible strategy for the brain to combine differing estimates between cues would be to take an average of the available estimates. However, as previously

stated, each cue has its own level of reliability, which can change dependent on a number of factors, such as the distance to the object and lighting. The question then becomes: how does the visual system combine these sometimes-conflicting estimates of depth? This is where Bayesian statistics can help.

Bayesian approaches are a branch of statistics that can take into account prior knowledge, known as a priori, and can combine this with information from sensory input to create posterior information. Returning to the example of a bottle of water, prior experience of common dimensions of such an object can be used to make sense of the sensory input. In this approach, differences in the reliabilities of cues can be taken into account when combining the information to produce a single estimate. This is achieved by acknowledging which cues have proven to be the most reliable in a certain situation, and weighting their estimates accordingly, with a weighted linear sum of the cues where the weight of each cue is inversely related to its variance (Alais & Burr, 2019).

In their text about the study of perception using the Bayesian approach, Knill and Richards (2008) provide a clear account of how this method can offer a weighted solution to the problem of noisy signals, a short introduction to which shall be presented here. Bayes' formula defines how weighted estimates may be calculated to take into account this a priori and posterior information:

Equation 1.1

$$p(\mathbf{S}|\mathbf{I}) = \frac{p(\mathbf{I}|\mathbf{S})p(\mathbf{S})}{p(\mathbf{I})}$$

Here, p is the prior information, $\mathbf{I}$ is the image viewed by the observer and $\mathbf{S}$ represents scene properties. By treating p(I) as the probability of the occurrence of

an image as a normalising constant (Knill, Kersten, & Yuille, 2008), the above equation can be simplified as:

Equation 1.2

$$p(\mathbf{S}|\mathbf{I}) \propto p(\mathbf{I}|\mathbf{S})p(\mathbf{S})$$

The concepts of uncertain and noisy cues, and combining the Bayesian prior with sensory input to create a posterior estimation are illustrated in Figure 1.9.

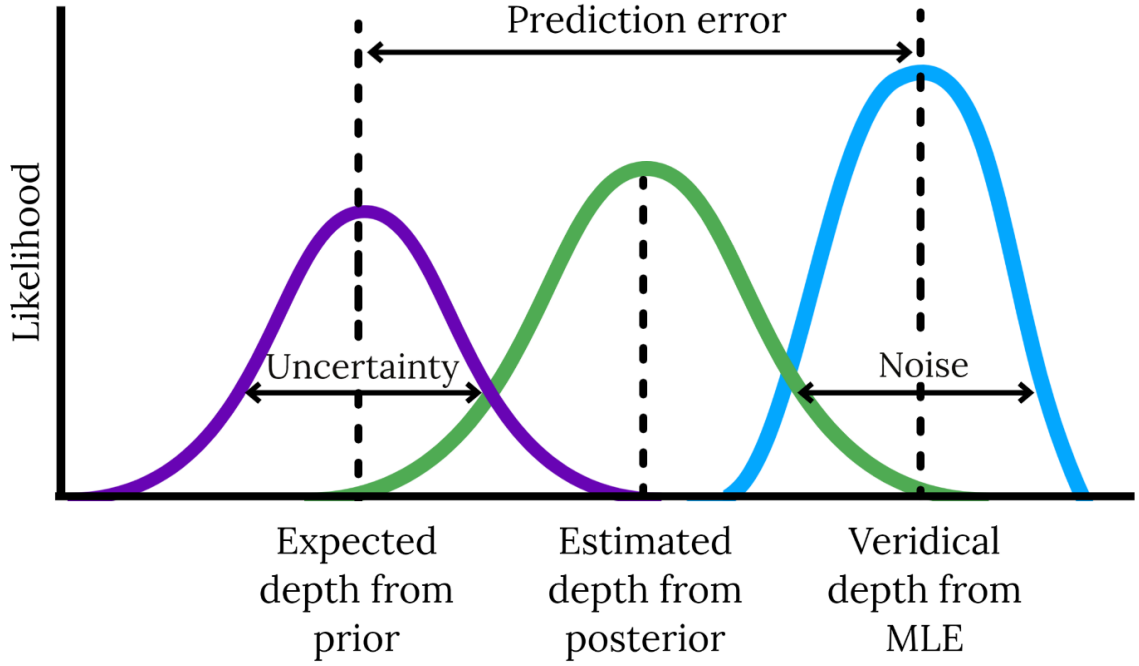


Figure 1.9: Bayesian inference. Illustrating how a noisy MLE and uncertain Bayesian prior are combined into a posterior estimate of depth.

Here, the purple slope shows the expected depth from the Bayesian prior. The spread of the curve shows how uncertain the cue is, with a steeper curve indicating increased likelihood of a cue estimating a certain depth. The blue curve shows the depth estimated from the MLE for the veridical depth. Here, variability represents noise in the sensory input signal, such that a noisier cue will have a shallower curve, and a less noisy cue will report a certain depth with a greater likelihood from a

steeper curve. The difference between the most likely depth estimates from the Bayesian prior and the MLE is the prediction error. Under Bayesian inference, the prior knowledge and sensory input are combined into a posterior depth estimate, as shown in green.

The information above provides a way to integrate cue weightings into the depth estimate average. Hibbard, van Dam and Scarfe (2020) present an equation that accounts for the individual weightings given to cues in the depth estimate, allowing for comparison between cues when bias is purposely introduced in order to measure the weightings. In their work, the weighting for binocular disparity is compared against all other available cues, but this same reasoning can be used for any combination of cues, and the equation can be adapted as such:

Equation 1.3

$$D_P = W_{\mathrm{a}} D_{\mathrm{a}} + (1 - W_{\mathrm{a}}) D_{\mathrm{b}}$$

This estimates the perceived depth (D_P), predicted by a combination of the depth estimated by the manipulated cue (D_{a}) according to its weighting (W_{a}) and the estimated depth from all other cues available in the scene (D_{b}). In some traditional studies, the other cues may only represent a single other cue, but for the work presented here, this denotes a combination of all other scene cues available. This premise of cue combination is illustrated in Figure 1.10.

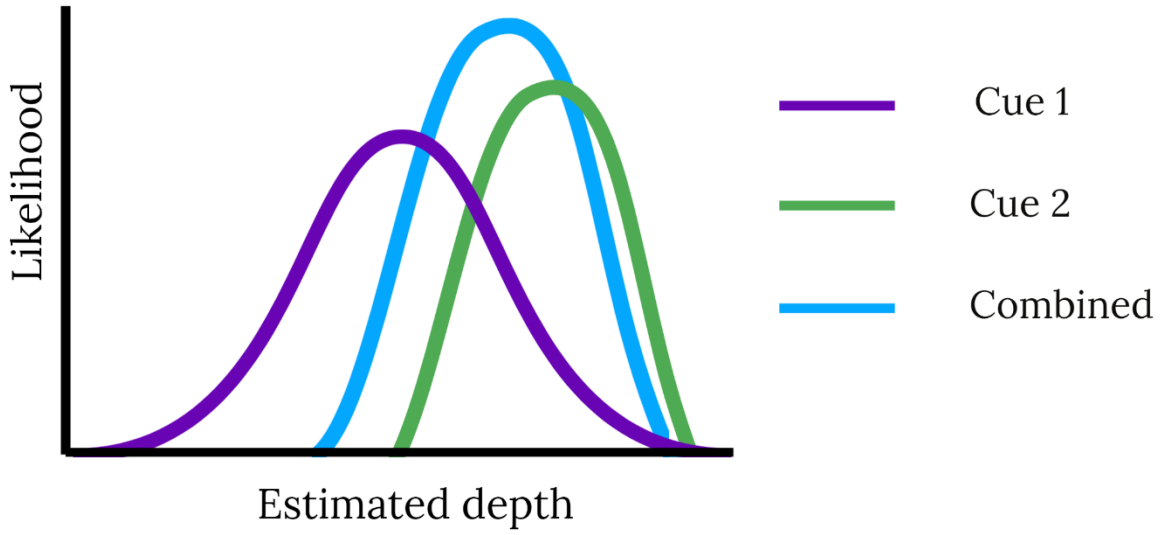


Figure 1.10: Cue combination. Illustrating how the information from two cues (or any combination of cues) can be combined into a more reliable estimate of depth through cue weighting.

Here, two cues are combined into a single weighted depth estimate. The purple curve shows a cue that is less reliable, denoted by the shallower curve. The green curve exhibits less variability, and therefore a cue with increased reliability, which would be weighted more in the weighted depth estimate. The blue curve shows the combination of depth estimates from these two cues based on their individual weightings.

The theory shown in this illustration can be expanded to include any number of cues, or groups of cues. For instance, if the purple curve is taken to represent a combination of pictorial cues, and the green slope represents more reliable binocular cues, it can be seen that the cue combination would result in an estimate closer to that of binocular cues given their increased reliability and therefore increased weighting, but also taking into account the depth reported by pictorial cues as well.

Weightings for cues are complementary and are summed to a value of 1 (Chen, McNamara, Kelly, & Wolbers, 2017). Equation 1.3 assumes that the cues are

weighted despite reporting conflicting estimates (Muller, Brenner, & Smeets, 2009), despite some evidence that the cue in conflict with the others may be weighted less (Landy, Maloney, Johnston, & Young, 1995), and that the weighted average may not incorporate cues that are too discrepant (Cheng, Shettleworth, Huttenlocher, & Rieser, 2007; van Ee, van Dam, & Erkelens, 2002). These equations form the basis of the cue perturbation and manipulation studies contained within this work, and allow for measurement of the individual weightings of cues when combined with multiple others to assess their contribution to the depth estimate.

1.3.1 The 'ideal Bayesian observer'

The Bayesian model makes certain assumptions about the information provided by cues (Knill & Richards, 2008). For instance, one assumption is that the likelihood functions for individual cues are Gaussian. Additionally, is it assumed that sources of noise in each cue are independent of one another. Therefore, under maximum likelihood estimation, each aspect of sensory input is assumed to report an unbiased or accurate, independent estimate, which can be compiled into a singular representation of 3D scene structure (Landy, Maloney, Johnston, & Young, 1995). These assumptions represent the ideal observer (Knill & Richards, 2008).

However, evidence suggests that these assumptions are not always met, and it has been argued that systematic errors in perception, rather than being indicative of noisy or impoverished signals, may instead represent misunderstood parameters of the encoding of sensory signals themselves (Vishwanath, 2022). Despite this, the Ideal Observer is a convenient shorthand for expected performance based on theoretical predictions. Therefore, this work will refer back to this mystery participant

when comparing expected and actual perceptual judgements, as a way of relating the empirical findings to the theory upon which it is based.

1.4 Limitations

Much work has been undertaken to examine how these cues work, and how reliable they are as a source of information to the brain. However, there are some limitations to these methods that will be discussed here.

Often these studies look at a single cue, or the interaction of two. This is often for practical reasons, as when designing an experiment, two cues can be tested for individually, and then compared to investigate how they interact. For instance, when investigating the cue of convergence, cue isolation allows for a direct study of the vergence cue, without including other confounding cues such as binocular disparity. Viguier and colleagues (2001), whose work is covered in depth in Chapter 3, presented observers with light-emitting diodes with either retinal disparity cues, or extraretinal vergence, available upon which to make distance judgements. This was achieved by asking observers to move the test diode to either the same, double or half the distance of the reference diode. By isolating extraretinal cues in this manner, they were able to work out the relative weightings of vergence compared to when retinal cues were available as well, given the difference in distance reported under each condition.

While studies such as this provide insight into relative weightings within the depth estimate, showing the interaction of two or more cues, little work has been presented that discretely measures these weightings. The experiments contained within this thesis are designed to create a direct measure of weightings for cues

within the depth estimate. Additionally, conflicts between cues, such as those purposely introduced in this kind of study, can result in unnatural or uncomfortable viewing conditions, creating visual discomfort for the observer (Lambooij, Ijsselsteijn, & Heynderickx, 2007), such as common symptoms like motion sickness, eye strain and double vision experienced when using a poorly-calibrated and unstable 3D stereoscopic display (Hwang & Peli, 2014).

Another criticism of studies of this kind is that the stimuli used are overly-simplified in comparison to naturalistic objects in natural scenes. In order to isolate information provided by individual depth cues, many studies present simplistic stimuli, such as dots or lines, often in a darkened laboratory, to reduce extraneous depth information such as that reported by accommodation of the eyes on the surface of a screen, or by knowing how far away a reference point is, such as knowing the dimensions of the room. Some psychophysical studies have utilised physical objects to create stimuli, such as a light emitting diode (Bradshaw, Parton, & Glennerster, 2000), or printed random dot stereograms, viewed with the use of a stereoscope device (Julesz, 1971). With advances in the power and availability of computers, studies can now present similar stimuli using computer screens, for instance, random dot stereograms can be rendered in 3D for stereoscopic viewing using advanced technology (Hibbard, Goutcher, Hornsey, Hunter, & Scarfe, 2023).

However, these simplistic stimuli presented to participants with reduced cue information, is far removed from natural viewing conditions, as the human visual system is highly complex (Nadenau, Reichel, & Kunt, 2002). In natural viewing, one has access to a full complement of depth cues that provide rich information on which to base estimates (Hibbard, 2021). In fact, it has been shown that experiments in a

laboratory setting can easily make perception of a scene challenging for the observer (Sedgwick, 1986). Yuille and Bülthoff (2008) make the highly relevant point that although previous research has found less reliable results for natural images compared to synthetic images, the human visual system actually performs much better and is more easily able to perceive depth using multiple cues in natural scenes than with isolated cues with synthetic images. This means there is currently relatively little understanding of how cues contribute to the perception of depth in complex natural scenes (Hibbard, Hornsey, & Asher, 2022).

The work presented here aims to address this criticism of previous work by presenting participants with stimuli that represent complex scenes of realistic objects under more natural viewing conditions, by utilising recent advancements in methods and technology, details of which are covered in the next chapter. The work follows the examples of Koenderink and colleagues (Koenderink, van Doorn, & Kappers, 1992; 1995; 1996; Koenderink, van Doorn, Kappers, & Todd, 2001), who presented observers with photographs of natural stimuli under multi- or full-cue viewing conditions, and measured the relative contribution of cues by introducing bias within a cue while holding the others constant, creating an ‘operating point’. This allows for manipulation of cues to assess their contribution, while using naturalistic stimuli.

However, these methods raise the issue of how to manipulate just the cue of interest while holding the others constant. Hibbard, Hornsey and Asher (2022) detail how experiments that probe pictorial space do not directly measure 3D space, given that pictorial space exists as a representation of 3D space within the mind of the observer. Additionally, while photographs contain rich pictorial information, they do not provide information from cues such as binocular disparity and vergence

(Hibbard, Hornsey, & Asher, 2022). The use of artificial stimuli viewed stereoscopically can afford sufficient experimental control (Rust & Movshon, 2005), while accounting for these issues, further details of which are covered in Chapter 2: Methods.

1.5 Perception of surfaces

Given the limitations outlined above, a major drive for this thesis is to explore the contribution of depth cues within complex scenes of naturalistic stimuli, presented under natural viewing conditions. The locations of points in space are denoted using the Cartesian coordinates of x, y and z. However, Gibson (1950) criticises this approach for the theory of vision, stating that we perceive surfaces, not just points in space. As such, the next step is to understand how we perceive these surfaces.

Perceiving surfaces is an important aspect of making judgements of depth, particularly for this thesis, where the focus is on natural objects in naturalistic scenes. Within the visual field, there are many different types of surfaces, each with differing considerations. For example, knowing distance and safe navigation of one's surroundings depends on successful interpretation of the ground plane.

Hugely influential in this area of visual perception is Gibson. He outlined his theory of surface perception as the identification of the distance, depth and orientation of an object purely by means of the difference in pattern of an array of surfaces (Gibson, 1950). He detailed a list of essential properties by which a surface may be considered 'determinate'; that is, that the surface is one which makes up part of the ordinary visual scene, such as the face of an object, the ground, or a building.

An indeterminate surface is classified as one beyond this visual scene, such as a cloudless sky (Gibson, 1950). The list of eight properties includes if the surfaces appears solid, the colour including the brightness, hue and saturation, whether the surface is illuminated or in shadow, and the amount and direction of the slant of the surface.

Another type of surface important to navigating the world is the surface of objects themselves. Accurate perception of object shape allows for recognition of the object, as well as being a vital part of reaching and grabbing in order to interact with the object (Melmoth & Grant, 2006; Watt & Bradshaw, 2000). While there are many methods of measurement for psychophysical studies (Anderson, 1970), methods such as a gauge figure task have proven successful for probing local surface attitude in pictorial space (Koenderink, van Doorn, & Kappers, 1992; 1995; 1996; Nefs, 2008). While the specifics for this method are covered extensively in Chapter 2: Methods, the premise of the task allows for the capture of the slant and tilt of surfaces, which define their orientation. Studies of this kind have used the orientation of responses to recreate a depth map of the perceived surface, showing how the local settings are consistent with global perception of the scene (Koenderink, 1998; Koenderink, van Doorn, & Kappers, 1995).

1.6 Measuring the contributions of depth cues in complex natural scenes

In summary, limitations of depth cue combination studies include overly-simplified stimuli and viewing conditions, with many studies reporting relative cue interactions, rather than providing a discrete measure of this weighting. As such, the overall aim

of this thesis is to measure the contributions of depth cues in complex natural scenes. Specifically, this is achieved through simulating natural full-cue viewing conditions, for complex scenes of naturalistic stimuli, and introducing conflict between cues in order to measure their contributions to the overall depth estimate. Each chapter provides key insight into a different topic highlighted in this introduction chapter. This introduction included a broad outline of the visual cues available for the perception of depth, and how they are recovered, in order to introduce the reader to the problems with cue combination research that form the rationale for this thesis. Each empirical chapter provides a more detailed discussion of the most relevant background concepts and literature for each study. The remaining chapters are summarised as follows:

- Chapter 2: A number of options for psychophysical study are available, but some are more suited to addressing the concerns outlined above. Chapter 2 considers these, and outlines the methods designed to address these issues raised in the literature, and how these can be used to explore depth cue contributions in complex natural scenes.

- Chapter 3: In order to scale disparity information appropriately to judge object depth, an estimate of object distance is required. This distance estimate can be provided by the angle at which the two eyes convergence on the object. Chapter 3 explores the contribution of the cue of vergence to shape constancy, for depth judgements over distance.

- Chapter 4: Shading cues provide pictorial depth information, and are interpreted relative to the light source in a scene. Chapter 4 explores the ‘Dark is Deep’ rule for shape from shading, investigating the contribution of pictorial and binocular cues under various luminance and viewing distance conditions for complex natural scenes.

- Chapter 5: The gauge figure task is an intuitive method for capturing perceived surface shape to assess depth perception. However, parameters of gauge size and distance scaling remain arbitrary within the literature. Chapter 5 explores the impact of gauge size and distance scaling on the captured shape for rough and smooth objects, and for local and global viewing of complex natural scenes.

- Chapter 6: Binocular disparity is often cited as one of the most reliable, and therefore most relied upon, cues to depth perception. Chapter 6 seeks to measure this contribution directly in a series of cue-perturbation experiments exploring binocular cues at close and far distances, and for natural objects in isolation, as well as in cluttered and complex natural scenes.

- Chapter 7: The previous chapter sought to directly measure the contribution of binocular cues. Chapter 7 presents a similar cue-perturbation method, this time for pictorial cues, seeking to measure their direct contribution to the weighted depth estimate for natural scenes.

- Chapter 8: The final experimental chapter presents a real-world application of some observations in this thesis, in the form of measuring the contribution of cues from a pictorial depth manipulation using a digital makeup filter.

- Chapter 9: The final chapter in this work summarises the experiments in the context of the overall aim of the thesis, presents the main findings, and proposes future directions for this work.

2 Methods

2.1 Limits of traditional methods

Having outlined the key theoretical considerations for this thesis in the previous chapter, here, an overview of the methods used to explore these are presented. This chapter contains specifics relating to the methods used, with summarised information covered in the context of the relevant chapters, where referring back would break the flow of the text.

The methods outlined here have been explored and selected to address some of the key issues raised in Chapter 1. In summary, this thesis addresses three main issues. Firstly, in order to measure the contribution of depth cues, many studies reduce those available down to either the cue of interest in isolation, or a pair of cues, neither of which simulates natural viewing and can in fact make decoding scenes overly-challenging for the visual system compared to naturalistic viewing (Sedgwick, 1986). Likewise, simplified stimuli under multi- or full-cue viewing do not capture the complexity of the problem faced by the visual system for natural viewing of complex naturalistic scenes (Yuille & Bülthoff, 2008). Finally, studies that do address natural viewing of natural objects do so with photographic representations, which present their own issues, and are often centred on smooth real or simulated objects that lack complexity or the context of a scene to replicate true-to-life viewing.

2.2 Use of advanced technology

The criticism of previous work highlighted above, in that stimuli were presented in an overly-simplified form, is due in part to the need to maintain a high level of

control over viewing conditions, which is not easily obtainable with traditional methods. For instance, it is important to control for the luminance level of the laboratory, size and shape of stimuli, and physical distance of the participant in order to measure exactly what is intended. In a typical laboratory setting, these issues can be addressed by implementing stringent control of these variables, such as by using simplified stimuli. Presenting natural objects in these conditions would clearly present a challenge; if a fruit or vegetable was presented to a participant, this would not be replicable much beyond the single laboratory session, as the item would change in colour and shape due to decomposition. Likewise, presenting a physical object which would appear different depending on the viewing angle, gives rise to problems of replicability if the object should fall or be nudged during testing.

An option previously used in depth perception studies to mitigate these issues was to use photographs of natural objects or scenes. For instance, Koenderink, van Doorn and Kappers (1992) presented observers with pictures of a smooth sculpture, and probed observers' perception of the local surface attitude via a gauge figure task, which is covered in more depth below. However, many of these objects are smooth and simplistic, and do not fully reflect the complexities of naturalistic scenes. They also do not allow for direct manipulation of certain parameters such as physical shape or size. The use of images in studies that incorporate binocular cues also presents issues with converging cameras, with misalignment potentially resulting in distorted stereoscopic depth (Allison, 2004).

One way to incorporate natural objects whilst maintaining a high level of control is by using several recent advances in technology, that allow for the presentation of natural objects that invoke natural viewing conditions, whilst ensuring uniformity of

size, shape and colour. Rendered objects presented in cluttered natural scenes in this way allow for full customisation and manipulation of physical aspects, including but not limited to size, shape, surface colour, surface texture, shading and scene lighting. The studies presented here take advantage of a range of different technologies to isolate and perturb various cues to depth in order to measure their contribution through their weighting in the overall depth estimate. These are covered in more depth within their relevant chapters, but illustrative examples will be presented here.

For instance, binocular cues are manipulated in two chapters within this work. Chapter 3 isolates the cue of vergence from other typically confounding cues such as binocular disparity, by presenting one image independently to each eye, ensuring no disparity information is present, and that results are from the cue of vergence. This is done using a 3D screen system called VIEWPixx™, the various products of which were designed to replace traditional CRT displays in vision science laboratories (VPixx Technologies, Quebec). The system is made up of a 3D screen, 120Hz LCD goggles and an infrared (IR) emitter that synchronises the glasses with the 3D stimuli on the screen, shown in Figure 2.1.

Figure 2.1: 3DPixx system. Image taken from https://vpixx.com/products/3dpixx/ showing the infrared (IR) emitter and 120Hz LCD glasses used to synchronise 3D stimuli on screen with the glasses.

One of the major benefits of using this technology over traditional display methods is the ability to easily display images dichoptically, where one image goes to the left eye, and one to the right. In doing so, it is possible to manipulate various depth cues and simulate unnatural viewing conditions in ways which would not be otherwise achievable.

A gain on binocular disparity is created in the stimuli in Chapter 6 by rendering left and right eye viewpoints and presenting these dichoptically using the VIEWPixx goggles. This simulates a different IOD for the observer, which is not physically possible, but is made easy and convenient with the use of this technology.

Pictorial cues are likewise manipulated using various technologies in this work. In Chapter 7, shape from shading and occlusion are manipulated by 'stretching and squashing' the 3D models, manipulating their depth relative to veridical settings. This chapter also manipulates the surface colour of objects by painting them uniformly grey in order to isolate the cues of interest, a process that is both convenient and reliable with the use of technology. Chapter 8 also manipulates shape from shading, this time by applying digital makeup 'filters' through use of software to apply targeted shadowing to manipulate shape from shading.

While this technology offers huge benefits to address the issues raised, there are some limitations. For instance, focus and accommodation cues are not manipulated in the above designs, and therefore would be in conflict with the perturbed cues, reporting the scene to be flat given the accommodation of the screen. This effect is known as vergence-accommodation conflict (Kim, Kane, & Banks, 2014), and has been observed for 3D screens (Karpicka & Howarth, 2013),

3D televisions (Hoffman, Girshick, Akeley, & Banks, 2008), and Augmented and Virtual Reality (Wang & Lin, 2021).

The online experiments presented in Chapters 5, 7, and 8 are presented on participants' own computer or laptop screens. Whilst binocular cues are not of interest and therefore are not manipulated in these experiments, information from binocular cues was still available, as these were not completed monocularly. When using a screen, binocular cues, similarly as above, would directly conflict with the information presented from pictorial cues, reporting the scene to be flat. However, it has been suggested that as this cue is not present within pictorial space, it may be ignored within the weighted cue estimate (Koenderink, 2012).

2.3 Creating naturalistic 3D models

2.3.1 Scanning technology and procedures

Most of the naturalistic stimuli presented in this work were captured using laser scanning technology. The exception is the faces stimuli used in Chapter 8, which will be covered in detail below. In total, three different scanners were used to capture the models, details of which shall be outlined.

2.3.1.1 NextEngine 3D scanner and multidrive turntable

The NextEngine 3D scanner was used to capture objects using a multidrive turntable. This scanner has settings to capture a single scan from a fixed angle, a 'bracket' which completes a scan of three angles, or a full 360-degree scan of the object achieved by using various tilt settings, starting at 0 degrees tilt and ranging

from -35 degrees to + 45 degrees tilt, to create a full mesh. Objects are gripped in place on the rotating platform, and captured with the scanner which consists of a flash, a camera, and a laser to detect depth of surfaces. The scanner has multiple division settings, which indicate the number of rotations the base will complete during the 360-degree scanning, where more divisions gives the scan more data but takes longer to complete. This is between four and 16 times, with the meshes created for this work using eight divisions. Scans are then processed using the proprietary NextEngine Scan Studio HD software to create 3D meshes for use in rendering software to create the experiments.

2.3.1.2 Artec Eva™

The Eva is a handheld 3D scanner suitable for making 3D models of medium sized objects such as the wheel of a car (Artec 3D, Luxembourg). It uses a flashbulb as its 3D light source, with an additional array of 12 white LEDs for the 2D light source. The device captures the image information using structured light scanning technology, which poses no health risks, so the scanner is even suitable for creating a human bust. The scanner retails for around €13,700, and claims to be suitable for scanning black and shiny surfaces which are often difficult to capture using light-based scanners, due to the possibility of glare and breaks in the surface captured (Kęsik, Żyła, Montusiewicz, Neamtu, & Juszczyk, 2023). The accuracy of the Eva is reported as up to 0.1mm, with accuracy reducing by up to 0.3mm per metre scanning distance. It has a 3D resolution of up to 0.2mm and is capable of capturing texture detail up to a resolution of 1.3 megapixels (Mpx) and colour depth information of up to 24 bits per pixel (bpp). Its working range is between 0.4m and 1m, and it is

capable of capturing up to 61,000cm^3. Objects are scanned on a smooth, flat surface, such as a table, by moving the scanner around the object in a steady arc. To create full meshes, both sides of the objects are scanned, and identifiable points on the objects referred to as 'landmarks' are used to align scans to fuse them into a water-tight model within Artec Studio, the proprietary scan processing software.

2.3.1.3 Artec Space Spider™

The Space Spider is a similar technology to the Eva, in that both scanners use structured light scanning technology to capture images. It differs from the Eva in that it is more suited to capturing small objects or high levels of precise detail, such as coins or a human ear (Artec 3D, Luxembourg). This scanner also uses an array of white LEDs for the 2D light source, but this time six compared to the Eva's 12, and it utilises a blue LED for the 3D light source instead of a flashbulb. At €19,700 it is more expensive than the Eva, but with this it is more accurate, giving an accuracy of up to 0.05mm, with a 0.1mm 3D resolution. The working range of the Space Spider is closer than the Eva at 0.2m to 0.3m and it is capable of capturing a volume up to 2,000cm^3, losing up to 0.3mm accuracy per metre of scanning distance. The Spider can capture texture information with a resolution of up to 1.3Mpx, and depth of colour information of 24bpp. Much like the Eva, scans are created by moving the Spider around the object in a smooth arc, capturing both sides of the model against a smooth, flat surface and aligning scans to create a water-tight 3D model within Artec Studio as above.

2.4 Producing 3D scenes

2.4.1 Creating 3D meshes

Figure 2.2 below shows a selection of results obtained from using the scanners detailed above.

Figure 2.2: Results of scanning. Showing a selection of naturalistic fruit and vegetable 3D models created using the scanners listed above. The sweetcorn was scanned using the Artec Spider, and the remaining models were scanned on the NextEngine scanner using the turntable.

Due to the availability of the different scanning technologies, and the perishable nature of the stimuli captured in this work, it was not possible to capture data from the same fruit or vegetable on each of the scanners used for the stimuli within this thesis. However, given the similar high resolution and accuracy levels between the types of scanners, this did not affect experimental design, and therefore a variety of stimuli created between the scanner types was used in this work.

2.4.2 Rendering the scenes

The majority of stimuli used in these experiments within this work were rendered in MATLAB using OpenGL and the Psychophysics Toolbox extension (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). OpenGL lighting modules were used to illuminate scenes, with the specific component magnitudes reported in each chapter.

Stimuli were either rendered for viewing binocularly, or from the 'cyclopean' viewpoint, shown in Figure 2.3, depending on the requirements of the experiments. For instance, some work was presented with differing left and right eye views, such as in Chapter 6 which explores the contribution of binocular disparity to the depth estimate. This was achieved by rendering two view points, one for the left eye and one for the right, separated by an inter-camera distance simulating observer interocular distance (IOD). For the work in Chapter 6, differing view points for both the left and right eye were rendered to cover a range of typical adult IODs between 52mm and 78mm as outlined by Dodgson (2004) for use in the experiments, creating stereoscopic pairs of images to accurately simulate disparity specified by the 3D structure of the scene.

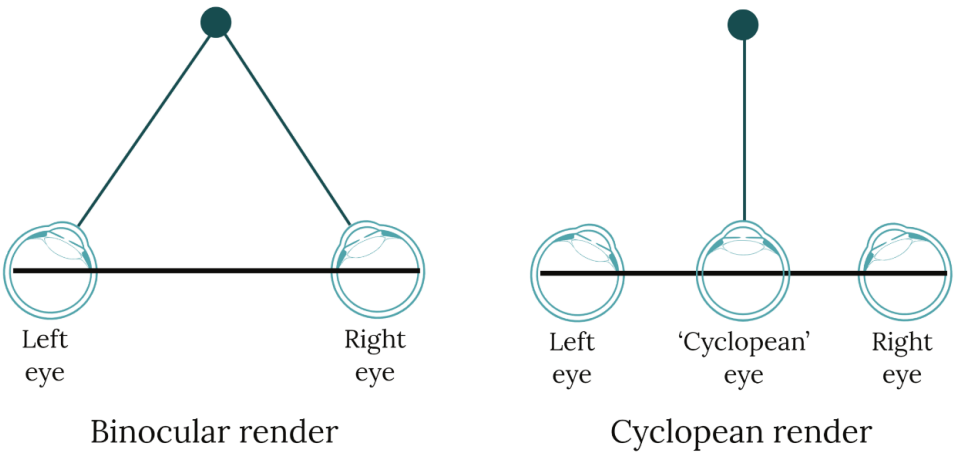


Figure 2.3: Binocular versus cyclopean viewpoint rendering.

The cyclopean view was rendered using the midpoint along the axis that connects the two eyes, and simulates monocular viewing from the imagined 'cyclopean' eye in the middle of head (Stidwell & Fletcher, 2011). This approach was taken where binocular disparity information was not isolated and manipulated, such as for the pictorial cue experiments in Chapters 5, 7 and 8.

2.4.3 Face stimuli

The faces used in Chapter 8 are the only artificially-generated complex stimuli used in this work, and were processed in a slightly different way to the other naturalistic models. These were generated using random face generating software FaceGen (Singular Inversions Inc, Toronto), and manipulated to create experimental conditions within image manipulation software FaceFilter Pro (Reallusion, California). These faces were rendered in 3D rendering software Blender (Hess, 2010) using a central light source placed directionally above and in front of the face to reduce additional shape from shading.

2.5 Introduction of experimental methods

Four main psychophysical tasks were used in this work to gauge perception of depth from various cues. Three of these are limited to individual chapters, and will be briefly outlined here, and covered in more depth within the relevant chapter. The fourth method is used extensively in four chapters, and will be explained in detail within this section to avoid repetition within these.

In Chapter 3, a nonius line task, typically used to measure the precision of vergence judgements (Jaschinski, 1997; Jaschinski, Bröde, & Griefahn, 1999; Chopin, Levi, Knill, & Bavelier, 2016; Chopin, Levi, & Bavelier, 2017), is employed to create a measure of how certain observers are with using the cue of vergence to make depth judgements. This measure of the certainty of vergence is compared to observers' depth perception by way of the second task in this work, recovering the depth of a triangle represented by three dots in 3D space (Bradshaw, Parton, & Glennerster, 2000). Judgements for these are compared over distance to assess shape constancy.

Chapter 4 introduces the third task of the work. Observers are presented with a scene upon which two dots have been superimposed, and are tasked with estimating the 3D distance between them (Lovell, Bloj, & Harris, 2012). This provides a measure of depth, used in Chapter 4 to investigate within- and between-object distance judgements under various luminance manipulations (Hibbard, Goutcher, Hornsey, Hunter, & Scarfe, 2023).

Here, the final task, which is used in Chapters 5, 6, 7 and 8, is a gauge figure task. This is covered extensively below.

2.5.1 Gauge figure task

One way to probe the local surface attitude of images is with a surface normal gauge figure task. A gauge figure consists of a circle and a rod (as per Figure 2.4:), and is superimposed onto the surface of an image so that it can be manipulated such that the circle element appears painted flat on the surface of the object in

pictorial space, with the rod sticking straight up from the object at a right angle (Koenderink, van Doorn, & Kappers, 1995).

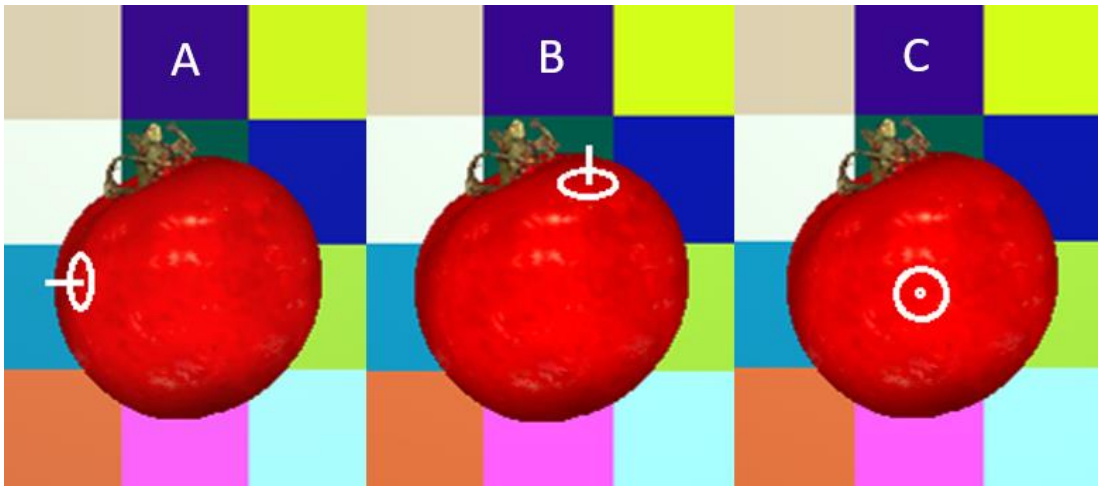


Figure 2.4: Gauge rotations. Image showing examples of how a gauge figure might be set at various locations on the surface of an object.

The orientation of the gauge figure to perceptually 'fit' the apparent surface (Koenderink, van Doorn, & Kappers, 1992) provides a measure of slant and tilt. The terms slant and tilt can have various definitions. To avoid confusion, these terms will be specifically defined here for use in this work.

Figure 2.5: shows possible gauge figure settings and their corresponding slant and tilt values. The slant value will be defined as the size of the angle in degrees by which the surface is rotated away from the observer on the horizontal axis, and includes positive and negative values, described by Gibson (1950) as extremes creating a ceiling and floor. These can range from 0 degrees, where the surface is exactly perpendicular to the line of sight, which creates a gauge figure that resembles a bullseye (as in Figure 2.4:c), through to 90 degrees where the surface runs parallel with the line of sight and is no longer directly visible, such as seen around the edge of the diagram in Figure 2.5:. When holding tilt constant, the same

slant value across an entire object would show a surface which is flat in all areas; where the slant values differ, this denotes curves or bends on the surfaces of the object (Gibson, 1950).

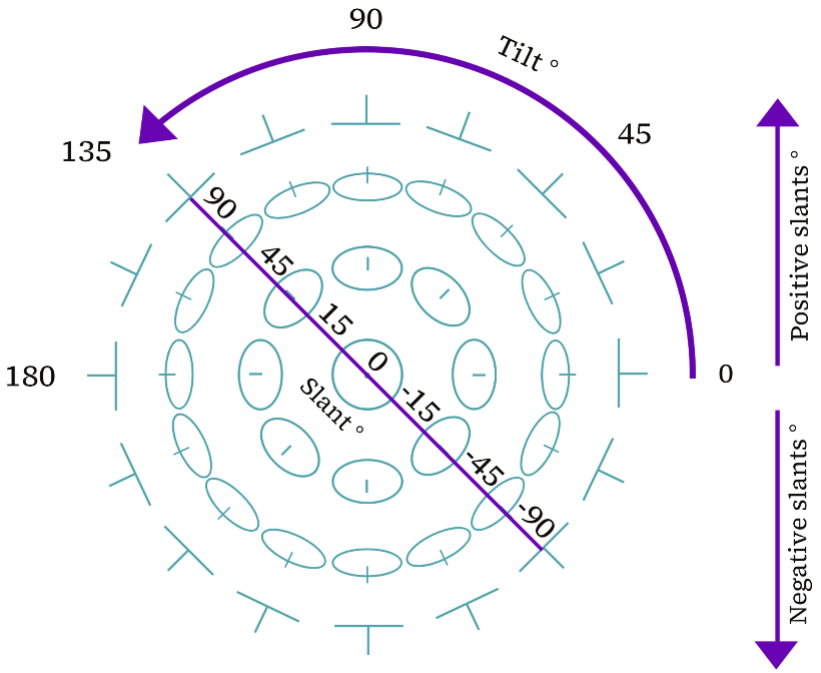


Figure 2.5: Gauge slant and tilt. Diagram showing slant settings, tilt settings, and how these can be combined to describe the local attitude of any surface.

The definition of tilt as used in this work is the axis around which slant rotates, as shown in Figure 2.5:, sometimes also referred to as the direction of slant (Gibson, 1950; Stevens, 1983). Here, a value of 0 degrees shows a vertical tilt axis around which a surface may slant by any degree as outlined above. This could be thought of as a flagpole around which a flag can move, depending on the direction of the wind. A 45 degree or 135 degree tilt creates a diagonal slant axis, and a tilt of 90 degrees creates a horizontal slant axis. A value here of 180 degrees also creates a vertical axis, but this is the opposite of that created at 0 degrees, as if the flagpole were upside-down. When taken together, the slant and tilt provide information on the

perceived local surface attitude, or depth gradient (Koenderink, van Doorn, & Kappers, 1996).

Stevens (1983) highlights the values of slant and tilt as the only two needed to identify the orientation of a patch of surface relative to the line of sight, referencing Gibson's (1950) statement that the magnitude and direction of slant are the two degrees of freedom of surface orientation. Gibson (1950) also proposed that identifying objects in a scene is less about identifying the form of the object itself, and more about consistent patterns between the object and the background, stating that perception in general can be reduced simply to the perception of a series of surfaces, and that orientation, distance and depth may all be derived from the details of such surfaces.

The gauge figure task is used in several chapters in this body of work due to several benefits over the other measurements listed above. Firstly, the gauge figure task is described in the literature as an intuitive task for measuring perception. Koenderink, van Doorn, Kappers and Todd (2001) report that observers rate the task as more 'natural' than several other tasks, a view echoed by Nefs (2008). Additionally, the task allows for the capture of perceived local surface orientation, which accounts for the argument that we perceive surfaces, not points in space (Gibson, 1950). Finally, as discussed below, the use of a gauge figure task to capture the perceived slant and tilt of surfaces allows for a reconstruction of the perceived global surface. Specifics of this shall now be discussed.

2.5.2 Recreating a 3D mesh

Several chapters in this work use a method of recreating a 3D mesh from the depth gradients calculated with slant and tilt settings in a gauge figure task. The specifics will be presented here in detail, and summarised within the context of the relevant chapters for ease of reading.

Nefs (2008) used a gauge figure task to probe the perceived surface of globular convex objects, with or without specular highlights, using pairs of stereo images. A red gauge figure probe was superimposed on the left-hand image of the pair, which observers rotated until it appeared painted flat to the surface of the object. In this paper, Nefs clearly outlines the method used to create a mesh surface from a set of depth gradients, which has been used successfully previously in other works (Koenderink, van Doorn, & Kappers, 1992), and will be used in the present work.

Nefs (2008) describes how the x and y coordinates for each of the gauge probe positions create a point, where three points refer to as a set of vertices, with the straight line connecting each pair of points an edge. Figure 2.6: shows how three vertices (V_1, V_2, V_3) define a face, as shown by the shaded triangle. The middle of this triangle, or the centre of mass, is the barycentre. In the experiments in this work that use this triangulated face mesh, this barycentre is the point at which the gauge figure probes the local surface attitude.

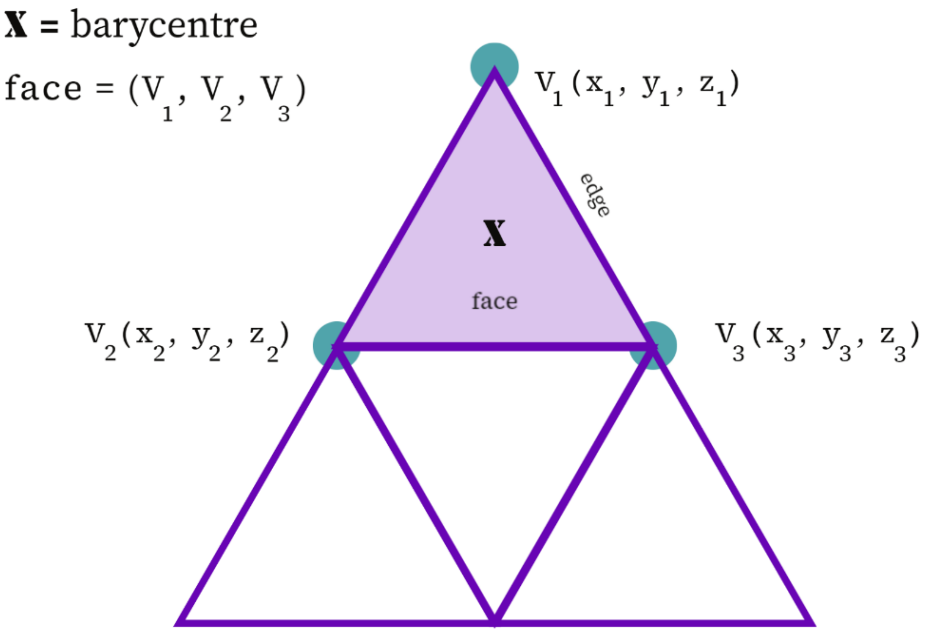


Figure 2.6: Barycentres. Diagram showing how a face (shaded area) is defined by a set of three vertices (V1, V2, V3), with the barycentre being the middle point or centre of mass of the created triangle.

Nefs used these gauge settings between two vertices to calculate the depth gradient, which shows the change in depth given a change in horizontal or vertical direction. This calculates the gradient as the change in a dimension, written as delta (δ). When calculating the steepness of a slope in two dimensions, this is calculated as the difference in the x and y dimensions between two points, see Equation 2.1.

Equation 2.1

$$slope = \frac{change\ in\ y}{change\ in\ x} \quad \text{or} \quad slope = \frac{\delta y}{\delta x}$$

This can then be extended to calculate the slope of a surface in three dimensions by incorporating depth (z) in Equation 2.2:

Equation 2.2

$$\{\delta z/\delta x, \quad \delta z/\delta y\}$$

Using this set of gradients, a best-fitting surface can be calculated in the form of a mesh grid, which can be directly compared against perceptions for other conditions. Chapters 5, 6, 7 and 8 all employ these methods of capturing slant and tilt settings with a gauge figure task and recreating meshes using Nefs' approach (2008). These are discussed in the context of the rationale within each chapter.

2.5.2.1 Comparing shape and depth of meshes

When converting slant and tilt settings to gradients, it is important to set a maximum angle. In the work contained in this thesis, a maximum angle of 85 degrees was applied. Constraining the angle in this way is important because of the non-linearity of the tangent of the angle of the gradients. As the slant approaches 90 degrees, the tan of the angle approaches infinity, as displayed in Figure 2.7:.

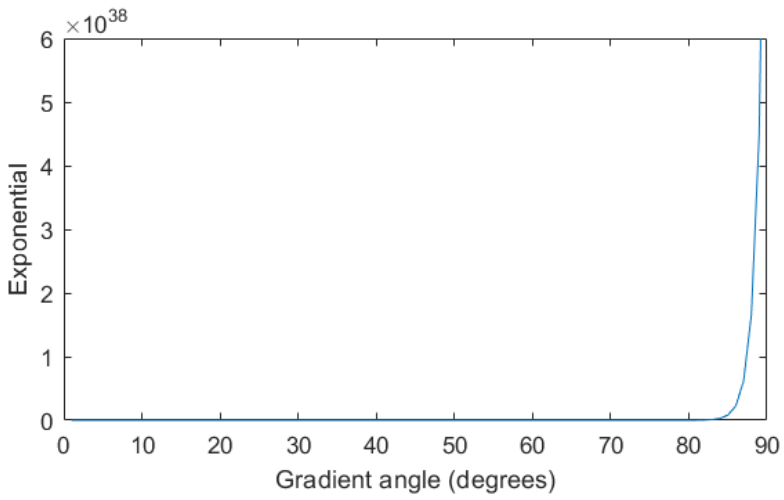


Figure 2.7: Exponential gradients. Graph showing exponential nature of the tan of gradient angles, approaching infinity for angles up to 90 degrees.

Given this relationship, very small changes in the settings for slant and tilt at these angles result in large differences in the fit of the mesh. Therefore, the angle was limited to 85 degrees to reduce this sensitivity.

Nefs (2008) describes how these mesh grid models may be compared through the process of affine transformation. This is a way to quantify the differences in the locations of points in space. The image below illustrates four qualities of affine transformation (Bazargani, Anjos, Lobo, Mollahosseini, & Shahbazkia, 2012).

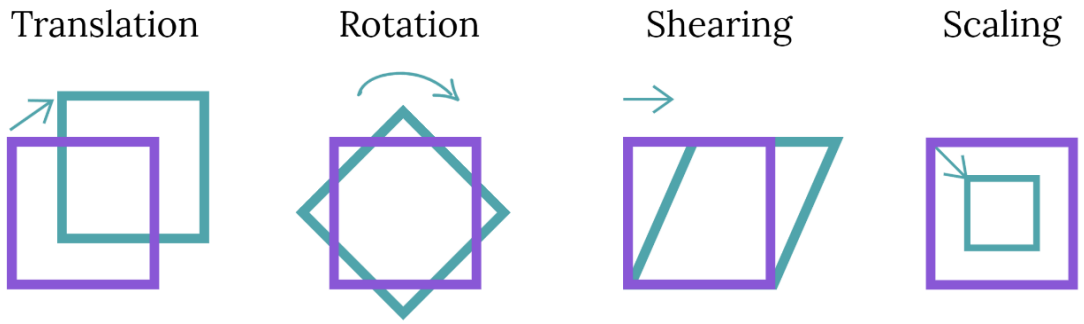


Figure 2.8: Affine transformations. Diagram showing four qualities of affine transformation, used to describe the changes between meshes.

Translation describes a shift in either the x or y coordinates of points, or both, with no scaling, meaning the location of the object changes but the overall size and shape remains the same. Likewise, rotation defines a shift with no scaling, here the object is rotated around its centre of mass so as to provide a different viewpoint. Unlike the previous two examples, shearing does not retain the original shape of the object, instead illustrating here how the points may change such that the edges connecting vertices are no longer at the same angle as before. Shearing is defined by both an orientation, shown as either a positive or negative slope, and a magnitude, from the steepness of the slope, and ranges from a value of 0, showing no shearing, to a maximum value of 1 (Koenderink, van Doorn, Kappers, & Todd, 2001). Finally, scaling shows how the overall shape of the object remains the same, but the size and therefore cartesian coordinate positions change. While these are illustrated here for two-dimensional images, the principle can be applied to three-

dimensional meshes, with the above qualities being combined to describe the transformation in any dimension.

Nefs (2008) uses the horizontal and vertical positions of x and y of the vertices sampled with the gauge figure, as well as the original depth, as a predictor in regression analysis to calculate the degree of affine transformation, using the following formula:

Equation 2.3

$$z' = az + bx + cy + d$$

Using this, the depth of a target model (z') may be predicted using the x and y coordinates, and the depth (z) of the surface in the reference model, where a shows the depth scaling between models, with a constant of d mediating the transformation in depth, and b and c shearing parameters for the x and y axes.

2.6 Analyses

Here, some background on the analyses used in this work is provided, including a discussion of the pros and cons of each method, and what they offer to this body of work. For readability, significant *p* values have been flagged in bold throughout the work, with bold denoting significance for at least a 95% alpha level, and the number of asterisks denoting the alpha level, where one shows significance at a 95% level, two shows 99% and three shows more than 99.9%. Some chapters contain a type of analysis limited to that work, such as Chapter 3 which presents a psi-marginal method for fitting a psychometric function, and these are covered in depth within those chapters.

2.6.1 Statistical power

To begin with, this section briefly discusses statistical power in the work contained in this thesis. Statistical power quantifies the likelihood of significance testing producing a statistically significant difference in results when the null hypothesis is false (Hallahan & Rosenthal, 1996). Power analysis is an important step in designing research to ensure effects of interest are observed (Abraham & Russell, 2008), although it depends on the type of analysis being conducted.

Chapter 3 uses correlation to explore the relationship between certainty of vergence and shape constancy. No significant correlation between the two was found, and using a correlation power calculator (Hulley, Cummings, Browner, Grady, & Newman, 2013) with a two-tailed alpha level of 0.05 for a 95% confidence rate, and a beta of 0.2 which equates to 80% power, the sample size of 35 participants in this experiment should have been able to detect a correlation up to r=.46, were one there to be detected.

Chapters 5, 6, 7 and 8 employ the gauge figure method, and for these types of experiment is has been argued that the number of trials per person is just as important as the number of participants when considering statistical power (Koenderink, van Doorn, Kappers, & Todd, 2001; Baker, et al., 2021). As such, the experiments within these chapters contain large numbers of trials, with the total sum of work representing nearly 85.000 individual gauge figure settings being made. Not only this, but each setting itself produces two numbers for slant and tilt. Koenderink et al (2001) purport that the vast quantity of data gathered using this methodology produce a rich set of data for analysis with sufficient depth.

In fact, relatively small numbers of participants, but large numbers of trials in experiments is the norm for psychophysics, with very few studies conducting power analysis. Much of the work cited in this thesis is based on as few as three participants, sometimes just the researchers themselves. The experiments within this thesis are tested on a relatively large number of participants compared with other traditional work in this field, to balance between participant numbers and data point numbers in consideration statistical power. However, this could potentially be a limitation of this work, and future studies in this field should aim to recruit participant numbers more akin to the standard in other fields of psychology to address this, making use of recent work on the ways to address this (Baker, et al., 2021).

2.6.2 Regression

Regression analysis is used in several chapters to compare the change in depth of reconstructed meshes to look for a linear relationship between variables using the following equation:

Equation 2.4

$$\hat{y} = \beta_0 + \beta_1(x)$$

This explores how much the predicted response ($\hat{y}$) can be predicted by the predictor variable (x), with the average unit of change being the difference between where x = 0 (β_0) and where x = 1 (β_1).

In chapters where meshes have been compared through affine transformation, the x, y and z Cartesian coordinate positions in the reference mesh were used as predictor variables to assess how well these parameters predicted the depth of points in the reconstructed mesh. This provides a measure of the change in

depth between conditions, derived from how well the depth of points in the reference can predict the depth of the reconstructed mesh. It also gives a measure of the change of shape perception between conditions, as the x and y parameters provide shearing information, that is how much the model is stretched in those dimensions compared to the reference. Finally, as the reconstructed meshes in these chapters are normalised to the average distance from observer to object, the intercept term provides a measure of how close or far the object was perceived to be. Therefore, this style of analysis provides a reliable measure of depth relief, and shape and distance perception by which to compare conditions.

Standard linear regression is used in the literature to compare the depth of a reference mesh with a target mesh, by examining how well the reference depth predicts the target depth. Egan and Todd (2015) present a comparison between linear correlations and affine transformations to quantify depth. They presented observers with randomly deformed objects under varying lighting conditions and probed the surface orientation with a gauge figure task. They found that 60% of variance in observers' judgements was accounted for by a linear correlation of relative reported depth and relative depth of the reference, which in their case was the simulated object. When an affine correlation was applied, using the x, y and z Cartesian coordinates, 89% of the variance was accounted for, showing that affine transformations provide a better fit of the change in 3D meshes than linear correlations of depth alone. Indeed, Koenderink, van Doorn, Kappers and Todd (2001) similarly compare linear and affine transformations for their experiment probing pictorial relief with a gauge figure task, and say that by their very design, the values for affine transformations always exceed those for linear regression, by building on the single parameter of depth with shearing and intercept

transformations. Given these benefits, affine transformations have been favoured over linear correlations in this work.

2.6.3 Linear mixed effects models

Many studies in this body of work have a repeated measures design. To account for this, linear mixed effects (LME) models have been utilised to fit the data to account for the fact that residuals between scores for one observer are more similar than the residuals between observers, and should therefore be nested by observer within the model given this heterogeneity amongst observers, which allows their intercepts and partial slopes to vary from the overall average (Morrell, Pearson, & Brant, 1997). This is because individual observers are expected to have a different baseline for depth perceptions and therefore their own intercepts, as well as differences in the scaling of their judgements across the gain conditions. For instance, Nefs (2008) found that scores between conditions for the same observer were more highly correlated than scores between different observers. Given this, individual observer's slopes are expected to vary, and this method can take account of this and model the variance accordingly. The required formula to include the correlated random intercept and slopes (Bates, Mächler, Bolker, & Walker, 2015) is:

Equation 2.5

$$z \sim 1 + x \ + (1 + x \mid y)$$

Where the response variable (z) is predicted using the predictor variable (x) as a fixed effect, with a grouping variable (y) to account for the correlated random intercept and slopes, and a default intercept of 1. Models may contain either just random intercepts ($1 \mid y$) which allows for observers' data to start at different points,

both random slopes and intercepts $(1 + x \mid y)$, or neither, by subtracting 1 to remove the default intercept. These different models are compared using the Akaike Information Criterion (AIC), which evaluates how well the model fits the variance in the data from which it was generated, with a lower AIC indicating a better fit to the data, including negative values. Models throughout this work were compared with their AIC values, and the results are shown in the Appendix for completeness.

While this method is used a lot on this body of work due to its many benefits, there are some instances in which it is not the analysis of choice. For instance, Chapter 5 explores observers' preference in a gauge figure task with varying sizes. This categorical type data is not suited to an LME approach as the categories are not linearly related and therefore the response variable needs to be continuous data, and as such a Chi-square analysis was conducted to compare preference. However, the LME approach does handle categorical predictor variables, and this is used in several chapters.

In summary, a number of advanced technologies and methods have been outlined to address the three main issues highlighted in the introduction. These allow for the creation of complex natural scenes of naturalistic stimuli. The experimental work based on these methods is now presented.

3 Vergence scales binocular depth estimates, but does not account for shape constancy

3.1 Abstract

The same object viewed at different distances will produce a different size of retinal image – the closer the object is, the larger the image on the retina. For a given depth, the binocular disparity also varies with distance. The visual system needs therefore to take account of the distance to an object in order to correctly estimate its 3D shape and size, a process known as constancy. This constancy is not perfect, and systematic errors are made in the estimation of size and depth across varying distances. Typically, objects viewed at far distances are estimated to be smaller and shallower than their physical size and depth. It has been proposed that these errors could be due to uncertainty in the estimation of binocular convergence, the difference in viewing direction between the two eyes when fixating an object. The current work was designed to test this explanation. In a darkened lab, participants were asked to complete two tasks. One task assessed the precision of vergence, the other assessed shape constancy across distance. It was predicted that the extent to which the apparent depth of an object decreased with increasing viewing distance could be predicted from the precision of vergence. When considering individual observers, we predicted a positive correlation, such that a participant with a high degree of precision in their estimate of convergence would make smaller systematic biases in shape constancy. Results showed that general vergence precision was better at far distances. We also found that observers had a failure of shape constancy with distance, but did not find a relationship between this failure of shape

constancy and certainty of vergence. A systematic fixation disparity bias was observed, but this was not of a magnitude high enough to explain the failure of constancy. From these results we conclude that there is no evidence of a relationship between the variation in certainty of vergence and shape constancy and that further work is needed to explain this failure by other means.

3.2 Introduction

Humans, like other animals, use visual information about the world around them in order to safely navigate and interact with it (Warren & Hannon, 1988; Rushton, Harris, Lloyd, & Wann, 1998; Servos, Goodale, & Jakobson, 1992; Watt & Bradshaw, 2000; Bradshaw, et al., 2004; Melmoth & Grant, 2006). For instance, in order to grasp something, such as a piece of food, one must use information about the distance from the self to the object, as well as the physical dimensions of the object itself to execute a successful reaching and grasping movement (Servos, Goodale, & Jakobson, 1992; Watt & Bradshaw, 2000; Bradshaw, et al., 2004; Melmoth & Grant, 2006; Hibbard & Bradshaw, 2003; Keefe, Hibbard, & Watt, 2011).

This chapter focuses on a very simple aspect of the visual judgement of 3D properties of scenes – the judgement of the depth separation between points, and how information about distance from binocular convergence contributes to these judgements. As discussed in the introduction, rather than relying on a single cue to provide depth information, in the Bayesian approaches to depth perception the brain combines the information received from several cues to make judgements about depth.

Studies have shown that combining the information from multiple cues can lead to more reliable estimates than depending on a single source of information (Kunnapas, 1968; Mather & Smith, 2004; Landy, Maloney, Johnston, & Young, 1995). As covered in Chapter 1, a number of cues to depth perception provide information directly from the retina, such as pictorial cues. As humans have two forward-facing eyes, a number of cues combine information from both eyes together, in the form of binocular cues, which will be the focus of this chapter. Given that a human's eyes are separated by around 6.3cm on average for adults, known as the interocular distance, or IOD (Dodgson, 2004), the image presented on each retina is slightly different, creating the cue of binocular disparity. These differences in the relative position of projections of the same points between the two retinal images provide information about distance and depth (Cutting & Vishton, 1995). The geometry behind this will be explored in more detail later in the chapter.

3.2.1 Calculating the vergence angle

In addition to information provided by the retinas, the brain also relies on extraretinal information, such as the cue of convergence (Cutting & Vishton, 1995). The rotation of the two eyes when viewing an object creates a vergence signal in the brain through the state of the extraocular muscles, which provides the central nervous system with an estimate of vergence (Tresilian, Mon-Williams, & Kelly, 1999). Convergence is measured as the angle between the optical axes of the two eyes when fixating an object. This vergence angle is largest for near distances, and reduces with distance. When viewing an object in the far distance, the eyes are effectively parallel with a convergence angle of 0 degrees.

Cutting and Vishton (1995), on discussing the measurement and assumptions of binocular disparity, highlight that the visual system requires at least a roughly accurate assumption about IOD, as well as the current state of vergence angle when viewing an object, as this information is used for the scaling of disparity by identifying the location of points which correspond with the sensory input. This is because disparity scales with viewing distance, so the brain needs a way to interpret this signal (Mon-Williams, Tresilian, & Roberts, 2000). The basic geometry of binocular distance and depth cues is outlined next.

The underlying principles of these cues are based in the geometry created by the triangulation of the two eyes and the viewed object. Linton (2022) describes early 1600s theories on triangulation of stereo depth perception by Kepler and Descartes. This principle is shown in Figure 3.1 below, where a fixation distance, *D*, may be calculated given the known interocular distance, I, and the rotation of the eyes, or vergence angle, *a*:

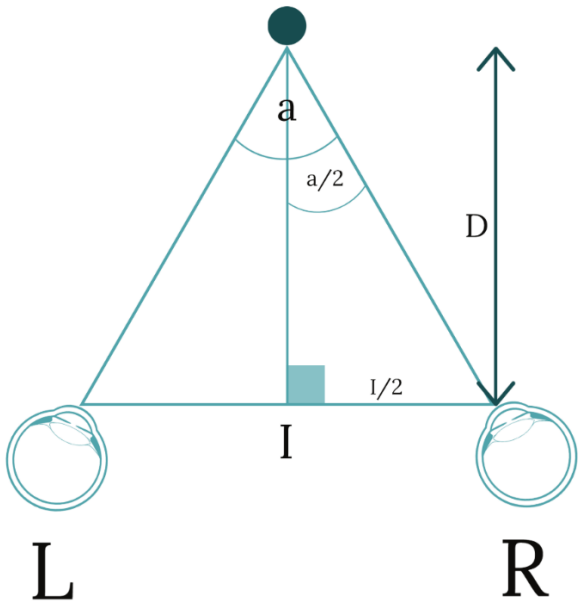


Figure 3.1: Triangulation. Diagram showing triangulation model of stereo depth perception, and how this geometry may be used to calculate distance (D), for a known interocular distance (I) and a given vergence angle (a).

As with many other depth cues, the vergence angle is measured in degrees of arc, with each degree being made up of 60 minutes of arc, and each minute of arc being made up of 60 seconds (Purves & Lotto, 2011). The vergence signal provides the brain with important information, as the brain can interpret the angle at which the eyes converge as an indicator of distance (Mon-Williams, Tresilian, & Roberts, 2000), see Figure 3.1. Indeed, one study has shown that external measurements of participants' vergence eye movements can in principle be utilised as a way of estimating distances, with an accuracy over short distances of over 90% (Inoue, Bounyong, Kato, & Ozawa, 2013). This is likely a similar mechanism employed by the brain to use the information it receives from the vergence angle in order to make an estimate about distance.

As can be seen in Figure 3.1, the triangle created between the viewed object and the two eyes can be halved, to create two right-angled triangles. From this area, denoted with a right angle, basic trigonometry principles can be applied. The formula for calculating the tangent of an angle, for instance half of the vergence angle (*a*/2), would be as follows in Equation 3.1:

Equation 3.1

$$\tan\left({}^{a}/_{2}\right) = \frac{{}^{I}/_{2}}{D}$$

This formula can then be rearranged to use this information to estimate distance (D) as shown in Equation 3.2:

Equation 3.2

$$D = \frac{{}^{I}/_{2}}{\tan\left({}^{a}/_{2}\right)}$$

For instance, using the above equation, for an observer with an interocular distance (I) of 6.3cm, with a vergence signal reporting a vergence angle (a) of 3.6°, the estimated distance (D) between themselves and their fixation point should be reported as 100cm. Likewise, should the same observer view another object at a different distance with a vergence angle of 9°, they would be expected to report a distance estimate of 40cm.

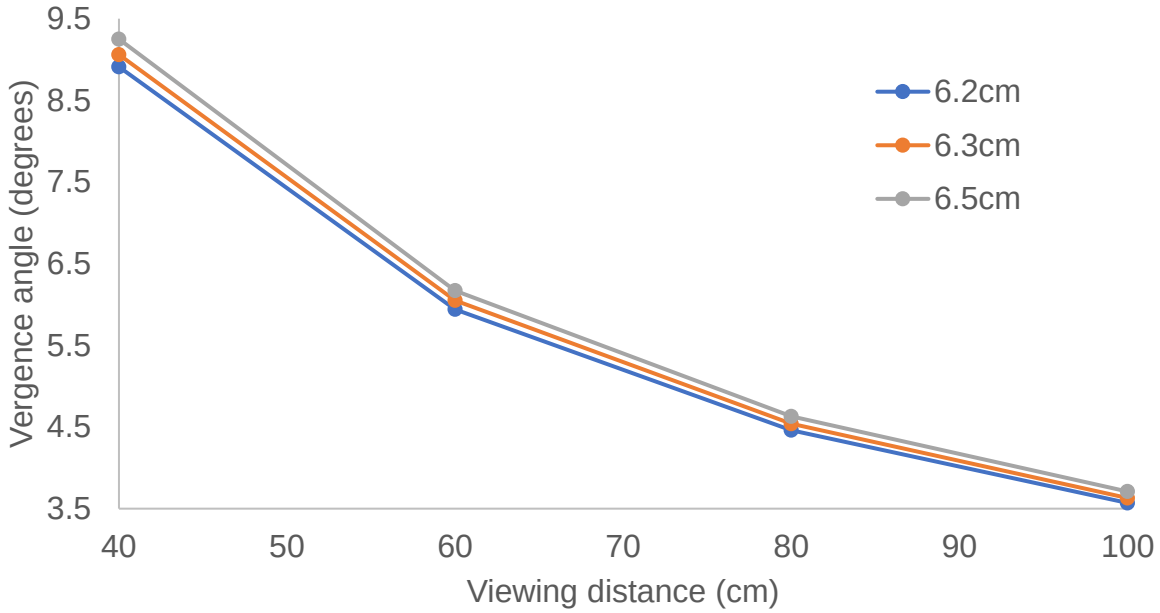


Figure 3.2: Predicted vergence angles. Plot of predicted vergence angle per viewing distance for average female (6.2cm), average adult (6.3cm) and average male (6.5cm) IODs.

Figure 3.2 plots expected vergence angles for differing viewing distances, using the average female IOD of 6.2cm, the average adult IOD of 6.3cm and the average for males of 6.5cm (Dodgson, 2004). Extending this, as can be seen in Figure 3.3, as distance (D) increases, vergence angle (a) decreases, to the point where the eyes are effectively parallel. If we assume a just noticeable change in vergence (v) of 10 arc min (Cutting & Vishton, 1995) and an interocular distance (I)

of 6.4cm, the distance (D) beyond which vergence is no longer useful can be calculated as such:

Equation 3.3

$$D = \frac{I}{tan\left(\frac{v}{60}\right)}$$

This shows that the maximum distance over which vergence is beneficial is 22m, which is considerably further than the distance of 6m that is often assumed (Gregory, 1973).

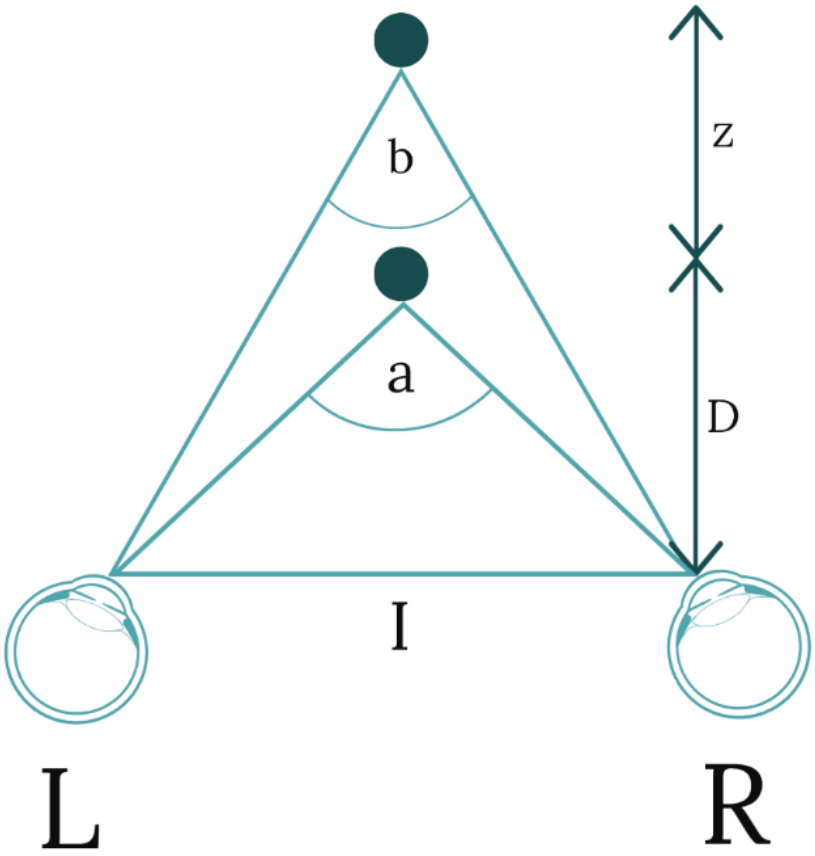


Figure 3.3: Difference between vergence angles. Diagram showing how the left (L) and right (R) eyes converge on objects. For a given interocular distance (I), a wider convergence angle (a) would be observed when fixating the closer object at distance D, than when fixating the further object at distance D+z, which creates a smaller vergence angle (b).

3.2.2 Vergence, disparity and depth

The geometry of vergence can also be used to calculate binocular disparity, and how this can be used to estimate depth. Figure 3.3 shows two points at two distances, D and $D + z$. Where $H = I/2$ is half the interocular distance, the corresponding convergence angles for the two points a and b are:

Equation 3.4

$$\tan\left({}^{a}/_{2}\right) = \frac{H}{D}$$

Equation 3.5

$$\tan\left({}^{b}/_{2}\right) = \frac{H}{D + z}$$

Binocular disparity can be derived from the formulae for the difference between these two angles:

Equation 3.6

$$\tan(a - b) = \frac{\tan(a) - \tan(b)}{\left(1 + \tan(a)\tan(b)\right)}$$

Therefore, binocular disparity can be derived from Equation 3.4 and Equation 3.5 as such:

Equation 3.7

$$\tan\left({}^{a}/_{2} - {}^{b}/_{2}\right) = \frac{{}^{H}/_{D} - {}^{H}/_{(D + z)}}{1 + ({}^{H}/_{D})({}^{H}/_{(D + z)})}$$

Equation 3.8

$$\tan\left({}^{a}/_{2} - {}^{b}/_{2}\right) = \frac{H(D + z) - HD}{D(D + z) + H^{2}}$$

Equation 3.9

$$\tan\left({}^{a}/_{2} - {}^{b}/_{2}\right) = \frac{Hz}{D^2 + Dz + H^2}$$

Assuming that both H and z are small relative to the distance *D*, then the disparity δ is given by:

Equation 3.10

$$\delta = 2\tan^{-1}\left(\frac{Hz}{D^2}\right)$$

This shows that the size of the disparity increases with the interocular distance, but decreases approximately with the square of distance. For an individual observer the interocular distance is fixed. This means that, for objects at the same distance, binocular disparity increases linearly with depth. However, the disparity of a fixed depth interval between two points on an object will decrease with the square of distance as the distance between the observer and object increases.

In summary, the difference between the two retinal images creates relative depth information in the form of disparity, but the brain needs a measure of absolute depth by which to scale this information. Here, the vergence signal can be used as a way to scale this information for depth estimates. Having outlined the underlying geometry, the notion of shape constancy will now be explored.

3.2.3 Shape constancy

Constancy describes the degree to which objects moving away from the observer appear to stay a constant size and shape, despite the image on the retina and the disparity reducing in size as the object gets further away (Wallach &

Zuckerman, 1963; Foley, 1980; Johnston, 1991). The change in the size of the image on the retina occurs because the angle in the visual field taken up by the object is greater for closer objects than for those further away from the observer. As outlined above, the binocular disparity of points on this image will also reduce with distance, but in this case with the square of distance. Shape constancy would require that, as an object moves further away from the observer, it does not appear to shrink and flatten, but maintains its constant size and shape, despite the changes in retinal size and binocular disparity.

This difference in dimension in the retinal image is called the visual angle. For example, the width of the thumb is said to extend a visual angle of approximately 2 degrees when held at arm's length (O'Shea, 1991). Figure 3.4 shows how an object with a height of H, at a distance of D, extends a visual angle (a), and that an object of twice the height ($2H$) can extend the same visual angle at twice the distance ($2D$).

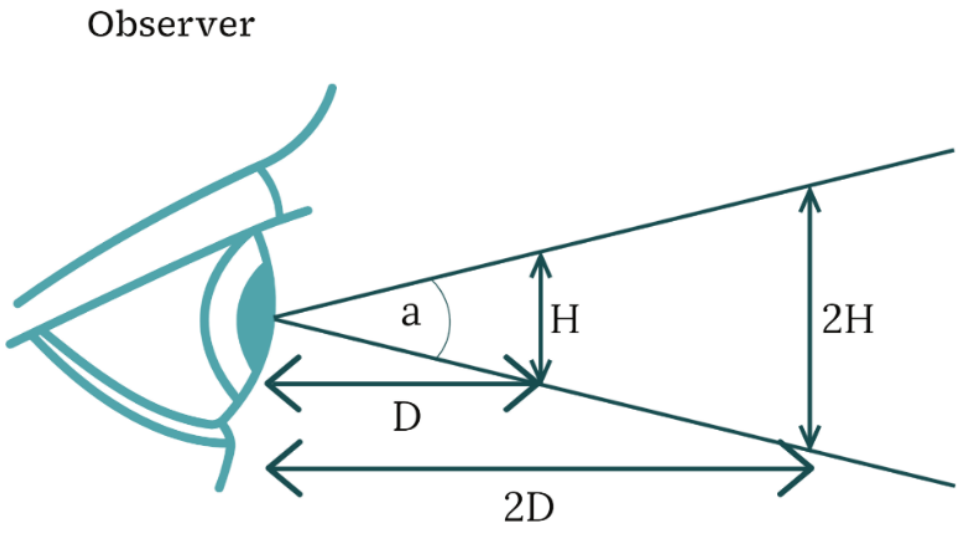


Figure 3.4: Visual angle. Diagram showing how an object with a height (H) at a distance (D) creates a visual angle (a), and that an object with twice the height (2H) may subtend the same visual angle (a) at twice the distance (2D).

Additionally, the same object with height H but at twice the distance ($2D$) would extend a smaller visual angle than at the closer distance D:

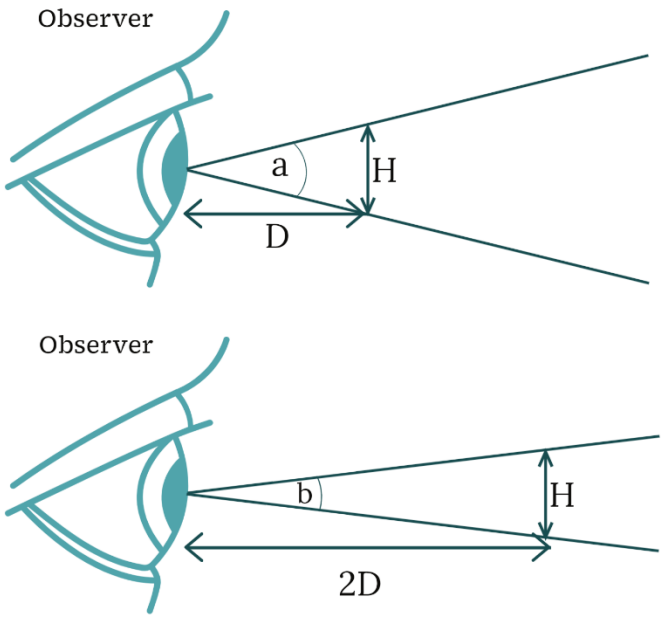


Figure 3.5: Change in visual angle. Diagram showing the change in visual angle with viewing distance increasing from D to 2D.

These diagrams provide an explanation as to what would happen if the observer were to over- or underestimate distance. Were the observer to overestimate distance and perceive that an object is further away than it truly is, say at distance $2D$ instead of D as per Figure 3.4, the object would be perceived to be twice its true height at $2H$ instead of H. The equation (Kaiser, 2017) to measure the visual angle (V) using height (H) and distance (D) for objects level with the line of sight is:

Equation 3.11

$$V = 2arctan\left(\frac{H/2}{D}\right)$$

Say an object of height H of 5cm is viewed at a distance D of 40cm. This would subtend a visual angle of 7.15 degrees. The equation can be rearranged to calculate the height that the object would then appear to the observer:

Equation 3.12

$$H = 2\left(tan\left(\frac{V}{2}\right) \times D\right)$$

Using this, an object of 5cm viewed at 40cm, erroneously believed to be at 60cm would be perceived as 7.5cm in height to extend the same visual angle, and therefore would be perceived as bigger than it truly is, creating a scaling distance error. Equally, if the distance is underestimated, the object would be perceived as closer and therefore smaller, in order to be congruent given the visual angle it subtends. This can be used to measure an effective scaling distance denoting the effective distance perceived by observers given their reported judgements (Bradshaw, Parton, & Glennerster, 2000).

Much work in this area has shown that the human depth perceptual system, like many other systems, is subject to noise. Additionally, it has been shown that judgements made at far distances are often underestimated (Baird, 1970; Brenner & van Damme, 1998; Johnston, 1991; Scarfe & Hibbard, 2006; Viguier, Clément, & Trotter, 2001) as space within the visual field tends to be squashed at further distances, so that objects are seen and reported as being closer than they really are, and do not scale correctly. Objects seen at a closer distance appear stretched, and only things seen at an intermediate distance appear veridical (Johnston, 1991). As discussed above, triangulation assumes the ability for the brain to accurately estimate vergence to make estimates of distance, but as highlighted previously,

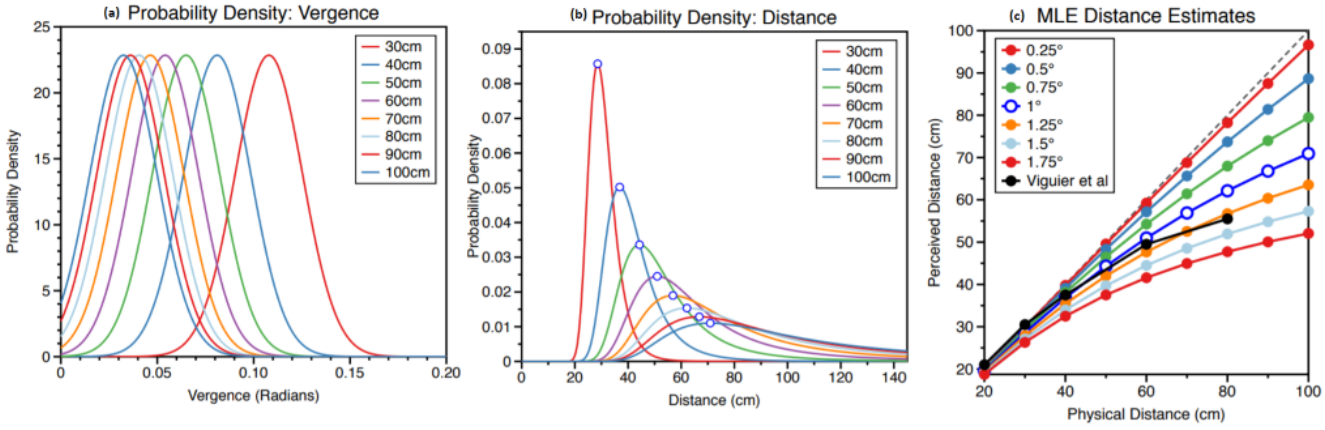
depth cues are subject to noise and uncertainty. It will now be discussed how this imprecision may affect the depth estimate.

3.2.4 Vergence uncertainty

Included in this array of work are several studies that have shown that an underestimation of perceived distance also occurs when specifically relying on binocular convergence (Mon-Williams, Tresilian, & Roberts, 2000; Viguier, Clément, & Trotter, 2001). Viguier and colleagues (2001) investigated participants' ability to perceive and estimate distances using either retinal disparity cues or only extraretinal information such as vergence. They presented participants with light-emitting diodes at distances ranging from 20cm to 120cm. Participants were tasked with reproducing this seen reference at either the same, double or half the original distance. Results showed that when retinal disparity cues were available to participants, they were able to successfully reproduce the seen reference with high accuracy across the range of distances. However, when only the extraretinal cue of vergence was available to provide information, participants were only able to reliably reproduce the reference distances when within arm's reach, but underestimated distances beyond 60cm. This shows that biases occur, and therefore errors are made, with increasing distance when relying solely on extraretinal cues compared to when retinal information is available.

One idea put forward to explain these systematic errors is that when relying on the convergence angle alone, biases in depth estimates stem from uncertainty of vergence, such that a person who is more uncertain of vergence will exhibit higher levels of bias, and therefore misestimation of depth, than a person who is relying

more on the measurement of vergence to aid in depth perception (Scarfe & Hibbard, 2017). As such, if the vergence signal contains a lot of noise, the prediction would be that the scaling of disparity and therefore shape constancy will be poor at estimating the true depth, as would be evidenced by a negative correlation between vergence noise and shape constancy.



Graphs reproduced with permission (Scarfe & Hibbard, 2017) from poster presented on a Bayesian model of distance perception from ocular convergence. Graph (a) shows probability densities of vergence angle for distances between 30 and 100cm, making the assumption that variance is constant across vergence angles. Graph (b) shows information from graph (a) as a function of distance, showing the probability of a distance being selected from the observed vergence angle as the most likely estimate at a given distance. Graph (c) shows perceived distance against physical distance for differing levels of noise (bias) in vergence angle, here shown by standard deviations, between 0.25° and 1.75°, as well as results from a previous study that is concordant with their work (Viguier, Clément, & Trotter, 2001).

Figure 3.6 shows a theoretical interpretation of distance underestimation proposed by Scarfe and Hibbard (2017), which assumes that the uncertainty of vergence is constant across distance. The theory they have put forward suggests that errors in the estimation of physical distances may stem from increased noise in distance estimates derived from the vergence signal, which leads to maximum likelihood errors. Relating this back to the trigonometry presented earlier, for an observer with an IOD of 6.3cm, with a vergence signal reporting a vergence angle of

3.6°, an observer should give an answer of distance estimated at 100cm. However, as the graphs show, there is noise in several stages of this process. When estimating the vergence angle, multiple neurons in the primary visual cortex are excited by any given vergence signal, including the correct responses, as well as many around it (Kaufman & Alm, 2003). As such, the brain selects the most likely vergence angle to have excited the neurons in such a pattern. The first graph shows the probability of vergence for the given distances between 30 and 100cm (Figure 3.6a). These are assumed to be normally distributed, with the mean indicating the most likely vergence angle for a given distance. As such, this represents an opportunity for errors to be made, due to a noisy signal.

These probability densities can be replotted as a function of distance, such that the next graph shows the probability density that a given distance will be estimated from the measured vergence angle (Figure 3.6b). The peaks of the curves represent the most likely distance to be estimated by an observer. This creates a potential for bias. At 30cm, the peak estimate is accurate, as this shows an observer would be most likely to estimate the distance at 30cm to be 30cm. However, the predicted bias (underestimation of distance) increases with increasing distance, so that a noisy vergence signal for an observer viewing an object at 100cm is predicted to produce a distance estimate of around 70cm. This can be seen in the third graph, which displays how maximum likelihood estimates of distances progressively cause observers to underestimate the physical distance of a fixation point (Figure 3.6c). Viguier et al. (2001) reported data that fit with this theory that increased noise in the vergence signal, indicating increased biases, leads to progressively less accurate estimates of physical distances, in that these are increasingly underestimated at further distances.

The purpose of the current study was to measure the precision of vergence, as well as the accuracy of shape constancy, and see if the first predicted the second. This was achieved through two tasks, one which tested participants' certainty and bias of vergence, and one which measured their shape constancy across distance.

3.2.5 Measuring certainty of vergence and shape constancy

The task designed to assess a participant's certainty of vergence was a nonius lines task which involved presenting participants with a pair of nonius, or vernier, lines which were presented dichoptically - one to each eye (Jaschinski, Bröde, & Griefahn, 1999). Presenting one image to each eye in this way removed binocular disparity information, allowing for measurement of vergence noise for the perceived fixation point, since uncertainty or bias in vergence leads to uncertainty or bias in the nonius line alignment judgement. The nonius line task is typically used to measure visual acuity, but instead here the task was used to assess certainty of vergence by measuring the bias and noise of the vergence estimates. Observers were presented with the pair of nonius lines and asked which line appeared rightmost. The lines were generated with differing horizontal offsets to explore certainty through the smallest just noticeable difference (JND) in the stimuli intensity, and the bias through the point at which the lines appeared lined up to observers. An observer with more certainty of vergence would therefore show better precision in this task than one who is less certain, who would present with a smaller just noticeable difference.

Chopin, Levi, Knill and Bavelier (2016) used a nonius lines task to measure noise of the vergence signal to explore the suggestion that vergence noise accounts

for the observed difference between accuracy of absolute and relative depth estimates. They estimated a value for vergence noise of 225 arc seconds. They also found a difference between absolute and relative disparity acuities but concluded that this could not be explained by vergence noise alone.

Similar methods were used by Chopin, Levi and Bavelier (2017) to study individual differences in binocular visual acuity by comparing dressmakers to a control group. They were interested to see if dressmakers, who need to be able to converge their eyes accurately for their work, and use disparity to perceive depth differences, would show less vergence noise than people of other professions. They presented dressmakers with pairs of nonius lines by flashing them on screen and asked them to report whether the line above fixation was to the left or the right of the line below. From this they measured how accurate observers were with vergence as a measure of bias, and looked at the spread of variability of their staircase procedure as a measure of noise of the vergence signal. They found that dressmakers had better stereoacuity than non-dressmakers, although vergence bias and noise were not found to differ significantly between the groups. We adopted the methods outlined in these two studies by Chopin and colleagues (2016; 2017) to calculate an estimate of vergence noise in order to find the JND as a measure of observer's certainty of vergence.

The second task was designed to measure accuracy of depth perception across distance, in order to assess the observer's shape constancy. Participants were presented with stimuli at four distances ranging from 40 to 100cm, and asked to make a judgement about the depth of the stimulus presented (Bradshaw, Parton, & Glennerster, 2000). In line with previous work, errors made here were expected to

increase with increasing physical distance, showing a lack of shape constancy. Specifically, it was predicted in the present work that the JND would increase with distance and also that perceived depth would tend to decrease with increasing distance (Johnston, 1991; Bradshaw, Parton, & Glennerster, 2000).

The stimulus for the depth task in the present work was a triangle in 3D space, defined by three dots in a vertical line. The top and bottom dot were presented at the same distance, and the middle dot at a closer distance to create a triangle in depth. Participants judged whether the depth of this triangle was larger or smaller than half its height as the standard stimulus. This allowed us to calculate a Point of Subjective Equality (PSE) indicating the point at which the presented stimulus appears to be exactly equal to the standard stimulus (Rajamanickam, 2002). A JND was calculated to quantify the precision of these judgements. A regression slope was then calculated from the PSE scores from the four distances, which provided a measure of the change in bias with increasing distance, showing shape constancy, or a lack thereof. With perfect constancy, the slope of this line would be zero. If participants underestimated depth at far distances relative to close distances, the slope of the line would be positive, since this would show that observers required an increasing amount of depth in the stimulus to maintain a constant perceived depth. Therefore, the bigger the slope value, the worse the shape constancy.

This approach is similar to the work of Bradshaw, Parton and Glennerster (2000), who probed the relationship between relative and absolute depth recovery in a shape constancy task. Based on Johnston's (1991) 'apparently circular cylinder' task, they presented participants in their shape task with a set of Light Emitting Diodes (LEDs) to create a triangle, and asked them to adjust the location of the

LEDs to match the distance between the base LEDs, so to set the height of the triangle equal to the width. The LEDs were placed along the horizontal line of sight of the observers to remove the potential of vertical disparity information. The separation of the two LEDs that formed the base of the adjustable triangle was fixed at 20cm for the closer condition viewed at 150cm, and 40cm for the further condition viewed at 300cm. This experiment was conducted both monocularly and binocularly, in a darkened room with no additional light sources, and observers used a headrest, that either kept their position fixed in the static conditions or allowed free movement side to side of 6.5cm in the motion conditions.

In order to complete this task, the viewing distance needed to be recovered from the scene by observers, in order to scale the depth information received between the relative disparities of retinal images from the two eyes such as for static binocular viewing. As the base height of the triangle scaled with viewing distance, the observers were expected to scale their responses accordingly between the further and closer viewing conditions if singularly taking into account viewing distance. However, they found systematic biases for shape constancy that had not been observed in their other experiments using the same viewing distances. An effective scaling distance was calculated as the distance at which the disparity of the stimuli set by the observer matches the expected correct response of either 20cm or 40cm for the 150cm and 300cm viewing conditions respectively. The difference between this scaling distance and veridical performance was found to increase with viewing distance, in that further distances produced a higher rate of underestimations of depth than at a closer distance, showing reduced shape constancy across viewing distances.

Given the expected differences in bias between observers, the current work was interested in individual differences of certainty of vergence and associated shape constancy, and correlated scores to look for a relationship between the two. Nefs, O'Hare and Harris (2010) used an individual differences approach to explore motion in depth perception. Their work presented participants with two random dot stereogram (RDS) frames, one above the other, that simulated motion in depth over time in the form of dots moving towards or away from the observer. They present two theories as to how depth from motion is deduced, and as such they presented stimuli containing information for either Changes in Binocular Disparity over Time (CDOT), Interocular Velocity Differences (IOVD), or information for both mechanisms together, and asked participants to report if they perceived the motion of the upper panel to be moving towards or away from themselves. They correlated scores between the different mechanism conditions to look for a relationship between stimulus type and one or more mechanisms underlying performance, and found evidence of two opposite patterns of sensitivity between participants. From this, they highlight the importance of consideration of individual differences for mechanisms underlying visual perception. Likewise, Hibbard, Bradshaw, Langley and Rogers (2002), looked at individual differences and the mechanisms that underlie these for the perception of stereoscopic surface slant. They correlated results of thresholds for orientation and spatial frequency discrimination with 3D slant perception, and found a positive correlation between the two, from which they argued that perception of surface slant is limited by sensitivity to the underlying orientation and spatial frequency differences.

Firstly, as the main focus of this work, we predicted that there would be a positive relationship between the slope of the PSE in the depth task, and the JND in

the vergence task, such that as the slope of the PSE in the depth task increased, the JND in the vergence task would also increase. This would show that if a participant is less certain of vergence, as shown by a larger JND in this task, they would also show less shape constancy through a steeper slope of PSE scores in the depth task. Equally, those with lower JND scores in the vergence task, showing greater certainty of vergence, were expected to show more shape constancy across the four distances, as indicated by a shallower PSE slope.

From the information outlined above, it was predicted that all participants would exhibit increasing levels of systematic bias for depth estimates with increasing stimulus distance, requiring a deeper triangle, and therefore larger PSE score, for further distances. As the JND scales with the size of the stimuli, we also predicted that the JND would naturally increase as the PSE increased in the depth task.

We were not expecting a systematic bias in the vergence task, as the PSE score should not have differed with increasing stimuli distance. This is because the PSE in the vergence task indicates the point at which participants viewed the lines as lined up. This point was not expected to change with distance. Although some observers may exhibit bias in this task, this bias should remain constant across the four distances. However, the JND was expected to decrease with increasing stimuli distance, as it scaled with the size of the vergence angle, which in turn reduced with distance.

In summary, observers' certainty of vergence was tested with a nonius lines vergence task to establish a JND, their shape constancy was measured using the PSE from the triangles depth task, and the two were correlated to look for a relationship between certainty of vergence and shape constancy. This work

predicted that observers with less certainty of vergence shown in the nonius lines task would exhibit reduced shape constancy and greater variability in the triangles task.

3.3 Methods

3.3.1 Participants

35 participants between the ages of 18 and 28 were recruited. 74% identified as female and 26% identified as male. All were screened prior to the start of the experiment for normal or corrected-to-normal vision, as well as stereoscopic acuity. Participants included one of the researchers, as well as 34 people naïve to the purpose of the experiment.

3.3.1.1 Recruitment

Participants were recruited through the University of Essex's online SONA system, as well as through word of mouth. Some participants who were enrolled as Psychology students received course credit for their participation, while others were compensated financially.

3.3.1.2 Screening

Two vision tests for normal vision and stereo acuity were administered to see if participants qualified for the experiment. The Stereo Optical Butterfly random dot depth test (Stereo Optical, 2020) was administered to screen for sufficient gross stereopsis, with the cut-off point for participation being if participants could view the

entire 3D butterfly, which equated to 700 seconds of arc. This was viewed through polarised glasses at a distance of 41cm (16 inches), as per the instructions. Participants were also screened for normal or corrected-to-normal vision using the Lighthouse Distance Visual Acuity Test. The cut-off point for participation was receiving a Snellen score of 32 or better, as this gave participants a visual acuity score of 90, with the Snellen ratio 20:20 being considered normal vision, which indicates a visual acuity score of 100. Participants who did not meet the screening criteria were thanked for their time and did not participate. Those who did pass the screening then underwent set up tasks. Participants' IOD was measured as the distance between the two eyes using a standard ruler. This was measured three times and average taken to ensure accuracy. The participant's dominant eye was assessed by holding up a pen at arm's length with both eyes open and aligning it with a mark on the far wall, then alternately closing each eye to see which remained aligned (Porac & Coren, 1976).

3.3.2 Apparatus

The stimuli for both tasks were generated and presented using MATLAB with the Psychophysics Toolbox extension (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) and were viewed on a 52 by 29cm VIEWPixx3D monitor with a resolution of 1920 by 1080 pixels. VIEWPixx 3D synchronisation LCD shutter goggles synchronised to the 120Hz refresh rate of the screen, along with a 3DPixx IR emitter, presented a different image to each eye individually, giving a total of 60 frames per second to each eye. Stimuli were presented in red and the crosstalk between eyes was measured to be 0.12% using a Minolta LS-110 photometer.

Participants sat with their head on a chin rest, adjusted so that the middle of the screen was at eye level for each participant, to minimise head movements during trials to eliminate additional depth cue information. Responses were recorded using either the ‘Up’ or ‘Down’ arrow keys on a standard computer keyboard.

3.3.3 Stimuli

The stimuli presented for the vergence task were a pair of red vertical nonius lines, 10mm tall and 4mm wide, set against a black background. These were presented one line to each eye as shown in Figure 3.7 using the VIEWPixx goggles.

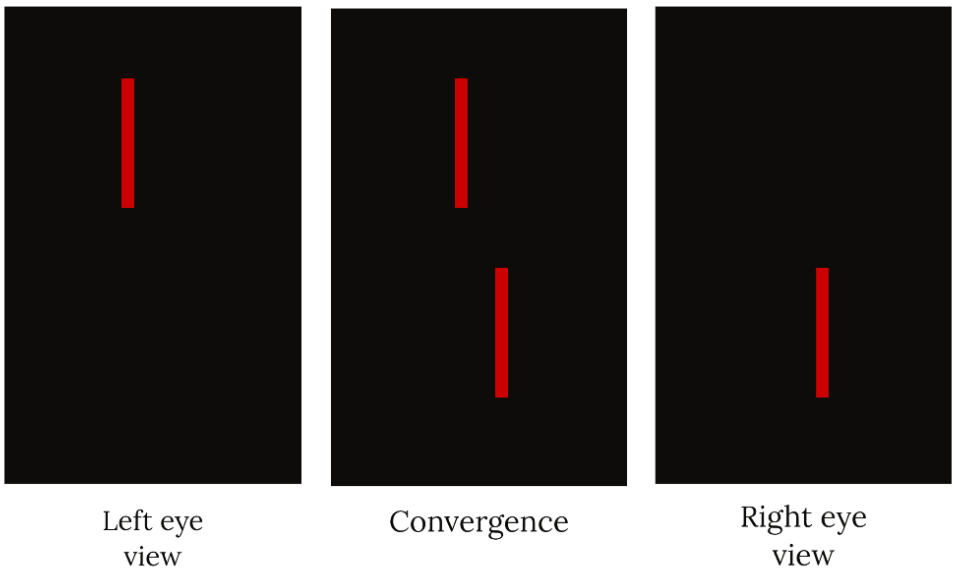


Figure 3.7: Nonius lines stimuli. Image showing left and right eye views of example nonius line stimuli, with the view of convergence.

The stimuli presented to participants in both the vergence and the depth tasks were generated using a ‘psi-marginal’ psychophysical staircase method (Prins, 2013), using the Palamedes Toolbox extension (Prins & Kingdom, 2018) within MATLAB. This calibrates the stimulus level based on participants’ response in the

previous trial to get a good fit for the psychometric function by positioning points along the curve to get a good measure of both the midpoint and the slope, and calculates the main parameters including the standard error for each trial, ensuring a robust result and giving increased confidence in the estimates it provides. A maximum of 20 steps in stimuli intensity were set, calculated using the angle of the stimuli in degrees. Figure 3.8 shows three example presentations of pairs of nonius lines.

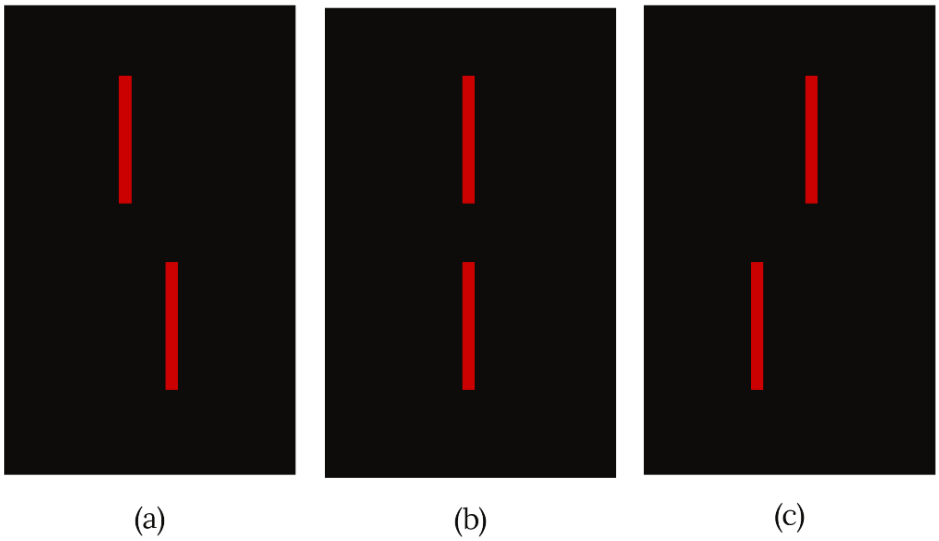


Figure 3.8: Differing nonius orientations. Representation of three different variations of stimuli that could have been presented to participants during the vergence task. Participants were presented with a single pair of lines where the top line was either on the left (a), in line with the bottom line (b) or to the right (c).

The psi-marginal method also allows for inclusion of differing guess and lapse rates, where other methods predetermine these. The guess rate, or gamma, is the starting point of the psychometric function, which in this work is always 0 as the starting point of the proportion of responses. The lapse rate, or lambda, is the point at which the psychometric function lapses, and can vary up to 100%. Ensuring these parameters can vary reduces floor and ceiling effects. This way, participants were presented with stimuli that measured their individual ability level in the tasks,

meaning individual differences were better recorded than if a set range of stimuli had been presented. This method is also beneficial in addressing issue of participants making a mistake. Were an observer to accidentally press up instead of down, this method would adjust the stimuli intensity, but this would then be corrected in subsequent trials. In other adaptive methods, mistakes made by the observer may result in stimuli intensities skewed by this error (Prins, 2013). Additionally, this method presents a number of trials with intensities away from the threshold, ensuring optimal slope estimation (King-Smith & Rose, 1997), a main benefit over traditional staircase methods (Cornsweet, 1962).

The stimuli used in the depth task consisted of three red dots presented against a black background in a vertical line, representing a triangle in 3D space (see Figure 3.9). The dots were 5mm in diameter and the base height of the triangle, as represented by the top and bottom dots, was 4cm. The dots were rendered as spheres within MATLAB so as to scale appropriately with distance, and no light source was added to the scene which ensured they appeared as dots and not spheres.

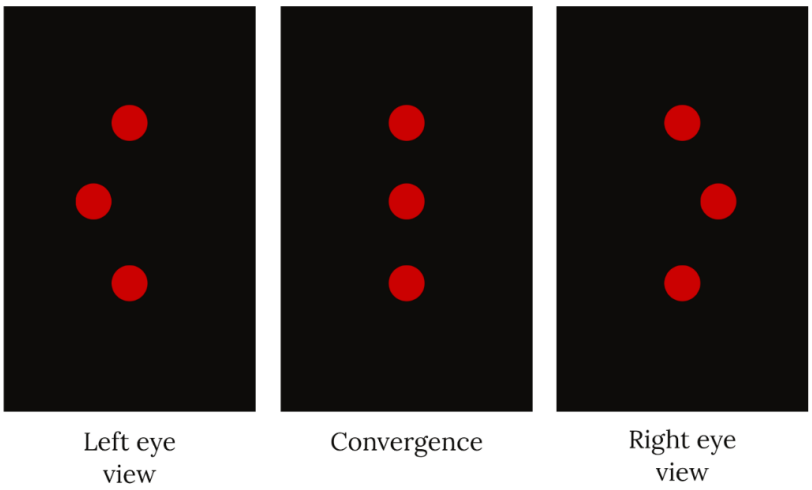


Figure 3.9: Triangles stimuli. Image showing stimuli in the triangles depth task, including the view presented to the left and right eyes, and image seen by participants when converging the scene.

100 trials were displayed per block. Although the range of stimuli seen by each participant was unique to them, there were limitations on the range of stimuli that the program was able to present. For the vergence task, the range for the distance between the nonius lines was between 0 and 2000 seconds of arc. For the depth task, two ranges were used. At the closer distances from the monitor (40 and 60cm), the depth of the triangle presented could be between 0 and 8cm, and for the further distances (80 and 100cm), the range of triangle depths was between 0 and 15cm. these ranges were tested during piloting of the study and found to offer a wide enough range of stimuli to capture sufficient data.

3.3.4 Procedure

Written consent was obtained from participants, and the screening tests were administered, with only those whose performance was better than the set criteria being invited to take part in the study. Both tasks took place in a darkened room, and stimuli were presented at a distance of either 40cm, 60cm, 80cm or 100cm by moving the monitor to these distances from participants' eye level. Each of the two tasks was presented at each of the four distances, giving a total of eight blocks. Blocks of trials were randomised between participants to avoid practice effects. Participants were seated in the darkened room with their chin in the chin rest prior to the start of the experiment. Between each of the eight blocks, the dark room was illuminated to allow setup of the next block, as well as reduce participants' adaptation to the lack of light, which may have provided additional depth cue information.

For the vergence task, a 10mm fixation cross appeared at the centre of the screen for 1 second. This was then automatically replaced by a set of nonius lines,

which were presented for 100ms. These were then replaced by a black screen and participants were given as long as they required to respond. Participants were required to report the location of the line which appeared to them to be on the right. For instance, if participants thought the line on the right was the top of the pair of lines (as per Figure 3.8c) they were instructed to press the 'Up' arrow key on the keyboard, and to press the 'Down' key when the line on the right appeared to be the bottom of the pair of lines (see Figure 3.8a). If participants were unsure of which line was right-most, or if the lines appeared to line up perfectly (as per Figure 3.8b) participants were instructed to guess either 'Up' or 'Down'. Once participants had pressed the key corresponding to their answer, the next trial began automatically, as indicated by the fixation cross.

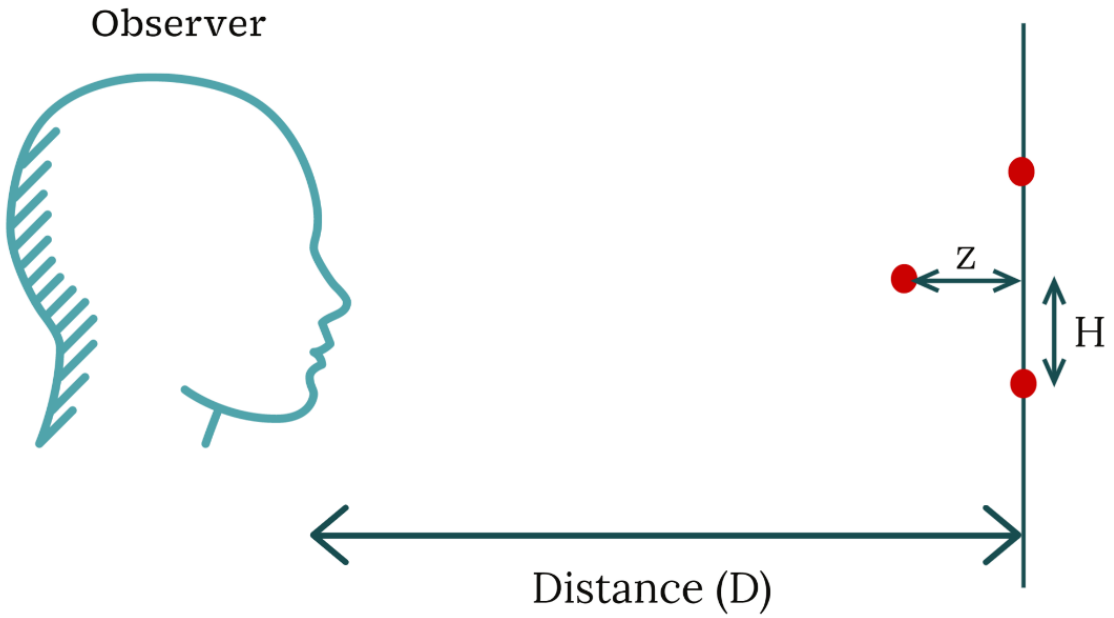


Figure 3.10: View of the experiment. Representation of a side view of the participant in the experiment, where D is the distance of the monitor, either 40cm, 60cm, 80cm or 100cm. Participants were tasked to decide if a presented triangle was too shallow or too deep, with the 'perfect triangle' being one in which the distance between the closest dot and the line created by the other two dots (z) was equal to half of the height of the vertical line (H).

For the depth task, again a 10 mm fixation cross appeared for 1 second at the centre of the screen, aligned to the depth of the base of the triangle. This was then replaced with the three dots. These were presented for 100ms, before being replaced by a black screen, and allowing participants as long as required to input an answer. Again, this was done by pressing either the 'Up' or 'Down' arrow key. Figure 3.10 shows a side view representation of the participant in the experiment. Participants were informed that a 'perfect' triangle in this task would be one where the distance of the closest dot (*z*) was equal to half the height (*H*). If the triangle presented on the screen looked too shallow to be a 'perfect' triangle, participants were instructed to press the 'Up' arrow key, and to press the 'Down' arrow key if the presented triangle looked too deep. Once participants had pressed the key corresponding to their answer, the fixation cross appeared once more to indicate the start of the next trial. Once the block of trials was finished, the room was illuminated and the screen moved to the appropriate distance for the next block of trials, as per the randomised order for each participant. At the end of the eighth block, participants were debriefed on the purpose of the experiment.

3.4 Results

3.4.1 Data treatment and psychometric functions

Raw scores for each participant were in the form of responses to the task stimuli. For the triangles depth task, this was a response of 0 from pressing the 'Down' arrow key if the depth was deemed deeper than the reference stimuli, and 1 from pressing 'Up' if it was deemed shallower. For the nonius lines vergence task,

this was a response of 0 from pressing the ‘Down’ arrow key to indicate the line on the right was at the bottom, and a 1 to indicate it was at the top of the pair.

The psi-marginal method used in this work fit a new psychometric function to the data and calculated four key parameters, as well as the standard error, for each of the 100 trials: alpha, beta, gamma and lambda, and used these to identify each participant’s individual level of performance. This process can be seen in Figure 3.11 where the slope estimates and standard errors start off high in the first few trials and quickly reduce with changing stimuli intensity as the psychometric function is refit each trial.

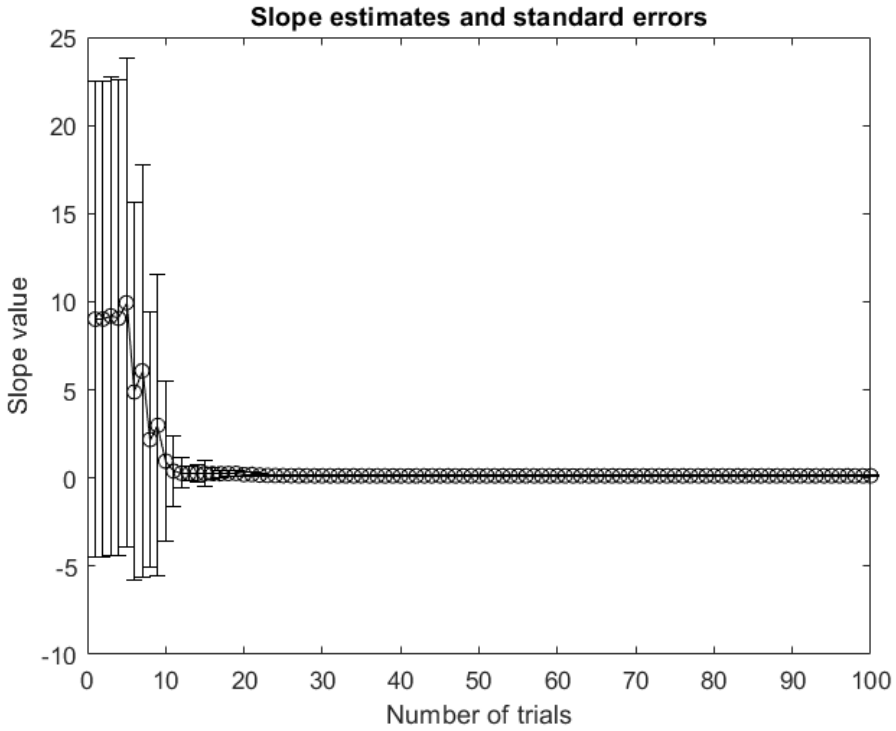


Figure 3.11: Psychometric fit. Graph showing example of change in slope estimate and standard error from refit of psychometric function across trials with the psi-marginal method for one participant in the vergence task at a distance of 100cm.

From these, cumulative Gaussian psychometric functions can be calculated for each participant at each distance. Here, two parameters are most important, the

threshold and the slope of the psychometric function (Prins, 2013). The threshold, or alpha parameter, indicates the location of the psychometric function and denotes the PSE stimulus intensity that specifies a desired level of performance has been achieved, which for this task represents the point at which observers were equally likely to press up or down. Values were expected to match the veridical 50% point, which is 2cm for the depth task, and 0 minutes of arc for the vergence task, if no bias was present. The slope value, or beta, denotes the function's rate of change, with a smaller result indicating a shallower slope and therefore less certainty of response.

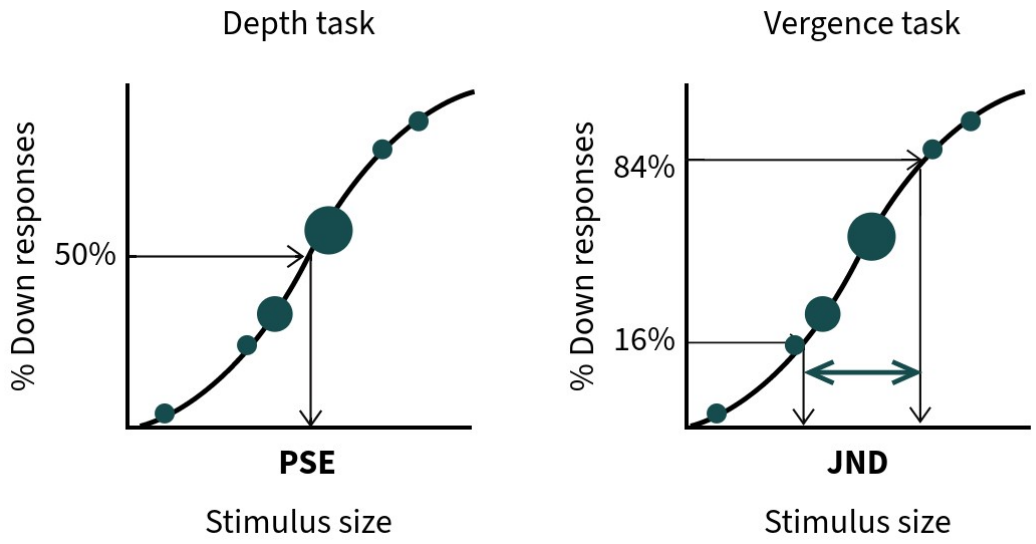


Figure 3.12: PSE and JND. Graphs to show how the PSE and JND were calculated for this experiment from the psychometric function. PSE in the depth task was calculated as the point at which 50% of responses were 'Down' and the JND in the vergence task was taken as one standard deviation which here gave a criterion JND of between 16 and 84%.

Figure 3.12 above illustrates how the PSE and JND were calculated from the cumulative Gaussian psychometric function curve, and the values used for the cut off. In this work, the graphs plot the percent of Down arrow key responses made by participants when viewing a stimulus of that size or intensity against the size of the

stimulus presented. The shape of the curve shows how participants change from confidently always indicating that the stimulus shown was shallower than the standard stimulus indicated by 0% of Down responses, to always pressing that the stimulus shown was deeper than the standard stimulus, indicated by 100% Down responses, when presented with a stimulus of a larger magnitude. This would represent, for instance, a triangle that is obviously deeper always causing the participant to press the Down arrow key. The middle of the curve represents a participant's uncertainty when presented with a stimulus closer to the veridical magnitude of the standard stimulus.

From the psychometric functions, two parameters were used for both the depth and the vergence tasks. The Point of Subject Equality (PSE) was calculated as the point at which participants gave each answer 50% of the time, indicating the size of the stimulus that participants believed to be the same size as the standard stimulus, which comes from the alpha parameter. For the depth task, this was the point that participants would have said that the presented triangle was the same depth as the standard stimulus, and therefore were equally likely to say deeper or shallower 50% of the time. For the vergence task, this is the point at which participants would have said that the lines were exactly lined up. Specifically of interest in this work is the PSEs in the depth task, which show how the bias changes with increasing stimuli distance, showing shape constancy.

The standard deviation of the underlying Gaussian was also determined as the inverse of the estimated slope. This value determines the Just Noticeable Difference (JND), the difference in stimuli magnitude required for participants to notice a

difference at a criterion level of reliability. The measured value of one standard deviation corresponds to a criterion JND of between 16 and 84%.

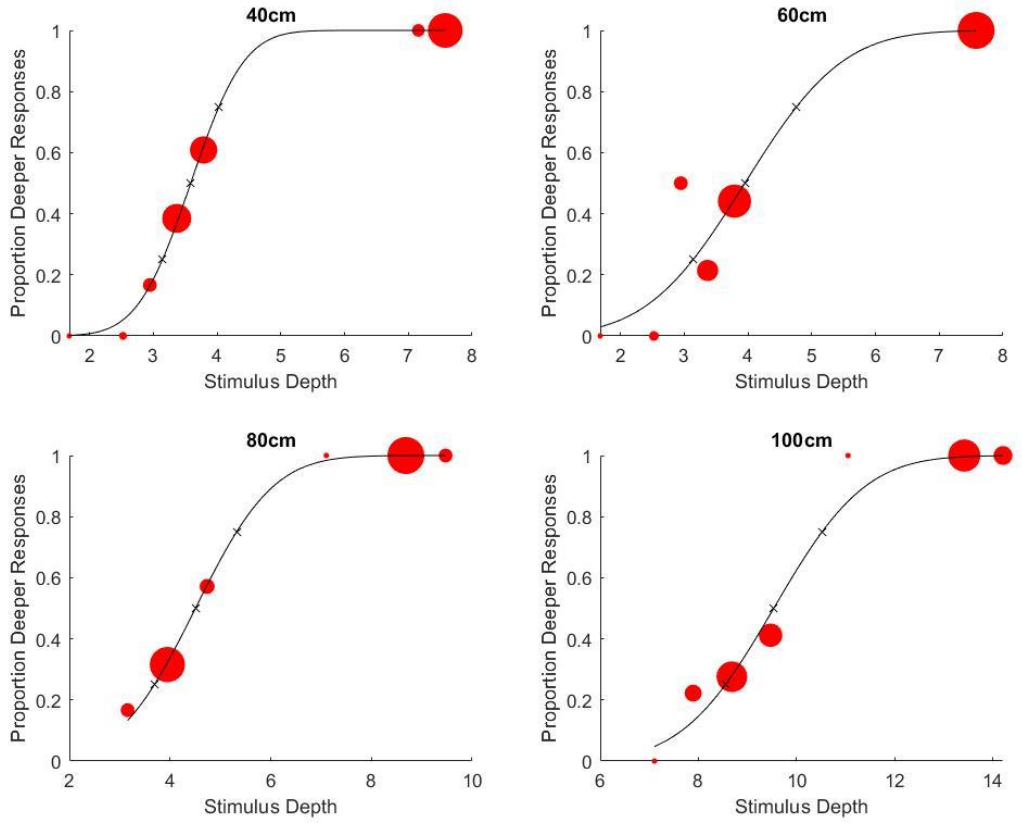


Figure 3.13: Psychometric curve. Example psychometric function slopes for one participant at all four presentation distances. Bigger circles indicate more trials of this stimuli intensity presented to participants.

This graph shows the percent of responses where the participant pressed the down arrow key to indicate that the line on the right was at the bottom of the pair of nonius lines. Here, the more trials presented to a participant by the psi-marginal method, the bigger the circle. The curve at 40cm is fairly steep, as seen from the sudden shift to 100% of responses resulting in the Down arrow being pressed at around 0.15 degrees of arc. This shows the observer confidently reported the rightmost line with a small difference in presented stimuli, showing less noise in the

vergence signal and therefore higher certainty of vergence. The PSE and JND were extracted directly from the psi-marginal output.

3.4.2 Effect of distance of PSE and JND

Means and standard errors (SE) for the PSE and JND were calculated for both tasks. These are plotted in Figure 3.14. The dashed line indicates what perfect performance would look like, without the expected systematic bias, here, 2cm for the triangles depth task, and a PSE of 0 seconds of arc for the vergence task.

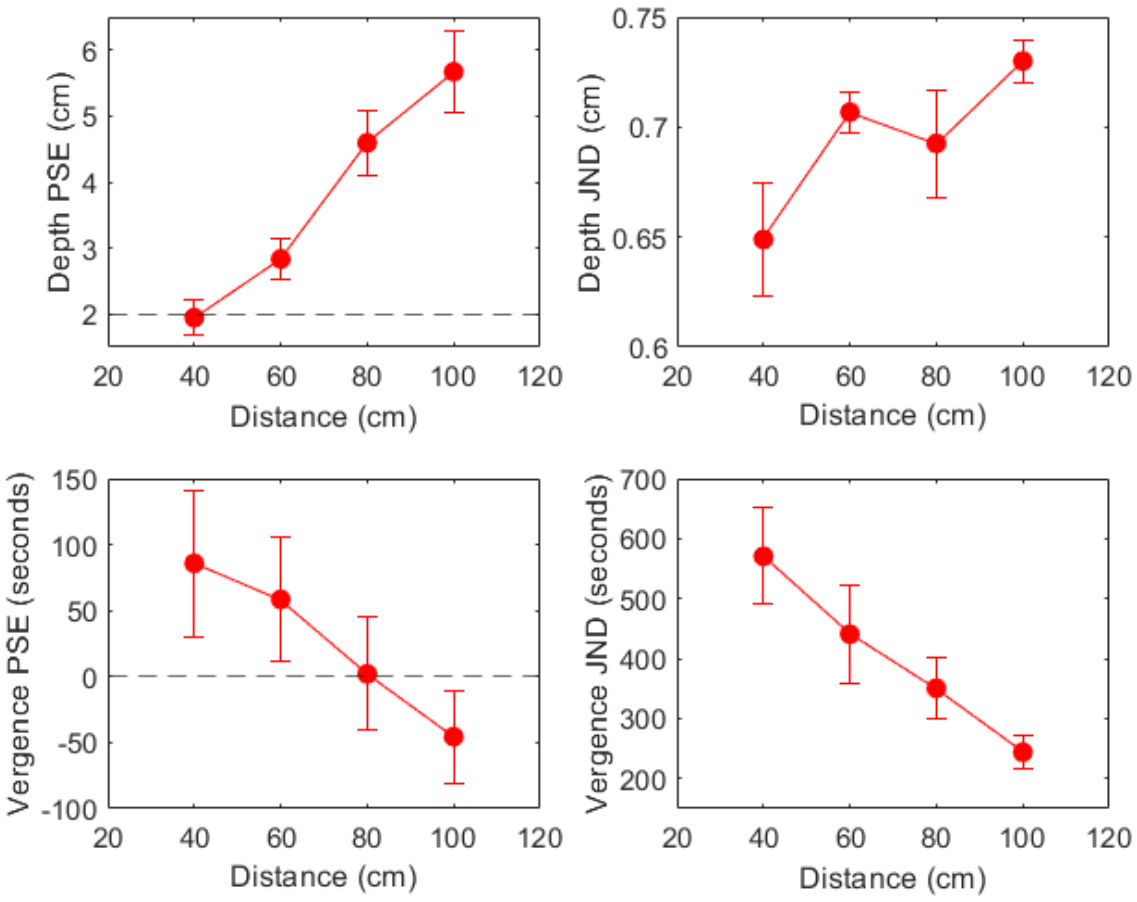


Figure 3.14: Vergence and depth task results. Graph showing mean (SE) PSE and JND scores for the depth and vergence tasks. Black dashed line shows predicted performance.

As this experiment uses a repeated measures design, linear mixed effects models were used to account for the fact that residuals between scores for one observer are more similar than the residuals between observers. By including a grouping variable in the model, this allows their individual intercepts and slopes to vary from the average (Morrell, Pearson, & Brant, 1997). The formula used for the linear mixed effects model is:

Equation 3.13

$$pf \sim 1 + d \ + (1 + d \mid o)$$

This tested whether the fitted psychometric function parameter (pf) of PSE or JND changed significantly over distance (d) as a fixed factor, with random slopes and intercepts and a grouping factor of observer (o). This model was fit four times to account for the depth task PSE (pd) and JND or sigma (sd), and the vergence task PSE (pv) and JND or sigma (sv).

All four parameters varied significantly with distance (***p*<.05** with confidence intervals not including 0), as shown in Table 3-1. This model takes into account the random slopes and intercepts, but it was found that changing this model to other combinations of random factors did not significantly change the estimates, and the full random slopes and intercepts model provided the best goodness of fit using the lowest AIC values.

Table 3-1: LME results for PSE and JND of both tasks.

Variable	Model	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
Depth PSE (pd)	$pd \sim 1 + d + (1 + d \mid o)$	0.0647	0.0082	**138**	**<.001 *****	0.0484	0.0809

Depth JND (sd)	$sd \sim 1 + d + (1 + d \mid o)$	0.00115	0.0004	**138**	**.013***	0.00024	0.00205
Vergence PSE (pv)	$pv \sim 1 + d + (1 + d \mid o)$	-2.252	0.638	**138**	**.001****	-3.514	-0.990
Vergence JND (sv)	$sv \sim 1 + d + (1 + d \mid o)$	-5.35	1.40	**138**	**<.001 *** **	-8.13	-2.57

As can be seen, the PSE for the depth task does indeed increase with increasing stimulus distance as predicted, showing a general failure of shape constancy. The participants were instructed to set the depth of the triangle to half the height of the 4cm base, and we find that observers set the depth veridically at 40cm, with increasing errors (underestimation) with increasing viewing distance. The JND in the depth task also increased with distance as predicted, although to a lesser extent than the PSE.

Figure 3.14 shows both the PSE and JND scores for the vergence task decreasing with increasing stimuli distance. A systematic bias was observed in the PSE scores in the vergence task. The dashed line indicates performance with no bias, as the PSE score here indicates the point at which the lines appear to be lined up, and this was not expected the change with increasing stimuli distance. However, it can be seen that, in general, participants were the most accurate at 80cm, with accuracy increasing with increasing stimuli distance, as shown by the slope of the red line. This suggests that participants were fixating on a point closer than the point at which the stimuli were being presented at the far distance, causing a slight

crossed disparity, and fixating on a point beyond the screen at a closer distance, creating an uncrossed disparity. This fixation disparity is shown in Figure 3.15.

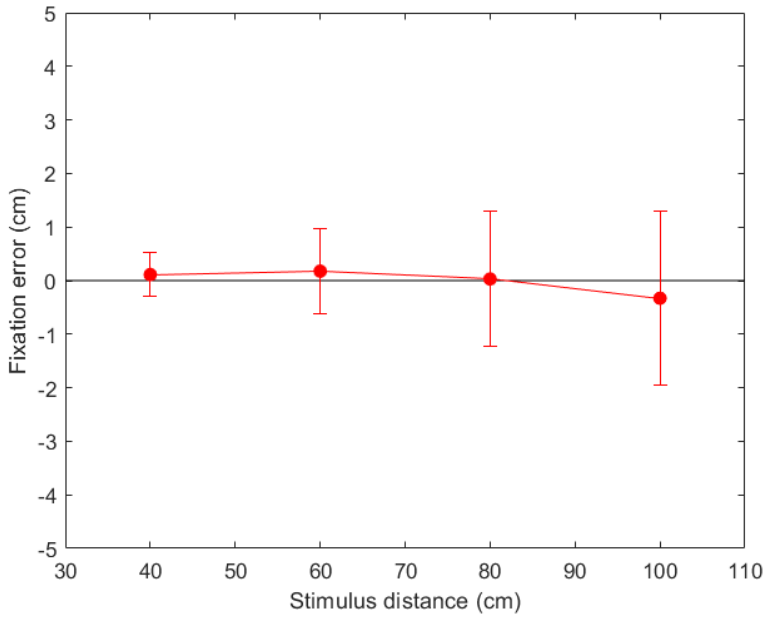


Figure 3.15: Fixation error. Graph plotting mean (SE) fixation error, showing change from crossed to uncrossed disparity with increasing stimulus distance.

To explore this potential fixation disparity effect, the bias and noise of vergence were correlated using the PSE and JND. A modest correlation was found between vergence noise and bias at the closer distance of 60cm, however confidence intervals denote a wide range:

Table 3-2: Correlation coefficients for PSE and JND in the vergence task.

Distance	R	DF	*p* value	Lower CI	Upper CI
40cm	0.19	33	.272	-0.15	0.49
60cm	0.37	33	**.028***	0.04	0.63
80cm	-0.04	33	.810	-0.37	0.30
100cm	0.03	33	.879	-0.31	0.36

An LME model with random slopes and intercepts was fit with the following equation to see if fixation distance (fd) is predicted by viewing distance (d), when grouped by observer (o).

Equation 3.14

$$fd \sim 1 + d + (1 + d|o)$$

The ideal observer would be expected to have a slope of 1. Here, stimulus distance was found to predict fixation distance very highly, as can be seen in the figure below, with an estimate of 0.99 (***p*<.001**, 95% CI [0.98 1.00]). As fixation distances, despite these small biases, were largely congruent with the veridical stimulus distance, fixation disparity alone does not explain the findings.

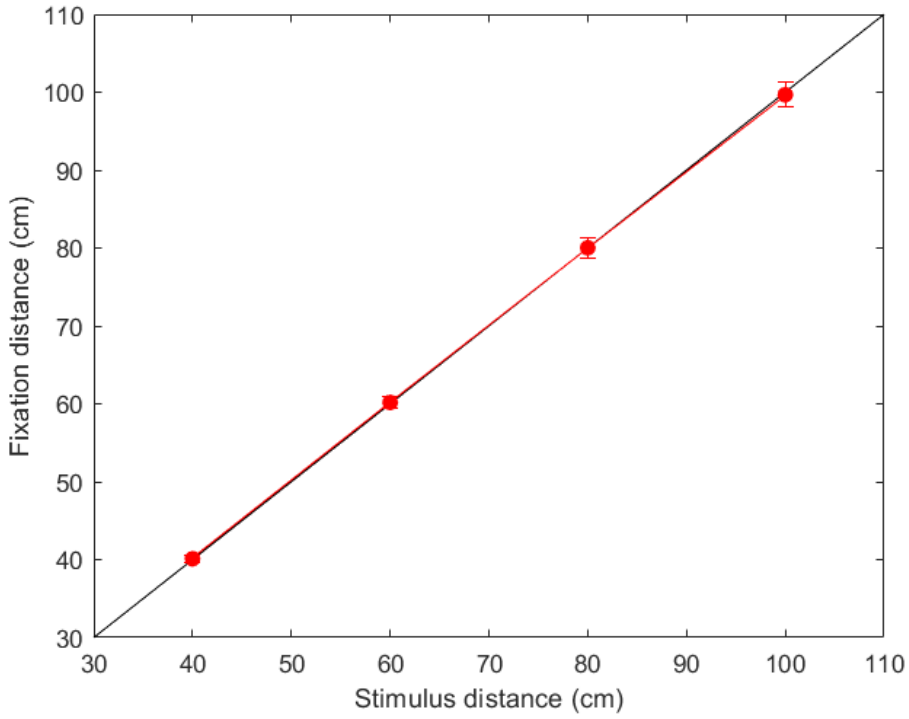


Figure 3.16: Fixation distance. Graph showing the mean (SE) calculated fixation distance against veridical stimulus distance.

3.4.3 Shape constancy

The disparity produced by a depth of 2cm will reduce with the square of distance. However, as stated above, the observed bias in the vergence task suggests that observers are fixating a point other than the target distance, known as fixation disparity. A possible source of error is if observers are using the fixated distance, rather than the target distance, by which to scale disparity. To estimate the extent to which this might explain the biases in depth seen relative to height, we calculated the fixation distance and plotted it relative to the target distance. This is shown by the black line in Figure 3.17.

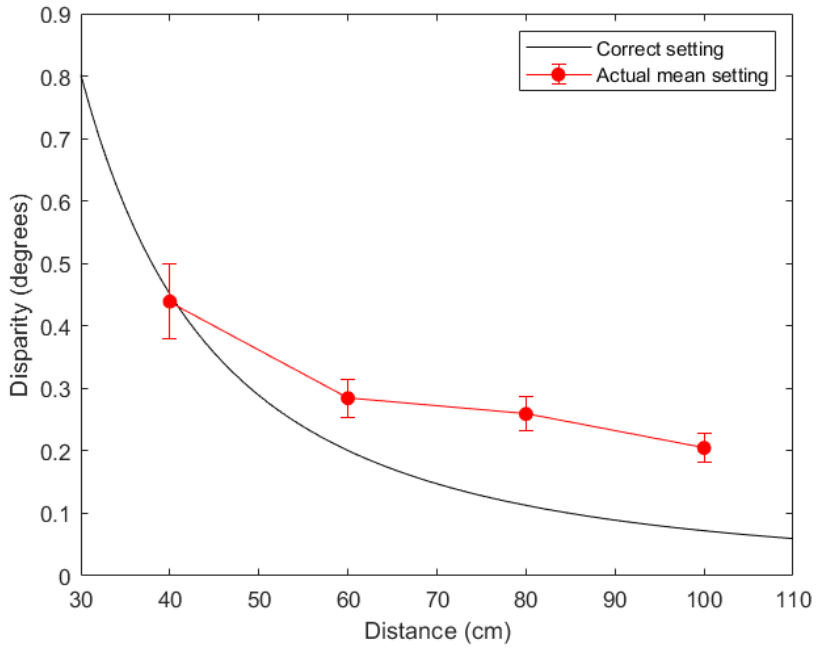


Figure 3.17: Fixation disparity. Black line shows the expected correct settings for the target distance against actual mean (SE) settings made by observers in red for fixation distance for each stimuli presentation distance.

The red data points show the mean disparity in degrees derived from the depth settings made by observers. Disparity does not remain the same over all viewing conditions, showing that observers do indeed use distance to scale

vergence information in order to make a judgement of depth. However, observers' actual mean settings do not match the expected correct settings beyond 40cm viewing distance. This shows that observers are scaling depth judgements from vergence information with distance, but that this scaling is incomplete as they are underestimating distance in the far viewing conditions.

The degree of disparity scaling can be quantified from the approximate geometrical relationship between disparity (γ), depth (z) and distance (D) by IOD (I):

Equation 3.15

$$tan(\gamma) = \frac{Iz}{D^2}$$

Taking the log of each sides gives:

Equation 3.16

$$\log\big(tan(\gamma)\big) = \log(Iz) - 2\log(D)$$

With full scaling, the slope of log disparity with respect to distance should have a value of -2. This can be used to assess the degree to which disparity settings actually scale with distance. Rewriting the equation above to include the free parameter k, and then taking the log gives:

Equation 3.17

$$tan(\gamma) = \frac{Iz}{D^k}$$

Equation 3.18

$$\log\big(tan(\gamma)\big) = \log(Iz) - k\log(D)$$

A linear regression was used to compare log(disparity) against log(distance) to get a measure of the degree of scaling. Here, we found a slope value of -0.65, considerably less than the value of -2 that would indicate full scaling, and also different to linear scaling of disparity with distance, which would give a value of k=-1.

3.4.4 Shape constancy and certainty of vergence

A correlation analysis was performed between the PSE in both the vergence and depth tasks to see if biases in vergence predicted biases in the depth estimates. No significant relationship was found between these, as shown in Figure 3.18 suggesting that the observed biases in vergence are not linked to the failure of shape constancy:

Table 3-3: Correlation coefficients for JND of vergence and PSE of depth.

Distance	R	DF	p value	Lower CI	Upper CI
40cm	0.08	33	.630	-0.26	0.41
60cm	0.06	33	.722	-0.28	0.39
80cm	0.06	33	.732	-0.28	0.39
100cm	-0.01	33	.958	-0.34	0.33

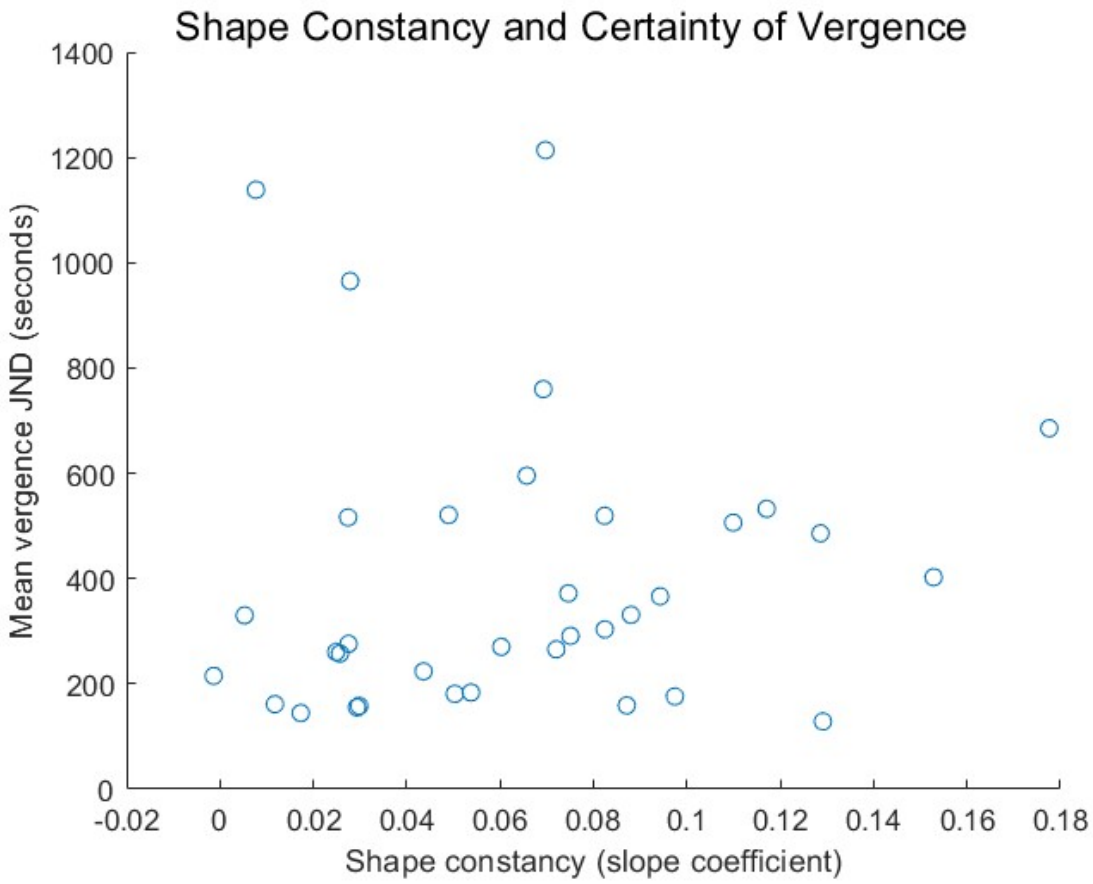


Figure 3.18: Shape constancy and certainty of vergence. Graph correlating shape constancy shown as the coefficient of the slope of PSE across the four distances, with the mean vergence JND across all four distances in seconds.

Likewise, the JND in the depth and vergence tasks were correlated to see if uncertainty in vergence predicts uncertainty in depth estimates. No correlation was found between the JND for the depth and vergence tasks at any distance, showing no relationship between the noise and therefore uncertainty for vergence and depth judgements:

Table 3-4: Correlation coefficients for JND in the depth and vergence tasks.

Distance	**R**	**DF**	***p* value**	**Lower CI**	**Upper CI**
40cm	0.22	33	.197	-0.12	0.52

60cm	0.06	33	.733	-0.28	0.39
80cm	0.06	33	.726	-0.28	0.39
100cm	-0.00	33	.990	-0.34	0.33

Evidence for a relationship between shape constancy and certainty of vergence was assessed. To evaluate change in participants' bias in depth estimates with increasing stimuli distance, and therefore a measure of the failure of shape constancy, a regression slope value was obtained for the four PSEs in the depth task for all participants. To measure certainty of vergence, an average JND score across the four viewing conditions was calculated for the vergence task for all participants. The regression slope value of the PSEs in the depth task was correlated with the average of their JND scores in the vergence task. This showed no evidence of a relationship between certainty of vergence and shape constancy, (R^2 = 0.09, p=.619, 95% CI [-0.254 0.409]).

3.5 Discussion

3.5.1 Findings

The aim of the work in this chapter was to test the explanation that errors in depth perception may be due to uncertainty when relying on the cue of vergence. This was achieved by measuring the change in depth perception across viewing distance and the certainty of vergence using a measure of vergence noise, and looking for a relationship between those two measures. The findings of this work will be explored here in detail.

To test the explanation for the main research question that failure of shape constancy could be due to uncertainty of vergence, the regression slope of the PSE scores for the depth task as a measure of shape constancy and the average JND score from the vergence task as a measure of vergence uncertainty were correlated. Results found no evidence of a relationship between these. Therefore, it is unlikely that those who have better shape constancy do so because they are more certain of relying on the cue of vergence.

Moving on to the other predictions, on average, observers displayed an increase in systematic bias in depth estimates with increasing stimuli distance as predicted. This supports the work of many previous studies that show that depth is underestimated at far distances (Baird, 1970; Brenner & van Damme, 1998; Johnston, 1991; Scarfe & Hibbard, 2006; Viguier, Clément, & Trotter, 2001). We also found that observers were able to accurately use vergence to recover the distance information to scale depth judgements at a viewing distance of 40cm only, in line with previous research that has reported accurate absolute estimations from vergence below 50cm (Foley & Held, 1972; Komoda & Ono, 1974; Mon-Williams & Tresilian, 1999). The JND for the depth task also increased with increasing stimuli presentation distance, which is predicted from the fact that disparity reduces with the square of distance (Stevens, 1975).

No significant relationship was found between the JND and PSE for vergence, meaning variation in vergence bias across individuals was unrelated to vergence noise. The JND in the vergence task was found to decrease with distance as expected. A systematic bias was observed in the PSE scores for that task, meaning the point at which the lines appeared lined up to participants shifted between viewing

distance conditions. This indicated that participants were fixating on a point closer than the point at which the stimuli were being presented at the far distance creating a crossed disparity, and at a point beyond the screen at a closer distance, creating an uncrossed disparity. An observed fixation disparity can be due to the effective scaling distance outlined earlier in the work, where the distance is misestimated resulting in judgements instead based on the scaled perception of the observer, not the veridical distance (Bradshaw, Parton, & Glennerster, 2000). However, fixation distances were found to largely match the stimulus distance, so this could not account for the findings.

This fixation disparity bias mirrors that of work by Jaschinski (1997), where observers' fixation disparities were measured using nonius lines at 20, 30, 40, 60 and 100cm. They found that fixation disparity changed from 1 arc minute eso to 3.5 arc minute exo with decreasing viewing distance, meaning that at far distances observers were converging on a point in front of the target creating crossed disparity, and at close distances converging on a point beyond, creating uncrossed disparity. This bias in fixation disparity could account for some degree of misestimation of distance.

In the present work, a small negative trend in disparity across viewing distance was found, which suggests that observers were indeed scaling disparity information with perceived distance. However, this change in set disparity was much less than required for full constancy, suggesting that observers were underestimating distance in the far conditions, and therefore not correctly scaling the vergence information, resulting in underestimations of depth at the far distances.

The measure of scaling from log transformations of disparity and distance did not find evidence of full scaling, but results also did not indicate that observers were scaling disparity linearly with distance. Rather, the scaling measured indicates that observers are scaling disparity with distance, but not quickly enough to conform to this pattern. It does not make sense to interpret this bias as scaling distance as other works have in this area (Johnston, 1991). Instead, we conclude that errors in vergence and fixation disparity cannot directly explain the observed failure of shape constancy.

No correlation was found between PSE in the vergence and depth tasks, indicating that bias in vergence does not predict the biases in depth perception. Likewise, JNDs did not significantly correlate between the vergence and depth tasks, showing that certainty of depth estimates cannot be predicted by certainty of vergence.

This chapter presented recent work by Linton (2022) to give an overview of the reasoning behind triangulation between the eyes and the fixated object, in order to explain the logic behind this experiment. However, instead of just presenting this work, Linton also raises some concerns with triangulation for stereo depth perception, arguing that triangulation is a poor fit for disparity and vergence information due to empirically observed biases. Indeed, the present work found evidence of a fixation disparity bias for vergence judgements.

Linton (2020) also finds that observers cannot effectively use vergence information to make absolute judgements of depth. This is important in the context of this work as, while observers were comparing the triangle presented to them against the ‘perfect’ triangle, which would be a relative depth judgement, this ‘perfect’

triangle was not shown to them, instead relying on them successfully recovering an accurate absolute measurement of the base of the triangle in order to picture the depth as half its height. This reliance on an internal standard may have caused some bias when comparing this against the stimuli shown. However, it should be noted that settings were on average unbiased at the closest distance.

Instead of traditional cue integrations which group information into a single percept, Linton (2022) instead suggests a two-stage model. In this model, he describes a two-stage process by which we first resolve the 3D structure of scenes through perception of visual experience, and a second stage where other cues such as motion and shading use this perceived 3D structure to understand the scene through 3D cognition, arguing that perceptual depth cues are innate, and cognitive depth cues are learned thorough experience. Linton (2022) goes on to claim that observed individual differences in perception are likely due to cognitive mechanisms.

For the present work, it could be argued that the depth task is a cognitive, not a perceptual one, and, indeed, some participants reported finding the task very challenging, due to having to 'imagine' the depth of the triangle by translating the object to picture the side view and recover the depth. Rather than getting a true measure of depth perception and therefore shape constancy, this type of task may instead be measuring participants' spatial reasoning skills, a set of cognitive abilities relating to the awareness and decoding of spatial concepts like distance and dimensions (Fowler, et al., 2021). If this were the case, the findings here may therefore suggest that spatial reasoning is not linked to certainty of vergence, rather than shape constancy. One potential change to explore this idea would be to replace

the depth task in this design with a more perceptual one and see if similar results are gained.

In summary, this chapter provides a key starting point for this thesis by examining the contribution of the cue of vergence in depth perception. Ultimately, we found evidence of a failure of shape constancy, although this cannot be explained with vergence behaviour. The findings of this research do, however, highlight areas where future work is still needed. Some suggestions for this are listed below.

3.5.2 Other future work

One potential change to the present methodology is discussed above in the context of Linton's (2022) theory on perceptual versus cognitive judgements. Further suggested work is presented here.

A possible variable to consider for future work would be the effect of repeated trials on participants' judgements. There are two areas that could have affected results here. The first is a practice effect, in that participants would improve with increasing number of trials. To address this potential issue, blocks were randomised between participants, but due to the physical nature of changing the distance of the screen, all trials were completed at the same distance before moving on to a new block. Comparing certainty of vergence across the experiment would be interesting to explore if practice using this cue improves certainty of relying on it. Future work could simulate different presentation distances between trials, and track certainty across presentation order to address this.

The second thing to note is the potential for a contraction bias, as observed by Tresilian, Mon-Williams and Kelly (1999). This bias is based on prior sensory

knowledge and experience. Here, observers would aggregate an ‘average’ stimulus from the pool of trials presented to them, which influences future decisions, so that depths bigger than the average are underestimated, and those smaller are overestimated. Within this work, this may mean that participants were influenced by the initial trials presented by the psi-marginal method, subconsciously using these depths as the average against which to compare judgements, rather than the ‘standard’ stimulus instructed. This has not formed part of the analysis in this work, as one way to combat a contraction bias would be to use a between participants design to ensure no range effect. However, the present work was interested in the individual differences of observers’ depth judgements across distances to show a measure of shape constancy, so this adaptation would not be appropriate for this work.

This experiment sat participants with their heads in a headrest for the duration of the experiment. This reduced any extraneous information from motion parallax as that was not of interest in this chapter. However, Bradshaw, Parton and Glennerster (2000) highlight the critical importance of motion parallax in depth from disparity tasks, and future experiments in this field should consider adding this depth cue into the design.

One criticism of the depth task presented to participants is the lack of a presented standard stimulus. Participants were told prior to the start of the experiment that the standard stimulus was a triangle where the depth was equal to half of the height of the base. However, this was shown through a diagram on the instructions sheet, and not displayed on screen. It was decided that the standard stimulus would not be shown, because, as the task is completed at four distances, a

standard stimulus reference provided at just one of these distances as an example for participants prior to the start of the experiment would have provided an extra benefit for that block compared to other distances.

However, it has been argued that providing this reference for participants in a yes/no forced choice task is important (Wixted, 2018). Some stimuli provide a natural reference point, such as motion or no motion. However, for the depth task presented in this experiment, there is no natural reference point. In such cases, participants are required to store an image of the standard stimulus in long term memory, in order to compare the presented stimulus against this remembered reference. As such, this type of task is better suited to a design where the participant can become familiar with the reference, such as through extensive training. Alternatively, to avoid this route, a reference stimulus may be presented at each trial against which participants can compare the presented stimulus. However, neither option is suited to the design of this experiment, although a possible alternative would be to present a reference stimulus once at the beginning of each block, which would tackle the problem with one distance receiving an advantage over others.

Another consideration is that by providing a reference standard stimulus at, for instance, a further distance, it would highlight to participants the bias associated with that increased stimuli distance. This is because, due to the systematic bias, participants would view the standard stimulus reference as too shallow at far distances as well, and would be aware that they would need to adjust their responses accordingly, which would eliminate the observed bias, not because participants would no longer be seeing objects as too flat, but because they would be matching their responses to the standard stimulus, now perceived to be too flat. It is

theorised that by providing participants with this reference triangle prior to the start of each block at each distance, the systematic bias observed would be reduced, or even eliminated. This may be an interesting avenue to explore for future work in this task, comparing performance when a reference stimulus is made available, compared to here where none was presented. Despite this, the pattern of results across conditions is what was of interest in this experiment, and therefore it was necessary for this study to allow participants to create their own image of the reference stimuli in their minds against which to compare presented triangles.

This issue would also arise for the vergence task, as if participants were presented with a pair of lines that were supposedly completely aligned, and told that this was a reference for the lines being lined up, but participants were subject to the systematic bias observed in the PSE for the vergence task, they would possibly view the lines as not exactly lined up, and know that what they were perceiving was not quite what was being presented. As such, they may attempt to recalibrate their responses to allow for this bias they could detect in their own perception. The benefit of the vergence task over the depth task, however, is that participants have experienced lines being lined up prior to the experiment, so the memory issues highlighted above do not apply, and pure perception may be measured in this task. To improve upon the design of this work, a depth task that does not rely upon long term memory of a standard stimulus may be better matched to the design of the vergence task.

4 Evidence that ‘Dark is Deep’, but only at far distances

4.1 Abstract

The rule that ‘Dark is Deep’ has been employed by artists for centuries to convey depth in a scene. This is built on the mechanism that surface luminance under diffuse lighting conditions determines that surfaces locally furthest away from a light source receive the least amount of light, as light rays are naturally occluded within crevices in a scene, leading to less light being reflected and therefore less luminance. One study reported an effect of surface luminance manipulation on the perception of depth in naturalistic scenes, but only for monocular viewing (Hibbard, Goutcher, Hornsey, Hunter, & Scarfe, 2023). However, these stimuli were only presented at a distance of 50cm, where binocular cues are more reliable and therefore weighted more highly under Bayesian cue combination rules. Therefore, the lack of overall luminance manipulation effect may be due to conflicts between the pictorial cues, such as shape from shading, which would be affected by the change in luminance, and the more highly weighted binocular cues, which would not report a change. The present study was designed to explore this possibility, and see if the surface luminance effect varied between close and far distances. Using the ‘Dark is Deep’ rule, pixels closer to the observer appeared more illuminated than further pixels on the 3D model. This rule was manipulated by either reducing or enhancing the luminance between the closest and furthest pixels. Estimations of depth were measured by the length of a line adjusted by participants to represent the distance between two dots presented on the objects. It was predicted that at a close distance, a similar effect of luminance manipulation to previous work would be observed.

However, luminance manipulation was expected to have an increased effect at the further distance, where binocular cues are less reliable, reducing their weighting in the average. This was expected to lead to an increase in the strength of the luminance manipulation, showing the contribution of shape from shading in the overall depth estimate. Results failed to replicate the main effect of luminance manipulation from the previous study, meaning in general, distance estimation did not reduce or increase in line with a reduction or enhancement of surface luminance as predicted. However, some evidence of an interaction between the ‘Dark is Deep’ rule and monocular or binocular viewing was observed at a further distance. This suggests the manipulation in this experiment was not sufficient to combat the estimates reported by binocular cues, except at the further distance where binocular cue weighting was reduced.

4.2 Introduction

4.2.1 Lighting and luminance

Directional lighting, such as that provided by the sun in a cloudless sky, illuminates surfaces so that their brightness varies with the angle at which the light source hits the surface plane (Horn, 1970). Under these lighting conditions, surfaces facing towards the light source reflect more of the light towards the observer’s eyes than surfaces which are not so directly facing the direction of the light source. This can be seen in Figure 4.1a, with Figure 4.1c illustrating the pattern of luminance.

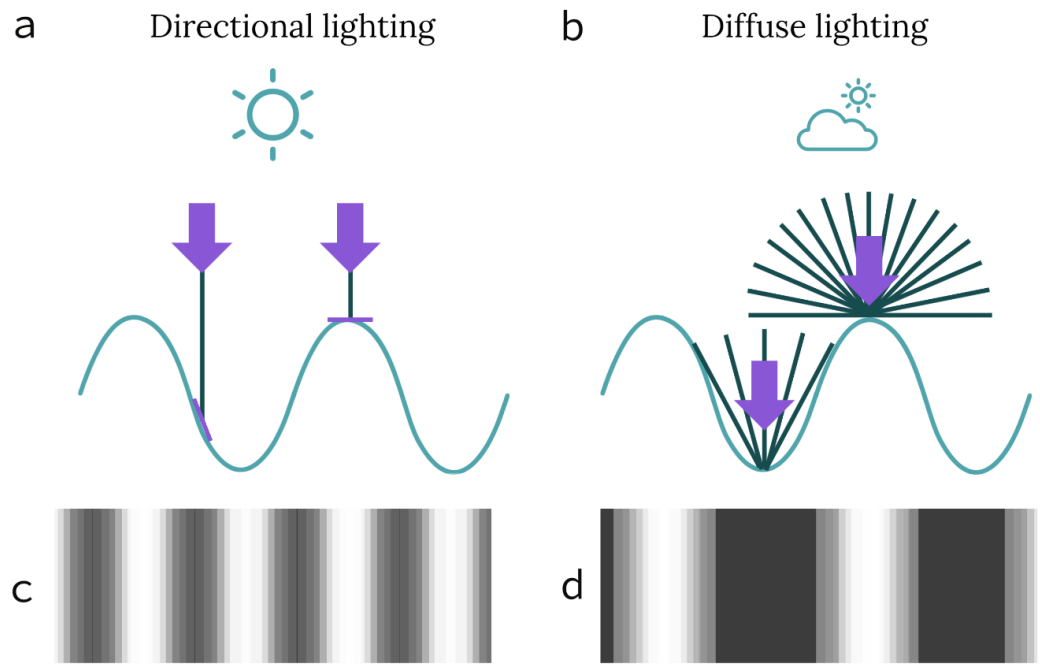


Figure 4.1: Shading by light source. Showing two sources of shading, from directional and diffuse light sources. Image (a) shows how light rays from a directional light source hit the surface of an object, and how the light reflected is relative to their orientation. Image (b) shows how light rays reflect from an object's surface under diffuse lighting conditions, and how these differ depending on the occlusion of the object surface, such as a crevice. Images (c) and (d) show the pattern of light reflectance and luminance for directional and diffuse lighting respectively.

In Figure 4.1b, the light source is diffuse, so the rays of light are scattered around the observable scene, as opposed to the example of the directional light source. With diffuse lighting, which can occur, for example, on a cloudy day, where the light rays are not all able to directly reach the surface, depth from shadowing occurs according to the rule 'Dark is Deep' (Tyler, 1998; Zhukov, Iones, & Kronin, 1998; Landis, 2002; Bredow & Imageworks, 2002; Méndez-Feliu & Sbert, 2009; Schofield, Rock, & Georgeson, 2011; Todd, Egan, & Kallie, 2015; Cooper & Norcia, 2014). Under this rule, points furthest from the light source appear darkest, due to reflectance of rays being naturally occluded by crevices in the surface of or between objects, as shown by the pattern of luminance in Figure 4.1d. Here, the amount of

light that reaches the surface determines how much luminance it appears to have, meaning that surfaces further away from the observer will appear darker, hence the rule ‘Dark is Deep’ (Potetz & Lee, 2003; Chen & Tyler, 2015; Scaccia & Langer, 2018; Hibbard, Goutcher, Hornsey, Hunter, & Scarfe, 2023).

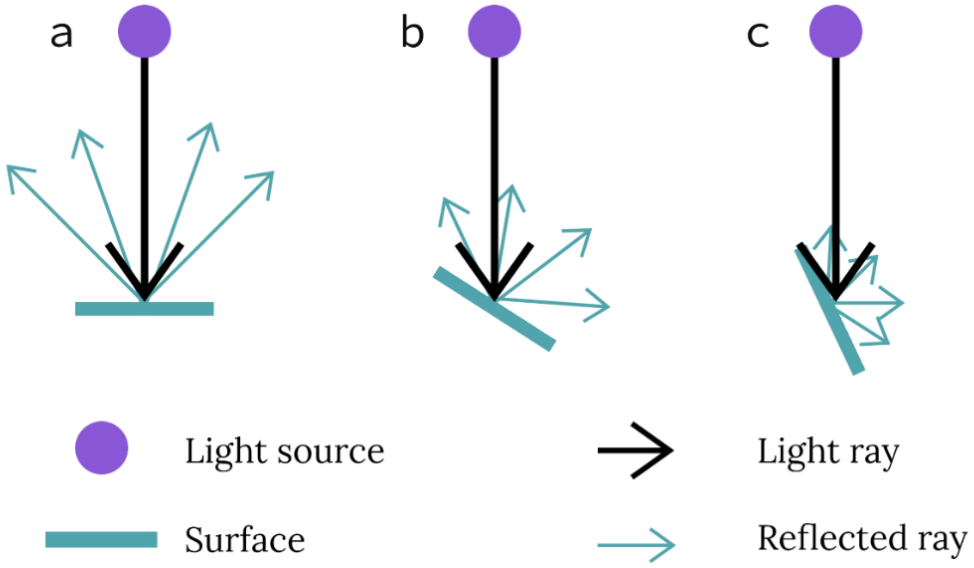


Figure 4.2: Angle of reflectance. Image showing how the direction of the light source relative to the orientation of the surface affects the intensity of light reflected from the surface.

At point Figure 4.2a, the surface of the object is directly facing the directional light source, with a slant of 0 degrees, and therefore will reflect more light into the observer’s eye. At point Figure 4.2b, the surface is less directly facing the light source, so less light will be reflected into the observer’s eye. Figure 4.2c shows a point which is rotated almost completely away from the light source and therefore very little light will be reflected directly from this surface, meaning that this surface will likely be in shadow to the observer. Therefore, shading under these lighting conditions is created by the angle of the surface relative to the directional light source.

This mechanism whereby dark indicates depth has been used for centuries in the art world to create the illusion of depth within paintings (Hibbard, Goutcher, Hornsey, Hunter, & Scarfe, 2023). Figure 4.3 depicts a charcoal drawing of a skull. Increased charcoal pigment to create dark areas has been used to create the impression of depth, such as in the eye sockets, which give the impression of a 3D object in pictorial space.

Figure 4.3: Charcoal drawing of a skull. Image showing how varying light and dark tones are used in artworks to create the impression of depth.

4.2.2 Previous work

Hibbard et al. (2023) explored the direct effect of the luminance contrast manipulation on the perception of depth. They presented observers with a cluttered scene of naturalistic rendered 3D models of scanned food items. These were

presented in grey-scale to remove any extraneous surface colour or reflectance information, leaving just surface texture for illumination manipulation. The surface luminance of objects was manipulated either across the objects themselves, or across the entire scene. For the scene luminance manipulation, they labelled each pixel as either belonging to the background or an object in the scene. The luminance of each pixel (L_i) identified as belonging to objects were manipulated with the following equation:

Equation 4.1

$$L_i = L_o\left(\pm\alpha\frac{Z_i - Z_{median}}{Z_{max} - Z_{min}}\right)$$

Here, L_o is the original luminance (measured in candela per square metre, cd·m^{-2}), the Z median, minimum and maximum relate to the distance (in cm) for that object, and α defines the degree of luminance manipulation gain, with a value of 0.1. In the enhanced luminance conditions, points further away relative to the medium distance in the scene had luminance subtracted and those closer received additional luminance, with the opposite manipulation applied for the reduced luminance conditions.

A similar luminance manipulation was applied for the object conditions, this time manipulating luminance across the objects themselves, using this equation:

Equation 4.2

$$L_i = L_o \pm \frac{Z_i - Z_o}{\beta}$$

Here, the original luminance (L_o) of points identified as belonging to each object are manipulated using the mean distance of pixels of the object (Z_i), the viewing distance to the object (Z_o) and the degree to which the luminance of each pixel is

scaled relative to distance (β), which had a value of 100cm relative to their viewing distance of 50cm.

They measured depth perception with a task where observers had to estimate the distance between two dots superimposed on the scene by adjusting the length of a line to match their separation in 3D space. In the between-objects scene manipulation these dots were on different objects, and in the within-object luminance manipulation they were on the same object. Regression analysis found that settings in the within-objects conditions showed overestimation of set distance compared to actual distance. However, slopes values were closer to 1 in the binocular conditions, showing that more veridical depth was perceived with binocular viewing compared to monocular. They observed that larger settings were made for points on the same object when viewed monocularly, as well as in the enhanced luminance condition compared to reduced luminance for monocular viewing. However, no significant interaction between the effect of luminance manipulation and monocular and binocular viewing was found.

Observers were found to underestimate depth overall in the between-objects conditions, but to a lesser degree than the bias for the within-objects condition, meaning settings were more accurate for between-object judgements. In addition, there was no general effect of luminance manipulation, but there was an interaction representing a bigger effect for the reduced luminance manipulation when viewed monocularly. However, this experiment was only conducted at a single viewing distance of 50cm, where binocular cues would be weighted highly under a cue combination model, due to being more reliable at closer distances.

4.2.3 The present work

The present study was designed to complement the work of Hibbard et al. (2023) and to investigate the outstanding question of the effect of varying viewing distance on the magnitude of the luminance manipulation effect and its interaction with binocular and monocular viewing. It covers four main areas of manipulation to explore this ‘Dark is Deep’ rule and therefore the contribution of shape from shading under diffuse lighting conditions to the overall depth estimate. The first area is the effect of viewing distance, to address the question arising from the work of Hibbard et al. (2023) outlined above. This was achieved by testing observers’ perception at both a close and a far distance. It was predicted that at the close distance, results similar to previous work would be observed, in that the luminance manipulation effect is reduced due to the conflict between pictorial cues and binocular cues, given that binocular cues are highly reliable at close distances and are therefore weighted heavily in the Bayesian cue combination model. This relationship is presented in Figure 4.4.

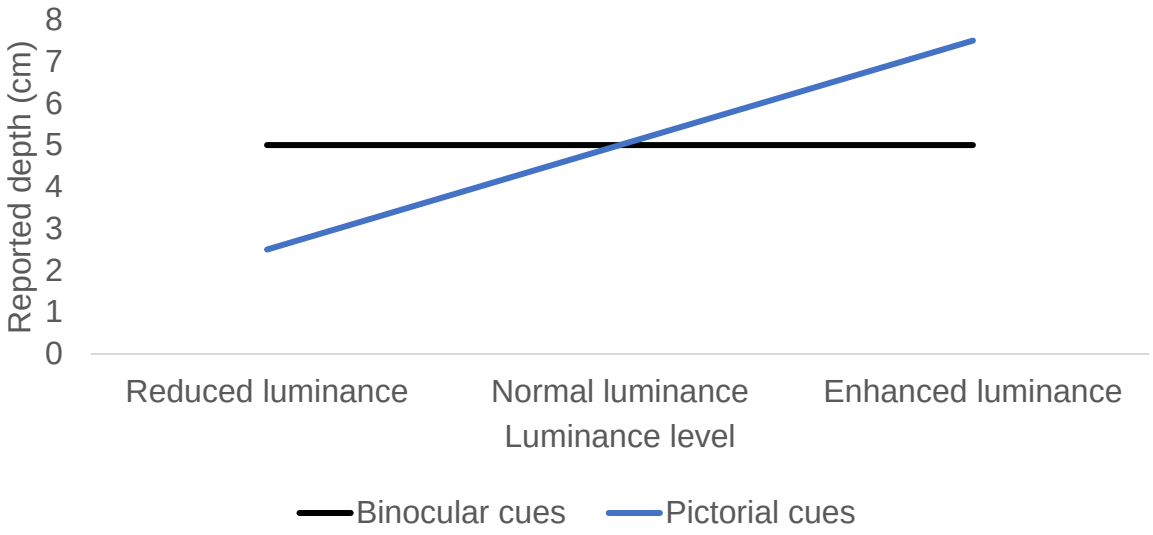


Figure 4.4: Binocular and pictorial cue conflict. Chart plotting expected reported depth from binocular and pictorial cues across luminance conditions when the cues are in conflict.

Here, depth reported by binocular cues would not be expected to change across the varying luminance conditions. However, depth from pictorial cues such as shape from shading would be expected to increase with increasing luminance manipulation. This therefore creates conflict between the binocular and pictorial cues. Under Bayesian cue combination theory for this type of cue perturbation method, presented in more detail in Chapter 1, the overall depth estimate would be expected to skew towards the estimate given by binocular cues at the closer distance, given their reliability and therefore increased weight in the averaging. As depth reported only by binocular cues does not change with luminance, this skewing effect would reduce the overall effect of the luminance manipulation, as seen by less variability between depth scores in the luminance conditions at a closer distance.

In a weighted cue averaging model, the depth would be estimated from the weighting given to each cue, relative to their individual reliabilities (Landy, Maloney, Johnston, & Young, 1995). For instance, the equation below, adapted from Hibbard et al. (2023) shows how depth (z) would be estimated from the weight of binocular disparity (w_D), the weight of luminance (w_L), and the weight of other cues available ($\mathrm{w_O}$):

Equation 4.3

$$\mathrm{z} = w_D s_B D + w_L s_C L + \mathrm{w_O} 0$$

Cues are weighted to total to one, such that here w_D + w_L + $\mathrm{w_O}$ = 1. s_B and s_C are scaling factors between depth, and disparity and contrast, respectively, which are necessary to transform sensory input from specific measures such as degrees for disparity into a standard measure of mm (Hibbard, Goutcher, Hornsey, Hunter, & Scarfe, 2023). Other cues in the scene may be limited such as in a cue perturbation

method experiment, or used to summarise the many other depth cues available for full natural viewing conditions.

Given the assumption of the weak fusion Bayesian model that cues are unbiased and have Gaussian uncertainty, consistent estimates between cues leads to a more precise estimate of depth, although does not increase the magnitude of judgements. When a conflict between cues arises, such as described above being artificially introduced between binocular and pictorial cues, the depth estimate skews towards that of the more reliable cue given its increased weight in the average.

The second manipulation uses the above effect of introducing conflicts between cues for the viewing condition, with stimuli being viewed either monocularly or binocularly. This aimed to replicate findings from the previous work of a difference at the closer distance between binocular and monocular viewing. A finding as such would support the theory that binocular cues are weighted more heavily and therefore the conflict between these and the pictorial cues would result in the depth estimate being weighted towards the depth reported from binocular cues under both monocular and binocular viewing.

The third manipulation in this work was the luminance-depth relationship, which was achieved using the methods set out by Hibbard et al. (2023). Here, the brightness of pixels was either enhanced or reduced. In the enhanced condition, the pixels furthest away from the observer were darkened, and the closest lightened relative to veridical surface brightness. In the reduced condition this effect was reversed, meaning closest points were darkened and further points lightened relative to normal. It was predicted that the effect of increased luminance manipulation would be more evident at the further distance, where pictorial cues are more highly

weighted, and therefore the magnitude of the effect would be inflated, as shown by an increase in depth with enhanced ‘Dark is Deep’ luminance.

The final manipulation in this chapter was whether the luminance manipulation was applied to the pixels across an object, so that for instance the furthest point of each object became darker, or whether luminance was manipulated across the entire scene, so that all pixels towards the back of the scene became darker. This exploration provides meaningful insight into the overall theme of this body of work, that is exploring the contribution of cues in naturalistic scenes, and how contributions of cues vary in the context of a cluttered scene. Following Hibbard et al. (2023), it was predicted that luminance manipulation would only influence depth judgements in the local, within-object condition. This is because the 'Dark is Deep’ correlation is a local cue to depth, and differences in luminance across separate objects may more naturally be interpreted as effects of surface reflectance.

In summary, this work aimed to explore the potential impact of viewing distance on depth from luminance manipulation to explore the contribution of the pictorial cue of shape from shading to the weighted depth estimate. It was predicted that more of a luminance effect would be found at the far distance, where binocular cues become less reliable and therefore weighted less compared to pictorial cues.

4.3 Methods

4.3.1 Participants

18 participants took part in this experiment, including 9 (50%) who identified as female, 8 (44%) male and 1 (6%) as non-binary. They were between the ages of 20 and 47 and were screened prior to the experiment for normal or corrected-to-

normal vision, as well as sufficient gross stereopsis. Participants included one researcher and 19 people naïve to the purpose of the experiment.

4.3.1.1 Recruitment

Participants were recruited through the University of Essex’s online SONA system, as well as through word of mouth. All participants were compensated financially for taking part.

4.3.1.2 Screening and set up

Two vision tests for normal vision and stereo acuity were administered to see if participants qualified for the experiment. Stereo Optical's Butterfly random dot depth test (Stereo Optical, 2020) was administered to screen for sufficient Stereo Acuity scores, with the cut-off point for participation being if participants could view the entire 3D butterfly, which equated to 700 seconds of arc. This was viewed through polarised glasses at a distance of 41cm (16 inches), as per the instructions. Participants were screened for normal or corrected-to-normal vision using the Lighthouse Distance Visual Acuity Test, with the cut-off point for participation being a Snellen score of 32 or better, as this gave participants a visual acuity score of 90, with the Snellen ratio 20:20 being considered normal vision, which indicates a visual acuity score of 100. Those who did not meet the criteria from the screening tests were thanked for their time and did not participate. Those who did pass the screening tests underwent two setup tasks. Observers’ IOD was measured as the distance between the two eyes using a standard ruler. This measurement was taken three times and an average used to input for the experiment. The observers’

dominant eye was assessed by holding a pen up at arm's length with both eyes open and aligning it with a marker on the wall. Observers then closed each eye in turn and reported which for which eye the mark remained aligned with the pen (Porac & Coren, 1976).

4.3.2 Design

The length of the line necessary to join together a pair of dots superimposed on a cluttered scene of naturalistic objects reported by participants was taken as a measure of the perceived 3D distance between the points (Wilcox, 2016). Four variables were manipulated: the two viewing distances of 50cm or 100cm, if the pixel illumination was reduced or enhanced, monocular or binocular viewing of the stimuli, and whether the luminance manipulation was applied across the object or the scene.

4.3.3 Apparatus

The stimuli for this experiment were generated and presented using MATLAB with the Psychophysics Toolbox extension (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) and were viewed on a 52 by 29cm monitor with a resolution of 1920 by 1080 pixels. VIEWPixx 3D synchronisation LCD shutter goggles synchronised to the 120Hz refresh rate of the screen, along with a 3DPixx IR emitter, presented a different image to each eye individually, giving a total of 60 frames per second to each eye. Participants sat with their head on a chin rest, adjusted so that the middle of the screen was at eye level for each participant, to minimise head movements during trials to eliminate additional depth cue information. Participants manipulated the length of the line on the bottom of the screen by moving a standard

computer mouse side to side, and clicked the left mouse button to confirm their judgement and move onto the next trial.

4.3.4 Stimuli

Participants were presented with a cluttered scene of six 3D models of fruits and vegetables at differing distances from the observer, as seen below in Figure 4.5. The items were arranged such that those closer to the participant naturally occluded those behind them, and were rendered in a 30 by 30cm bounding area on the horizontal plane. The models were rendered in grey-scale with no colour and only surface texture information available. The stimuli were presented against a grey background in full screen with an image resolution of 1920 by 1080 pixels. Stereoscopic image pairs were rendered for a range of adult IODs (Dodgson, 2004).

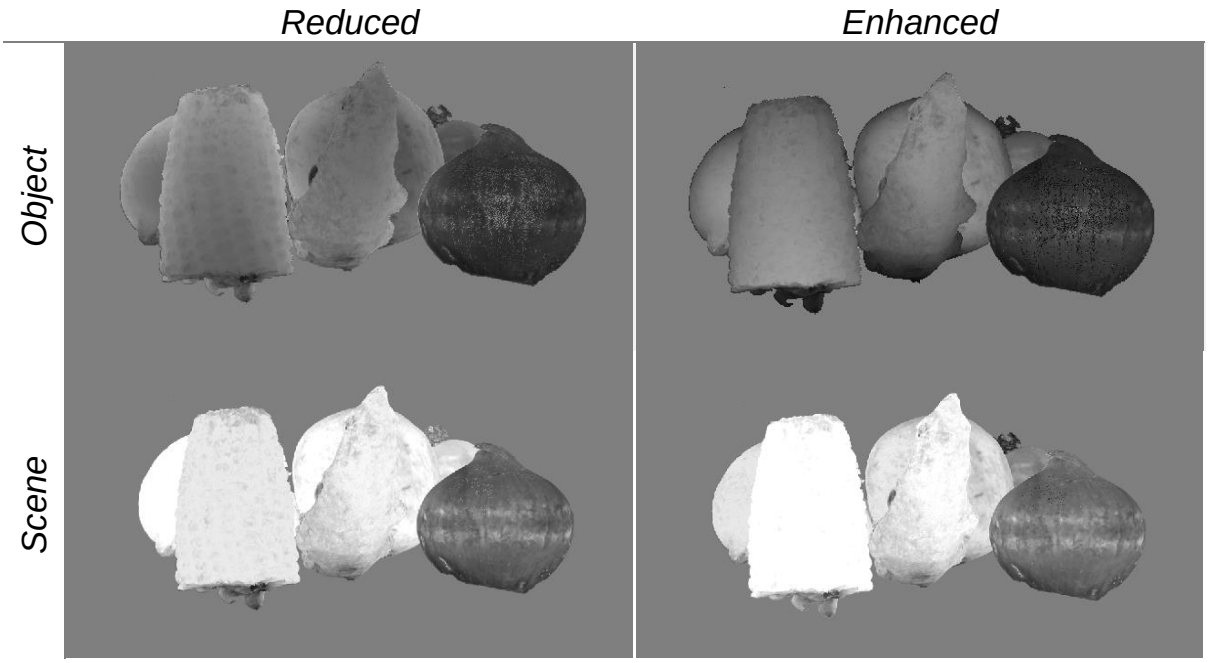


Figure 4.5: Stimuli scenes. Images showing the four stimuli conditions – reduced or enhanced luminance across either the objects or the scene. These were shown at two distances and both monocularly and binocularly, giving a total of 16 blocks.

The correlation between luminance of pixels and distance was artificially manipulated using the equations set out by Hibbard et al. (2023) presented in the introduction, which is shown in Figure 4.6. This shows that the correlation between distance and luminance gets more negative in the enhanced condition.

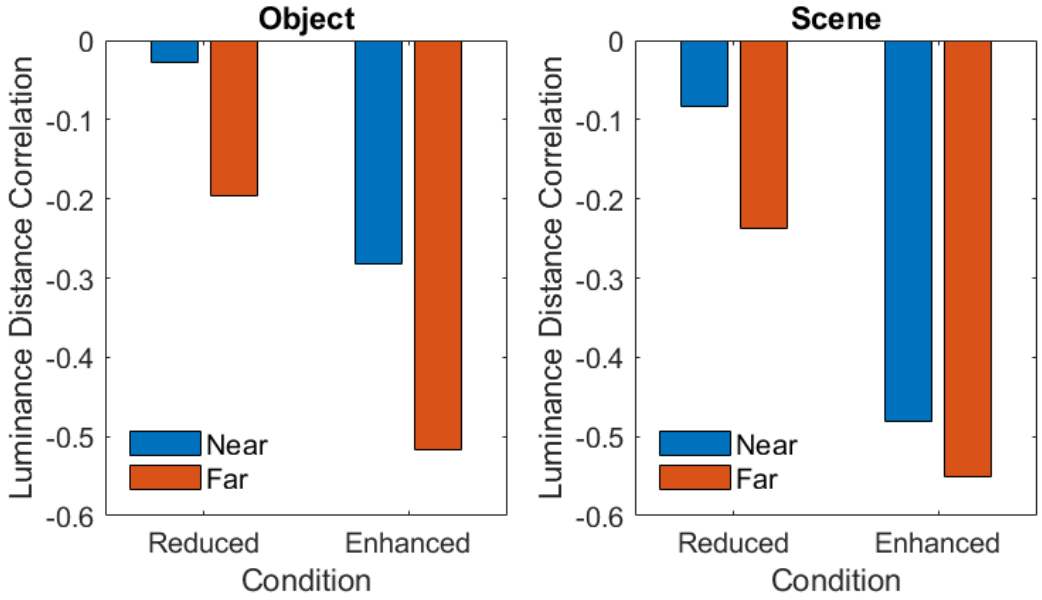


Figure 4.6: Luminance to distance correlation. Graph plotting the correlation between luminance and distance used in the experiment.

In the enhanced luminance within-object conditions, the luminance of the pixels was increased for relatively close points on the object, and decreased for relatively far points. In the scene conditions, the luminance was adjusted for all pixels across the entire scene relative to the distance of the object from the observer. For instance, in the enhanced scene luminance condition, the pixels on objects closest in the scene to the observer were lightened and the pixels on objects furthest away in the whole scene were darkened.

For each trial, a small pair of red dots was superimposed onto the scene, presented monocularly to the dominant eye. 20 pairs of dots were presented in each

block. For the object luminance manipulation condition, the pairs of dots were presented in random locations on the same object, with a mean separation in the screen plane of 2.4cm, and 2.6cm in 3D space. For the scene manipulation condition, the points were located individually at random between two objects in the scene, with a mean separation in the screen plane of 7.7cm and 11.7cm in 3D space. 6.1cm below the scene, a line was rendered, with the starting length randomised for each trial between 0 and 4cm.

4.3.5 Procedure

After obtaining written consent, participants were screened using the two measures outlined above. Those that scored under the required threshold were invited to take part. Participants completed the set-up tasks, and were then seated in a darkened laboratory, with their chin in a chin rest, facing the 3D screen. All blocks were completed in a darkened laboratory to reduce extraneous depth information, which was once again illuminated between blocks to reduce dark adaptation. Using the 3D goggles, participants were presented with the scene of scanned objects at differing distances. Somewhere in the scene, as outlined above, two red dots were placed on the objects. Participants were tasked with manipulating a line at the bottom of the screen so that it matched the distance between the two dots in 3D space, that is that the line would cut through any object in its way, much like a laser. Participants manipulated the length of the line to match the perceived distance by moving the computer mouse, moving to the left to decrease the line and to the right to increase the length. Once they were happy that the line represented the distance between the two dots they clicked the left mouse button, at which point the next set

of dots would appear to mark a new trial. The monocular conditions were conducted by using the VIEWPixx goggles to transmit the image to the observers' dominant eye, with a black screen presented to the other. In total there were 20 trials per block, and participants completed a total of 16 blocks altogether, corresponding to the four manipulated variables, that were randomised between participants. At the end, participants were thanked for their time and debriefed.

4.4 Results

4.4.1 Data treatment

Raw scores in the experiment were recorded in the form of the length of the line indicated by participants to represent the distance between the two red dots. As the objective ground truth of the location of the objects was known from the 3D rendered scenes, the veridical distance between the two red dots was calculated. These were compared to the settings made by observers to get a measure of how accurate judgements were overall.

4.4.2 Accuracy of judgements

As this experiment had a repeated measures design, a linear mixed effects (LME) model was selected to account for covariance of scores between observers, using the following equation:

Equation 4.4

$$d \sim 1 + g + (1 + g|o)$$

Here, a slope of 1 would indicate perfect performance. The ground truth settings (g) were found to strongly predict the distance settings made by observers (d), (β=1.117, ***p*<.001**, 95% CI [1.044 1.191]), although results show that in general there was an overestimation of distance by around 12%.

From the ground truth distances, an error score was calculated for each distance judgement made by participants. This was calculated as the distance reported divided by the ground truth distance, to give a relative error score for each trial. A score of 1 would indicate an accurate response, with scores less than 1 indicating an underestimation of distance compared to the veridical distance between points, and scores greater than 1 representing an overestimation. If the luminance manipulation acted as a depth cue as in Hibbard et al. (2023), the ratio should be smaller for the reduced luminance condition, and larger in the enhanced luminance condition.

To begin with, the overall error in distance estimates was compared for the close and far distances. Significantly less error was observed at the far viewing distance compared to the close distance, (t(286)= 2.062, ***p*=.040**, 95% CI [0.005, 0.225]). Additionally, the general within-objects and between-objects error rates were compared. Significantly less error was observed for local within-objects judgements than for global between-object judgements, (t(286)= -3.901, ***p*<.001**, 95% CI [-0.322, -0.106]).

The remaining analysis was split into four categories to replicate Hibbard et al. (2023). Firstly, as with their study, split by either object or scene manipulation, and secondly, for our extension of their work, split by close or far distance. Therefore, the groups are the within-objects conditions at the close and far distance, and the

between-objects, or scene, conditions at the close and far distance. Mean (SE) relative distance errors for each of the four groupings are plotted in Figure 4.7.

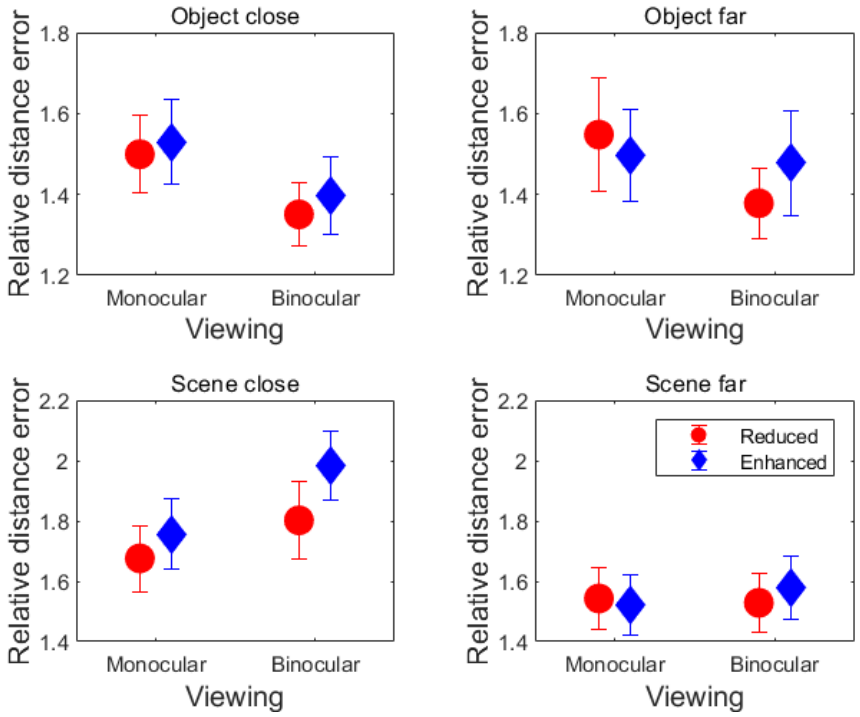


Figure 4.7: Relative distance errors. Plots show the mean (SE) relative distance errors for each of the four groups, plotting monocular and binocular viewing and reduced or enhanced luminance condition.

4.4.3 Object close

First, the judgements for the within-object luminance manipulation at the close distance were analysed. An LME model was used to see if enhanced or reduced luminance condition (c) or monocular or binocular viewing (e) predicted the relative error in distance estimation (err):

Equation 4.5

$$err \sim 1 + c * e + (1 + c + e|o)$$

This model compares the two main fixed effects, allowing for an interaction between them, and random intercepts and slopes for both condition and monocular or binocular viewing. Results are shown in Table 4-1:

Table 4-1: LME results for predicting relative error in distance estimations in the object close group.

Predictor	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
Condition (c)	0.030	0.034	68	.374	-0.037	0.097
Monocular or binocular (e)	-0.149	0.033	68	**<.001*** **	-0.214	-0.083
Interaction ($c * e$)	0.016	0.034	68	.644	-0.052	0.084

For the object close group, reducing or enhancing the luminance of pixels did not significantly affect the relative error rate. Neither was there an interaction with viewing condition. However, there was a significant negative effect of monocular or binocular viewing condition, showing that 15% smaller errors were made in the binocular conditions compared to the monocular conditions.

4.4.4 Object far

Next, the judgements were compared for the within-object conditions at the far distance, using the same LME equation as before in Equation 4.5. Results are shown in Table 4-2:.

Table 4-2: LME results for predicting relative error in distance estimations in the object far group.

Predictor	Estimate	SE	DF	*p* Value	Lower CI	Upper CI

Condition (c)	-0.052	0.043	68	.224	-0.137	0.033
Monocular or binocular (e)	-0.170	0.046	68	**<.001*** **	-0.261	-0.078
Interaction ($c * e$)	0.151	0.059	68	**.012***	0.034	0.269

As before, there was no significant main effect of reducing or enhancing luminance condition. As with the object close results group, monocular or binocular viewing condition was found to significantly effect the relative distance error, with 17% less error in the binocular conditions. Here, a significant interaction between luminance condition and viewing condition was also found, with a reduction in error between monocular and binocular viewing, but only for the reduced luminance condition.

4.4.5 Scene close

Here, the results are presented for the between-object, or scene, luminance manipulation at the close distance. As before, the results were compared with Equation 4.5. Results are shown in Table 4-3.

Table 4-3: LME results for predicting relative error in distance estimations in the scene close group.

Predictor	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
Condition (c)	0.080	0.070	68	.259	-0.060	0.221
Monocular or binocular (e)	0.126	0.086	68	.148	-0.046	0.298
Interaction ($c * e$)	0.101	0.058	68	.086	-0.015	0.217

Like the object close results, no significant effect of luminance condition was observed, or an interaction with viewing condition. Here though, there was no significant effect of monocular or binocular viewing for the between-object scene luminance manipulation.

4.4.6 Scene far

Finally, the effects of luminance manipulation for the between-objects scene condition at the far distance are presented. Once again, these were compared with Equation 4.5. Results are shown in Table 4-4.

Table 4-4: LME results for predicting relative error in distance estimations in the scene far group.

Predictor	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
Condition (c)	-0.021	0.024	68	.383	-0.070	0.027
Monocular or binocular (e)	-0.014	0.034	68	.669	-0.082	0.053
Interaction ($c * e$)	0.071	0.026	68	**.007****	0.020	0.123

As with all other groups, no significant effect of reduced or enhanced luminance condition was found. There was also no significant effect of monocular or binocular viewing here. However, a significant interaction was observed between the luminance and viewing conditions, with enhanced luminance resulting in increased error, but only under binocular viewing conditions.

4.5 Discussion

4.5.1 Findings

The present work was designed as a replication study of Hibbard et al. (2023), with an extension of a viewing distance condition, in order to test the theory that the observed effect of monocular or binocular viewing condition on the luminance manipulation effect may be due to the increased weighting of binocular cues at the closer testing distance from that work.

To start, the distances set by observers were found to be highly accurate overall, with a slope value close to 1 which would indicate perfect performance. This shows that observers were making sensible judgements during the task, and were able to recover the relative depth in the rendered scene in order to make accurate distance estimations.

In general, less error in distance estimation was observed for the further distance. This is an unexpected result, as from the findings in previous chapters, increasing errors would be expected at a further viewing distance. This is most evident in the scene condition, in which smaller distances are reported at the further viewing distance. One possible interpretation of this would be that it is an example of imperfect constancy, in that overall the objects and distances between them appear smaller at the further distance.

Additionally, the present work found less error in distance judgements for local within-object judgements than for between-object scene judgements, which is the opposite effect observed in Hibbard et al. (2023). One potential reason for this finding in the present work is the location of points. Several participants voiced being unsure which object a point close to the occluding contour or edge of an object was

placed on, especially at the closer distance. This could account for the observed increase in errors in the scene condition. When making the within-objects judgements, the participants knew the points would be on the same object. When making the between-objects judgements, the points could be on any two objects in the scene. Mistakenly believing that a point placed on the edge of an object was instead on the object behind, could result in large distance estimation errors, depending on how far apart the two objects were in 3D space.

Considering the four groups for analysis, an effect of monocular and binocular viewing was found for both the object close and object far conditions. Were results to follow the predictions based on the previous work, the effect of monocular and binocular viewing would have only been found at the closer condition. This suggests that binocular cues are still rated highly in the depth estimate at a far distance, despite binocular disparity having reduced reliability with increasing viewing distance (Harris, 2004; Keefe, Hibbard, & Watt, 2011).

However, this experiment failed to replicate the luminance manipulation findings of Hibbard et al. (2023). In their work, enhancing the luminance of pixels was found to significantly increase distance errors. Here, no main effect of luminance manipulation was observed in any condition. Luminance condition was found to interact with monocular or binocular viewing in both further distance conditions, however. This may suggest that the luminance effect in the present work was ineffective at manipulating distance perception at the closer distance due to heavily weighted binocular cues, which would not report a change in depth of objects or distance between them, meaning any small effect of luminance gain reported by pictorial cues would not weight the average such to reduce or increase depth with

reduced or increased luminance. However, as binocular cues are weighted less with increasing viewing distance, an interaction between luminance manipulation and viewing conditions may provide some evidence of the 'Dark is Deep' rule in action, where binocular cues are less reliable and therefore less able to weight the average such as to align them more with veridical distance.

4.5.2 Future work

This work used 50cm as a closer viewing distance, and 100cm as a further viewing distance. Distances beyond this point were not feasible with the setup of equipment in the laboratory. However, it has been shown that binocular disparity is a highly reliable cue *up to* one metre. As such, it is possible that binocular cues were still weighted so as to reduce depth in the overall estimate at this distance. One way to explore this further would be to extend the viewing distance beyond that tested in this work, such as including a 150cm viewing distance, or beyond.

One issue with how the data were gathered that may have had an unforeseen effect on the results is that the blocks were randomised between participants such that the four variables were randomised, and within these four blocks were completed in a row before moving on to a new variable. This was chosen as a way to simplify the design for data collection. However, it is possible that this could have caused unforeseen confounding variables which affected the data. If this work were to be repeated, this could be addressed by randomising across all blocks, to see if a similar effect occurs, which might point to something else interesting happening.

This work required participants to make estimates of the depth between objects in the scene, which Koenderink (2012) argues is somewhat of an arbitrary

measure given the volatility of determining depth. Instead, he advocates for methods that allow the relation of depth across the object to be measured, by way of spatial attitude of a surface. As such the following chapters will make use of methods that allow these judgements of surface shape to be measured.

Koenderink, van Doorn and Kappers (1995) argue that there is a difference between monocular viewing, and viewing stimuli through one eye when both are open. They put forward that even though the optical data are the same, the way that the visual system processes the information may not be, due to it having the knowledge that two eyes are open and thus should be receiving information through both. In this experiment, monocular viewing was achieved by way of using the VPixx goggles, and only streaming data to one lens. For some, it was more comfortable for them to shut the eye not being fed information. However, it cannot be ruled out that some participants had both eyes open during the monocular viewing, which may have affected the way in which their visual systems interpreted the information. Should further work be done using these monocular conditions, covering the eye to ensure true monocular viewing would remove this possibility.

5 Size and sampling scale both affect the texture and shape of objects captured with a gauge figure task

5.1 Abstract

This work explores the importance of measurement scale when capturing the shape of natural objects with a gauge figure task, and looks at the impact of scale and surface texture on judgements of depth. Participants in Experiment 1 were asked to capture the shape of 3D rendered images of natural objects using three different size of gauge figures. This was completed for a high surface texture object (a sweetcorn cob) and a low texture object (courgette), and also as part of a cluttered scene providing full context, or as if through a letterbox showing just a patch of the object of interest. It was predicted that the size of the gauge would affect the shape of the mesh captured, and that this effect would interact with object texture. Specifically, it was expected that more surface detail would be captured with the smallest gauge for the high texture object (the sweetcorn), leading to more surface roughness in the reconstructed 3D mesh, but that gauge size would not affect surface roughness for the low texture object. In Experiment 1, where we varied just the gauge size, rather than the sample locations, neither global shape from the depth range or local surface texture from a measure of roughness varied significantly with different gauge sizes. When asked which size of gauge they felt best captured the shape of the objects, participants consistently chose a smaller gauge figure for the high texture object than the lower texture one as predicted, showing an overall preference for the small gauge for this task, compared to a medium one for capturing the shape of the courgette. More overall depth was perceived when the full global image context was

available than in the local ‘letterbox’ condition for the sweetcorn, due to the presence of additional depth cues by which to scale the surface texture information, which better simulates natural viewing conditions. Experiment 2 followed up on these findings by adjusting the distance between gauge settings as well as the size. Results show a highly textured reconstructed surface with the smallest gauge, with less definition captured with increasing gauge size and distance. A measure of surface roughness found a significant negative effect of gauge scaling on surface roughness, indicating that increasingly less surface detail was captured with increasing gauge size and sampling distance. Combined with the lack of effect of gauge size on either global shape or local surface texture in Experiment 1, we conclude that both gauge size and sample point distance need to be scaled accordingly for sufficient global shape and local surface texture capture in a gauge figure task. This work highlights the importance of considering the scale of judgements when designing a gauge figure task. Consideration of the scale at which judgements are required, and the surface texture of the object, all need to influence the size of gauge selected for the task, as well as the distance between sampling locations. Additionally, the work reinforces the use of complex naturalistic stimuli in cluttered scenes to simulate natural viewing for depth perception tasks.

5.2 Introduction

Many previous studies have utilised a surface normal gauge figure task as an intuitive way of probing local surface attitude in order to capture global pictorial relief in depth perception experiments. However, there is little in the literature on the rationale behind the spatial scaling chosen for these methods. The present work

aimed to explore the importance of scale in a gauge figure task, with thought to how scale plays into consideration of surface texture, as well as for global and local judgements of shape.

The previous chapter utilised naturalistic stimuli, but the experimental methods only captured the distance of a line between points on the stimuli. Classical geometry defines a point as having a dimension of zero, and a line as having a dimension of one (Wahl, 1991). However, in the case of visual perception, Gibson (1950) states that we do not perceive points, but surfaces. As such, this work expands on the previous chapter by adapting the methods to capture the shape of a local surface in order to probe the shape of a global area, which has a dimension of two, and combine these into a 3D mesh of the percept to recreate an object with volume – a dimension of three, to elevate findings from simplified stimuli to more natural real-world viewing. This was achieved with a surface normal gauge figure study, which tasked observers to fit a circle 'pushpin' style gauge figure to perceptually 'fit' the surface of an object in pictorial space (Koenderink, van Doorn, & Kappers, 1992). To begin with, pictorial space will be defined.

In the art world, pictorial space is defined as the illusionary space within a two dimensional picture, where the artist, through the use of techniques such as perspective and varying dark and light tones, has attempted to convince the viewer of the presence of depth (Squire & Platt, 2017). Koenderink (2012) describes this ability to 'look into a picture' as being separate entirely from visual space and is in fact something mental rather within than a physical plane.

In defining pictorial space, Koenderink (2012) presents the work of artist Kazuki Takamatsu, who creates 'depth map' paintings by layering differing

monochromatic tones, so that the closest points are lighter and the further points are darker, as per the ‘Dark is Deep’ rule outlined in Chapter 4. In doing so, Koenderink argues that the artist has created (what renown sculptor Hildebrand, cited by Koenderink) refers to as a ‘depth flow’, allowing the viewer to sense the depth created as a smooth flowing topological landscape. As outlined in Chapter 2: Methods, a surface normal gauge figure task can be used to probe local surface attitude of these ‘hills and dales’, by rotating a gauge figure so that the circle element appears painted flat on the surface of objects in pictorial space, with the pin normal to the surface (Koenderink, van Doorn, & Kappers, 1995).

The orientation of the gauge figure when set to perceptually fit the surface provides a measure of slant and tilt. This chapter employs methods outlined by Nefs (2008), whereby the slant and tilt settings can be converted into depth gradients which provide information about the orientation and magnitude of surfaces. These depth gradients can be modelled into a mesh of the entire perceived object, which will form the basis of analysis for this study. More information on this is in Chapter 2: Methods.

While the gauge figure task is a popular one for depth perception research, Koenderink, van Doorn and Kappers (1996) have put forward the idea that the results of such a task are operationally defined, in that measurements are defined by the way in which you measure them. Expanding on this, Koenderink (1998) argues that the pictorial object exists not in regular geometric space, but in the mind of the observer, and therefore cannot be measured using typical geometric tools, such as a compass. However, he points out that the fit of a measure, in this case a gauge figure, can be judged in pictorial space just as easily as it can for actual space, by

using the example of how well a person's body appears to fit on a seat when viewing a picture of them sat on a chair. For instance, in earlier studies that utilise the gauge figure (Koenderink, van Doorn, & Kappers, 1992; 1995; 1996), pictures of stimuli were presented with a gauge figure superimposed on top of the image. This allowed capture of the change in perception of the object between experimental conditions, such as by varying the position of the light source in the image while holding all other variables constant, creating an 'operating point' of all other cues around which the cue of interest can be studied.

As tasks are operationally defined, this places increased importance on the task chosen, as well as its parameters. Indeed, the same task may even garner differing results based on changes to the chosen limits, such as scaling. The next section will outline the geometric theory behind this, and explore the reasoning behind why this is an important area of questioning for this research.

5.2.1 Importance of scale

While much work has been done using the gauge figure methodology (Koenderink, van Doorn, & Kappers, 1992; 1995; 1996; Nefs, 2008), the sizing of the gauge figures remains somewhat arbitrary given that the surfaces of stimuli used in the experiments were predominantly smooth and therefore would be unaffected by varying the size of the gauge, as sufficient local surface information would be captured with the size of gauge used in the experiment. Studies in this area often report that the gauge figures are placed at regular intervals (Nefs, 2008), without defining what a regular interval is, and this is not measured or reported. Indeed, Koenderink (2012) highlights the lack of conventional unit by which to measure depth

and discusses the complexity of this. The present work plans to explore this by investigating the importance of the size of the gauge figure in relation to the surface it is being used to capture.

Highly relevant to this work is the area of maths known as fractal geometry. Fractal geometry defines the repetitive self-similar patterns and structures often seen in the natural world - snowflakes, trees, lightning, amongst others are all made up of fractal patterns. Given that this body of work aims to explore viewing of natural objects, how we observe and measure objects with varying levels of surface complexity is an important consideration. For instance, when measuring something with a fractal structure, such as a coastline, the measured length is dependent on the equipment used to measure. This is known as the coastline paradox, where the measured length of the coast increases as the size of the ruler measuring it decreases, as more of the intricate detail is captured.

Applying this to capturing the shape of natural objects with a gauge figure task, the depth captured will be dependent on the scale of the gauge used to capture it. As such, the work here compared the shape of natural objects captured with differing sizes of gauge figure. We hypothesised that the overall range of depth of natural objects would differ when captured with varying sizes of gauge figures; specifically, that decreasing sizes of gauge figure would allow for more surface complexity to be captured, and therefore the local surface attitude was expected to vary across scale. To explore this idea, natural objects with varying levels of local surface complexity were tested using several sizes of gauge figure.

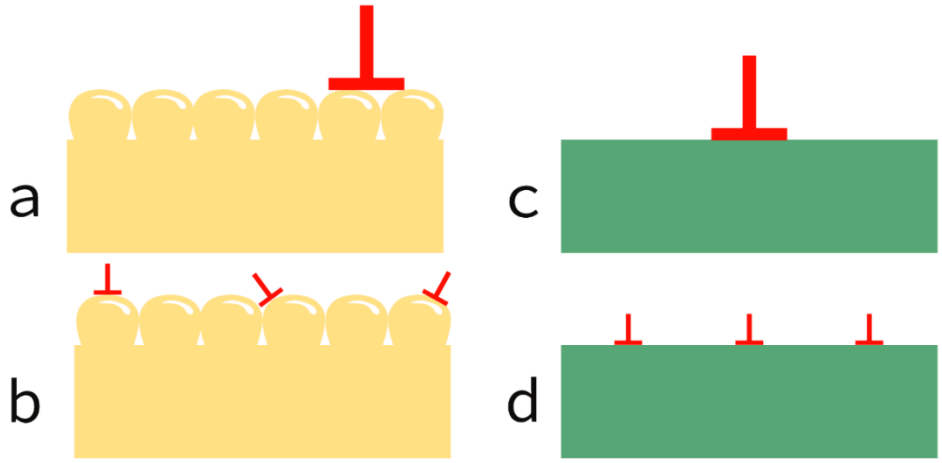


Figure 5.1: Gauge size consideration. Diagram showing how varying sizes of gauge figure would capture the shape of objects with varying levels of surface texture. Images represent sweetcorn on the left, and courgette on the right.

To visualise this concept, Figure 5.1 shows how differing gauge sizes would capture the shape of objects with varying levels of surface texture which denote fractal geometry. The images on the left represent a section of a sweetcorn cob, and the images on the right represent a section of a courgette. These objects have been chosen to compare as their underlying geometry is similar, in that they are both predominately cylindrical in overall shape. However, they have differing levels of surface texture, with the skin of courgette being largely smooth, and the kernels of the sweetcorn creating a surface with local texture complexity.

In a gauge figure experiment to capture the shape of these objects, the larger gauge in Figure 5.1*a* would be set such that it rests across the surface, and the intricate, complex surface detail of the sweetcorn kernels would not be captured. By using a smaller gauge, as per *b*, the gauge figure could be rotated so that it is normal to the surface, and more of the surface texture can be captured. The pictures in *c* and *d* show that decreasing the size of the gauge here would not affect the surface captured for the courgette.

The present work tested the hypothesis that varying the scale of the gauge figures would vary the local attitude captured for objects of differing surface complexity by sampling the surface of both the courgette and the sweetcorn using three differing sizes of gauge figures, to explore how the scale of the gauge affected the shape captured. It was predicted that the range of slant and tilt settings captured by the gauge figure in the sweetcorn conditions would increase with decreasing gauge size. It was also predicted that no significant difference would be observed in the courgette conditions between the varying gauge sizes. Relating back to Koenderink's (2012) comment that tasks are operationally defined, results were expected to show that global pictorial relief is not fixed but instead is dependent on the scale of the measurement used to capture it.

In addition to measuring the change in depth reported across the conditions, participants were also asked which gauge figure they felt best captured the shape of both the sweetcorn and the courgette respectively. By doing so, comparisons could be made between popularity of gauge size between differing conditions with participants, who were expected to make a fair assessment of more pragmatic considerations such as time and effort.

Extending the example of measuring the coastline with ever decreasing sizes of rulers, a balance must be struck between accuracy of measurement, and more pragmatic considerations of the time and effort required to undertake such a task, weighed against the usefulness of the level of definition required. For instance, using a scale where measurements are made using a ruler smaller than the length of a step may gather more accurate information, but would not be useful in practical

terms when walking the length of it. This work considers the results in the context of these considerations.

This chapter is an important milestone in this work as it represents a move from laboratory-based studies to online data collection. In this shift, the technology used to present stimuli in the first two experimental chapters is replaced instead with participants' own computer monitors. However, this move is not expected to impact perception of stimuli on the screen. Koenderink (2012) argues that when looking into a painting, neither the frame nor the surface of the canvas belong to pictorial space. Viewing a painting in pictorial space does not require optical assistance, as the pictorial cues, or monocular cues, used by the artist are often stronger than binocular disparity, or physiological cues such as depth of field or monocular parallax (Koenderink, 2012). In the case of these experiments, when viewing the natural objects in pictorial space, participants were not expected to set the gauge figure as if flat to the surface of the computer monitor, here replacing the canvas in the example of viewing a painting. Instead, it should be expected that the settings would be made as if flat to the surface of the objects in pictorial space if the participants view them as belonging to such.

5.2.2 Global versus local viewing

When considering the scale of capture in these experiments, further consideration is needed on the scale and context for the overall composition of the image presented. As highlighted above, Koenderink (2012) proposed that surfaces are viewed as a continuous relief map as part of 'depth flow'. This points to the importance of making judgements as part of a coherent global relief percept, rather

than single points in isolation. However, gauge figure tasks are used to survey the local attitude of pictorial relief. The present work will examine the importance of this idea when making judgements using the gauge figure task, which by its nature forces the participant to make a judgement of local surface orientation at a single point.

Koenderink, van Doorn and Kappers (1995) asked participants to set gauge figures on the surface of a sphere. They concluded that it was not possible to say which cues were dominant in the scene, but speculated that the circular contour of the object provided a lot of information, posturing that occluding the contour in part or entirely would dramatically change the results of the local relief. To explore this, the present work presented participants with both global context, in the form of a full cluttered scene of fruit and vegetables, as well as a local context condition, achieved by displaying only a patch of the full image, with the full contour of the objects occluded.

Here, the prediction was that the additional information provided by additional depth cues in the global context condition such as occlusion and scale of other objects, as well as additional information about the source of light in the scene by which to interpret shading information, would lead to greater variation overall in the local surface attitude reported.

A final consideration for this chapter is whether judgements are affected by the order in which global and local context conditions are presented. The Bayesian model to visual perception purports that rather than making entirely new judgements, the visual system ‘speeds up’ the process by incorporating prior knowledge to determine estimations of scene properties from sensory input (Battaglia, et al.,

2010). In the context of this work, this suggests that viewing the full global context before the local context condition would provide some additional information on which to base judgements, as participants may recall the original full image when presented with the local context only.

Additionally, the shape of objects is not arbitrary, objects within classes are known to have certain regularities constrained by natural laws (Knill, Kersten, & Yuille, 2008). This Bayesian knowledge combined may affect reported depth of objects when previously viewed, or known to fit certain physical properties. By removing these two aspects, i.e. by presenting a patch of an object rather than the entire model, this will force judgements to be made using available pictorial cues only, such as shading and surface colour.

Indeed, prior experience of stimuli has been found to enhance performance in a depth perception task; Hartle and Wilcox (2022) presented observers with stimuli on which to make depth judgements in two virtual environments, using a stereoscope and a Head Mounted Display (HMD) for Virtual Reality (VR), as well a physical viewing condition. Post hoc analysis revealed that performance in the virtual environments was significantly improved having viewed the stimuli in the physical viewing condition. These results in the context of this work led us to hypothesise that prior exposure to the stimuli could improve performance in the local ‘letterbox’ context condition having previously completed the global context condition of the full cluttered scene.

It was predicted that participants who received the global context first would perform better and make more accurate depth judgements than observers who completed the local condition first. This is because the lack of context in the local

condition was predicted to result in a flattened percept. We expected more overall depth reported in the image in the local conditions for those who viewed the global condition first than those who viewed the local condition first, as these observers would have had prior information of the full stimuli on which to decode the visual signal and make an estimation of depth, therefore reducing the flattening expected for the local-first observers.

5.2.3 Measuring the change in captured shape

Here, an overview of the three methods to measure the change in shape between reconstructed meshes are presented. Further information and detail on these methods is covered in Chapter 2: Methods.

Firstly, the present work examined the effect of gauge size on the global shape captured in the gauge figure task. As discussed above, methods employed by Nefs (2008) were used here to recreate the perceived surface from observer settings. These meshes can be compared using a process of affine transformation to give a measure of the effect of gauge manipulation on the depth and shape of the settings made. This affine transformation provided us with a measure of global shape change between local and global conditions. Secondly, we measured the range in depth across reconstructed meshes, between the closest and furthest pixels. This provided a broad measure of the depth of perceived objects on a global scale.

Finally, the experiments here explored the level of detail captured for surface texture using a measure of roughness. This provided us with a measure of the roughness of the surface of the reconstructed mesh to explore the difference in

levels of texture captured between the varying gauge size and texture conditions. To achieve this, we made use of a mesh simplification scheme presented by Wu, Hu, Tai and Sun (2001). Their original work used this measure as a way to compare 3D meshes that have been simplified using a variety of processes which all aim to reduce 3D mesh file size by adjusting triangles or vertices within a mesh. However, during this process sometimes important shape information can be lost, such as mesh decimation whereby vertices are entirely removed, resulting in a change in mesh topography.

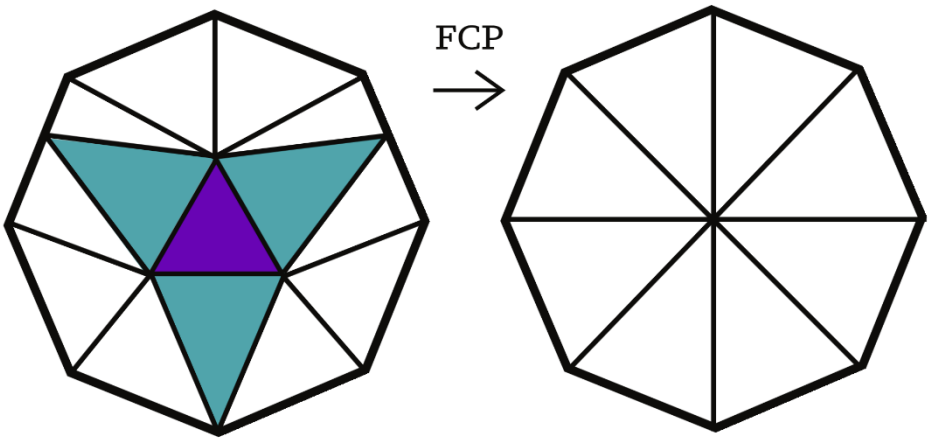


Figure 5.2: Face Constriction Process. Showing the Face Constriction Process (FCP). The purple triangle shows the face identified for removal. The turquoise triangles are the adjacent faces which are degenerated and removed. The original face is therefore converted into a new vertex.

Figure 5.2 shows the Face Constriction Process (FCP). Here, the purple triangle denotes a face targeted for removal. The neighbouring faces, shown with the turquoise triangles, are degenerated and deleted. This converts the original face into a new vertex, as seen on the right. In order to be a suitable candidate for removal, the FCP must not invalidate certain restrictions, namely that the face must not have more than three adjacent faces, meaning it is non-manifold, must not lie on the

boundary of the mesh, must not contain a cycle of three mutually-adjacent triangles, and must not result in triangles folding upon one another.

The FCP process creates a weighting for each of the surfaces identified for removal, and uses this weighting to report the data loss of a model when faces are removed. This is done by calculating the dihedral angles between a given triangle, and those connected to its vertices. Their intended use of this method is to compare meshes before and after removal of a face to see if the resulting mesh retains sufficient feature information, with a high score indicating that removal of the face would result in increased loss of surface information. They then use that figure to assess the cost of removal. Here, we have used this approach as a measure of surface roughness, by comparing this number across faces to assess the measure of change across the surface. For instance, under our use, a smooth plane would have a lower score than a textured plane given the lack of difference between dihedral angles for triangles and those surrounding them, indicating a smoother surface. This measure of roughness was calculated using the following equation:

Equation 5.1

$$GT = \frac{G_1 \cdot VAR_1 + G_2 \cdot VAR_2 + G_3 \cdot VAR_3}{VAR_1 + VAR_2 + VAR_3}$$

Here, GT is the weighted sum of average dihedral angle, with G_1 and VAR_1 showing the weighting and variance of the angles of vertex 1 respectively, and so on for vertex 2 and 3. The calculated ‘surface roughness’ score gave us a measure we could use to compare reconstructed meshes and assess the average texture captured across conditions.

To summarise, this chapter covers two experiments that explore the importance of gauge size, object texture and viewing conditions on the perception of depth and surface shape and texture. It was predicted that increasing depth would be captured with decreasing gauge size, and that this effect would differ depending on surface texture of the objects captured, as well as by viewing condition.

5.3 Methods

5.3.1 Participants

5.3.1.1 Experiment 1

16 participants aged between 19 and 63 took part in this 20 minute online study. Seven (44%) identified as female, eight (50%) as male and one (6%) as a trans man. Two were psychology students at the University of Essex who were recruited online through the University of Essex's SONA system and received 0.5 course credits for their participation. 12 participants were recruited through word of mouth and via social media and received £5 for their time. In total 14 participants were naïve, and two were researchers. Participants had to be over 18 years of age. All participants were required to have normal or corrected-to-normal vision (glasses or lenses) and were asked to wear their glasses or lenses if they typically would do to use a computer.

Although participants received thorough written instructions with diagrams, as well as practice trials, due to the online nature of the experiment five incomplete data sets were discounted from analysis with an exclusion criterion of more than 40% trials not completed, as shown by the slant and tilt settings remaining unchanged over these trials. One participant reported back that despite the instructions and

practice trials they had not understood the experiment sufficiently to complete it. The other four participants did not provide feedback. This will be discussed further in the discussion section. In addition, one participant had not completed the initial screen calibration task (as described below) correctly, so their data was discounted due to potential stimuli scaling issue effects.

5.3.1.2 Experiment 2

Three participants aged between 32 and 52 completed this 2.5-to-3-hour experiment, two who identified as male and one who identified as female. Two of the participants were researchers and the other was a naïve participant who volunteered their time. All had experience of previous experiments of pictorial depth. All had either normal or corrected-to-normal vision and wore glasses for the experiment if they required them.

5.3.2 Apparatus and materials

The two main 3D models used in both experiments, the courgette and the sweetcorn cob, were created by scanning real vegetables with an Artec Spider™, a light reconstruction handheld scanner (Artec 3D, Luxembourg), which captured the shape, colour and texture of the objects at a 3D resolution of up to 0.1mm, a point accuracy of up to 0.05mm, a texture resolution of 1.3 megapixels and depth of colour information of 24 bits per pixel. Objects were scanned around a full 360 degrees to create a scan of the complete object at a distance of around 25cm. Objects placed around the two main objects in the cluttered scene was scanned using the NextEngine 3D Scanner and Multidrive turntable, with 360 degrees of rotation, two

tilt settings of ± 20 degrees and a resolution of 62 points per mm^2. The scans were edited into meshes inside the NextEngine Scan Studio HD software and fused with a tolerance of 0.064mm for a water-tight model. Further information on the scanners can be found in Chapter 2: Methods.

The 3D models were rendered in using OpenGL in MATLAB with the Psychophysics Toolbox plugin (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Scenes were created using a cluttered collection of these 3D rendered objects. Scenes were rendered as if presented to the 'cyclopean eye' (Stidwell & Fletcher, 2011), where left and right eye positions were identical as if placed centrally. For more see Chapter 2: Methods. OpenGL lighting components for ambient and diffuse lighting with a magnitude of (0.7 0.7 0.7) were used, with a spotlight placed centrally and directed at the middle of the scene for the diffuse lighting element.

JavaScript code to render the gauge figures on the scene was adapted from work by Wijntjes and van Zuijlen (2004). The experiment was created using the P5.js library (https://p5js.org/) in Qualtrics (https://qualtrics.com) and completed online using the participants' own desktop or laptop computer. Responses were recorded using the participants' desktop or laptop computer mouse, where participants would rotate the gauge figure by moving the mouse, clicking the left mouse button to confirm their judgement and move onto the next trial. The apparatus and creation of materials for both experiments were identical.

5.3.3 Stimuli

5.3.3.1 Experiment 1

Rendered cluttered scenes of two main 3D objects created using scanned fruits and vegetables were presented to participants. One object, the ear of sweetcorn, had a high level of surface complexity in the form of its kernels, and the other object, the courgette, had a low level of surface complexity, being mostly smooth. However, both objects were similar in their overall geometric shape, in that they are both predominantly cylindrical, so as to be able to make a comparison of overall shape. These were presented either as an image of a full 3D model within a cluttered scene (Figure 5.3), or as a patch as if viewed through a 'letterbox' (Figure 5.5), as in the pictures below. All four images were 512 by 512 pixels square.

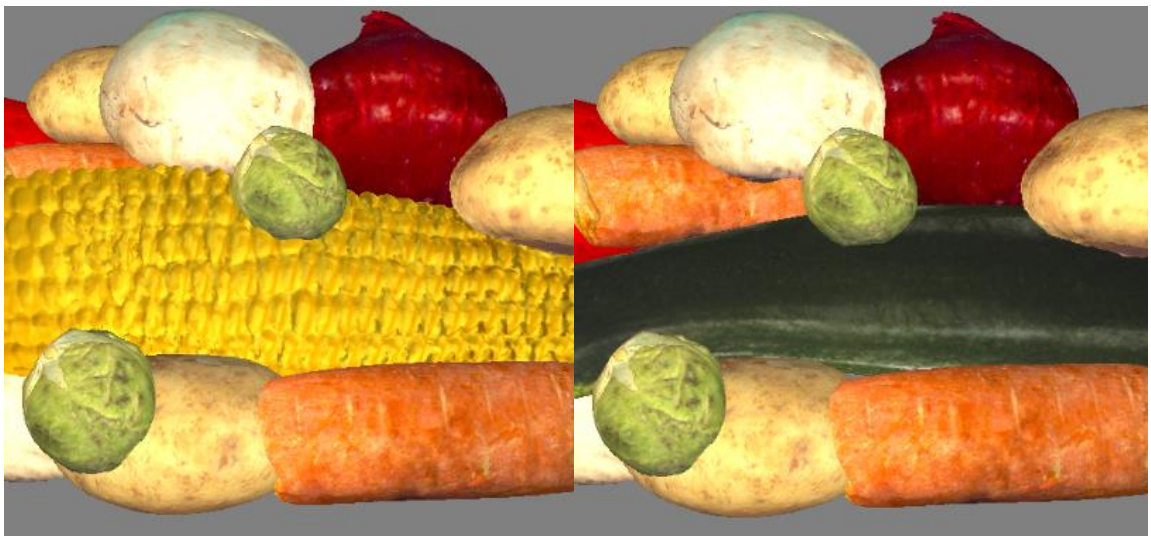


Figure 5.3: Stimuli. Images of stimuli presented for the high texture (sweetcorn, left) and low texture (courgette, right) objects within a cluttered scene.

For consistency, the same points were sampled on both the sweetcorn and the courgette, shown in the Figure 5.4 below, which shows a triangulation of the visual field. This triangulation was not visible in the experiment itself, but the gauge

figures could only appear at the centre of the triangles. For more on the ‘barycentres’ method, see Chapter 2: Methods. Gauge figures were superimposed onto the objects with three different base size diameters, and different colours used to help distinguish them from one another to help participants recall which they had felt best captured the shape of the stimuli at the end of the experiment: a 5px blue gauge in the small condition, a 10px pink gauge for the medium condition and 15px orange gauge for the large.

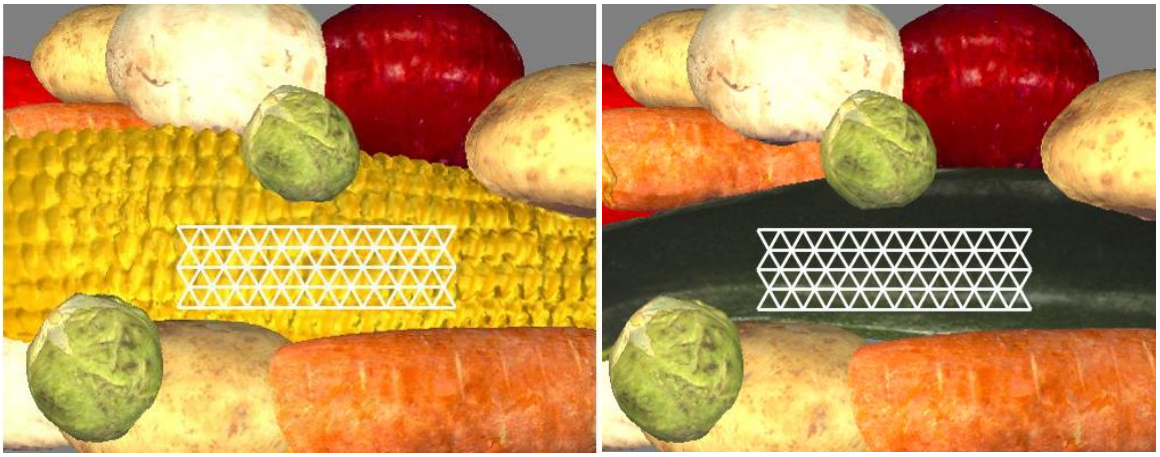


Figure 5.4: Gauge locations. Images showing possible gauge figure locations sampled on the sweetcorn and courgette in the cluttered scenes. Gauges were sampled in the middle of the triangular faces at the barycentre.

The ‘letterbox’ was identical in both the sweetcorn and courgette conditions. It was placed centrally and measured 95 pixels high by 266 pixels wide and was framed by a 5-pixel wide brown border to aid contrast to the black background and reinforce the idea that it is a small patch of a larger scene being seen through a viewfinder, as seen in Figure 5.5. The points sampled in the letterbox condition were the same as those sampled in the cluttered scene to allow for direct comparison.

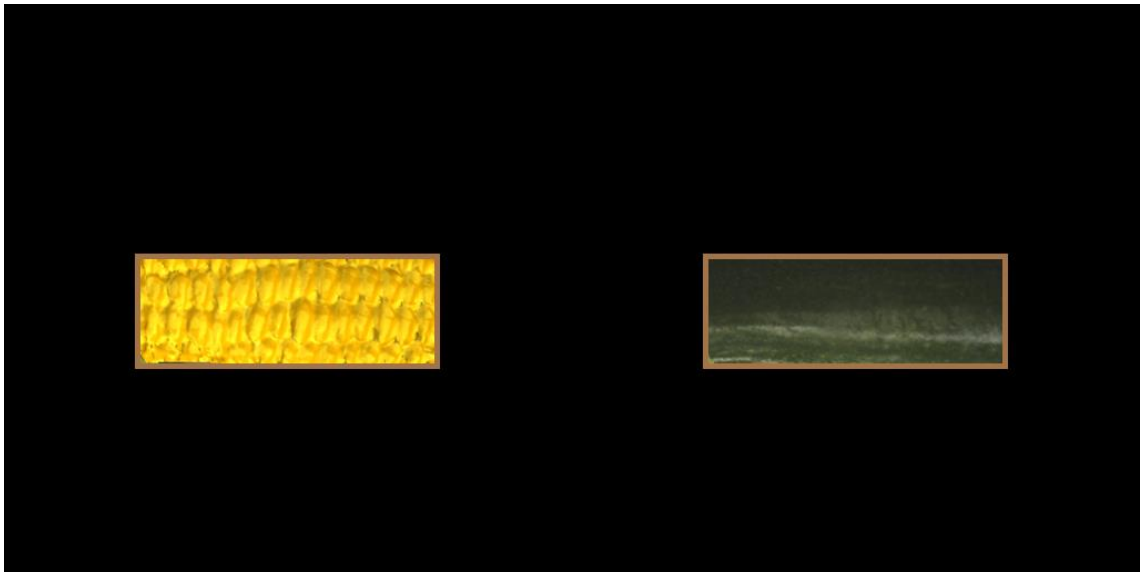


Figure 5.5: Letterbox conditions. Images showing the 'letterbox' conditions for the sweetcorn (left) and the courgette (right).

5.3.3.2 Experiment 2

The stimulus for this experiment was a rendered simple scene of a sweetcorn, the same used in the cluttered scene in Experiment 1, this time presented in isolation on a grey background, shown in Figure 5.6. The image was 1000 pixels high by 1800 pixels wide. The same picture was used in all three scale conditions.

Figure 5.6: Sweetcorn. Image of the 3D model of a sweetcorn cob scanned from a real object, presented in Experiment 2 for all conditions.

The gauge figures superimposed onto the sweetcorn in this experiment were white. All possible measurements across the sweetcorn were taken to create full meshes at each level of scale condition. In the fine detail condition, the sampled points were spaced 20 pixels apart. The gauge figures had a base diameter of 7 pixels. This was achieved by triangulating a mesh across the surface of the entire object. The image below shows the vertices (white crosses) and barycentres (red circles) which represent the locations of the gauge figures.

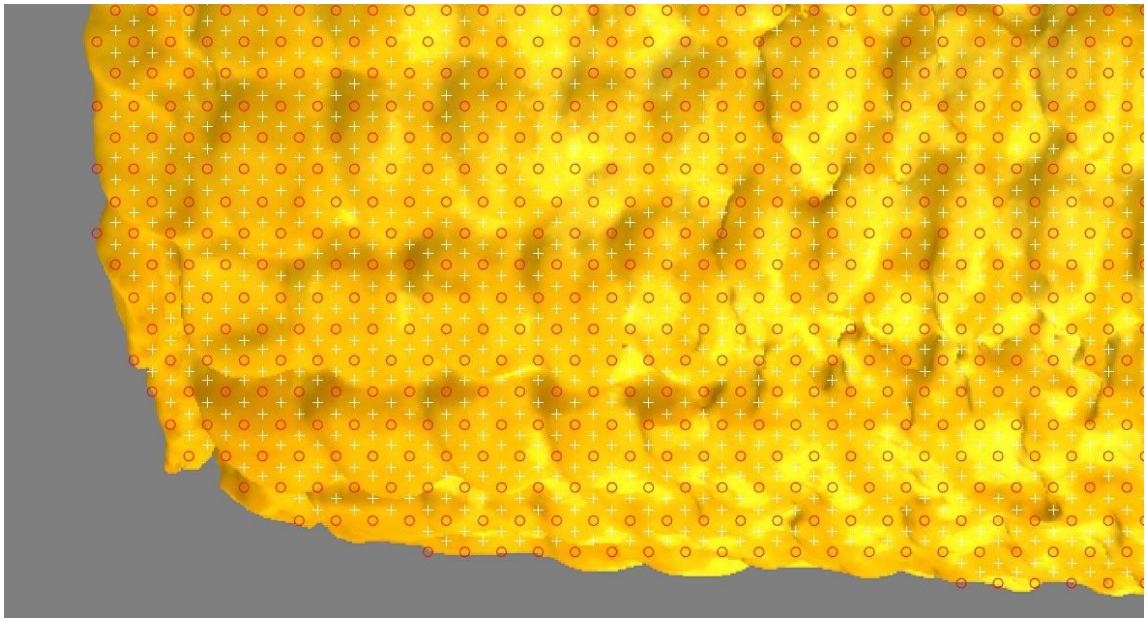


Figure 5.7: Sample locations. Image showing vertices (white crosses) and barycentres (red circles) for the fine detail condition.

In the medium detail condition the points were spaced 40 pixels apart and the gauge figure had a base diameter of 14 pixels. In the coarse detail condition the points were spaced 60 pixels apart. The gauge figure had a base diameter of 21 pixels.

5.3.4 Design

5.3.4.1 Experiment 1

In this chapter, no direct manipulation of depth was applied to the stimuli itself, and as such the same four images of rendered scenes were used in the relevant conditions for the full global context sweetcorn cluttered scene, full global context courgette cluttered scene, and a letterbox condition for each of the two main objects. The manipulation was in the form of the size of the gauge with three levels: small, medium and large. These were delineated to participants by both their size and their colour. In addition, the objects were either high textured, the sweetcorn, or low textured, the courgette. This gives six blocks. We also manipulated the viewing context with two levels: stimuli were either presented as a full global context cluttered scene, or as a local patch as if viewed through a letterbox, which was a further two blocks, both of which used the medium size gauge figure. Therefore, there were eight conditions in total. All participants completed all eight conditions in a random order, which was recorded within Qualtrics. The final manipulation was whether observers saw the global viewing or the local viewing condition first. This was randomly assigned within Qualtrics. Due to some observers not completing their dataset (outlined above), 11 participants saw the global viewing condition first, and five saw the local condition first.

5.3.4.2 Experiment 2

This experiment consisted of three conditions: a fine detail scale condition, a medium scale condition and a coarse scale condition, with the gauge figures and the distance between judgements increasing to two and three times the fine detail

measurements respectively. All three participants completed all three conditions in a randomised order.

5.3.5 Procedure

5.3.5.1 Experiment 1

Central binocular viewing was assumed. Participants completed an online consent form and began by calibrating the size of their screen by holding up a regular size bank or club card and drawing a box that size on the screen using the mouse, in order to correctly scale the presented stimuli. Next, they were provided with written instructions and images to explain the task.

The experiment began with a few practice trials to ensure participants understood the task. Participants were presented with an image of a tomato, and asked to capture the shape using gauge figures. A gauge figure would appear in a randomised location defined by the grid of seven possible locations, and participants were asked to rotate the gauge figure until it appeared to be painted flat to the surface of the object (Koenderink, van Doorn, & Kappers, 1995), with the stick portion of the gauge figure pointing outwards, perpendicular to the surface. Once participants were happy with the placement of the gauge figure, they clicked the left mouse button to place it and the next trial began with a new gauge figure in a new location. When participants were happy that they understood the task, they began completing the experimental trials.

The main experiment consisted of 40 trials per block. When all trials had been completed, the block ended, and participants were offered a break and prompted to begin a new block when ready. There were the two different objects (sweetcorn and

courgette), with three varying sizes of gauge figure (small, medium, large), presented as two different viewing conditions (six cluttered scene conditions and two 'letterbox' conditions), giving a total of eight blocks. Participants were randomly assigned by Qualtrics to view either the global or local context conditions first, and blocks were also randomised between these. At the end of the experiment, participants were asked to select which of the three sizes of gauge figure they felt best captured the shape of the object for both the courgette and the sweetcorn. They were thanked for their time and invited to contact the researchers for more information about the study if they wanted.

5.3.5.2 Experiment 2

The procedure was much the same as Experiment 1, except there were three conditions, capturing a fine, medium and coarse scale. These were split into 22 blocks of around 300 trials per block, with a total of 6,435 judgements, allowing for breaks in between blocks. The experiment lasted between 2.5 and 3 hours in total.

5.4 Results

5.4.1 Data treatment

Raw scores were produced in Qualtrics of phi (slant) and theta (tilt) settings made by observers related to the x and y positions of the points sampled with gauge figures. Stimuli and gauge figure location points had been scaled during the screen calibration task, so these were first standardised to allow comparison between participants by reversing the scaling factor. As the points were sampled in a random

order, these were then reordered for analysis. These were then converted into depth gradients using the methods outlined by Nefs (2008). Next, a mesh was fit to these gradients using a least squares deviation approach (Koenderink, van Doorn, & Kappers, 1992). Reconstructed meshes for each stimuli surface for each participant were first visually verified to ensure participants had understood the task and made sensible judgements, an example from one participant can be seen in Figure 5.8. A maximum angle restriction was added to the mesh reconstruction of 85 degrees to reduce sensitivity to extreme gradient values. Further details on these methods can be found in Chapter 2: Methods.

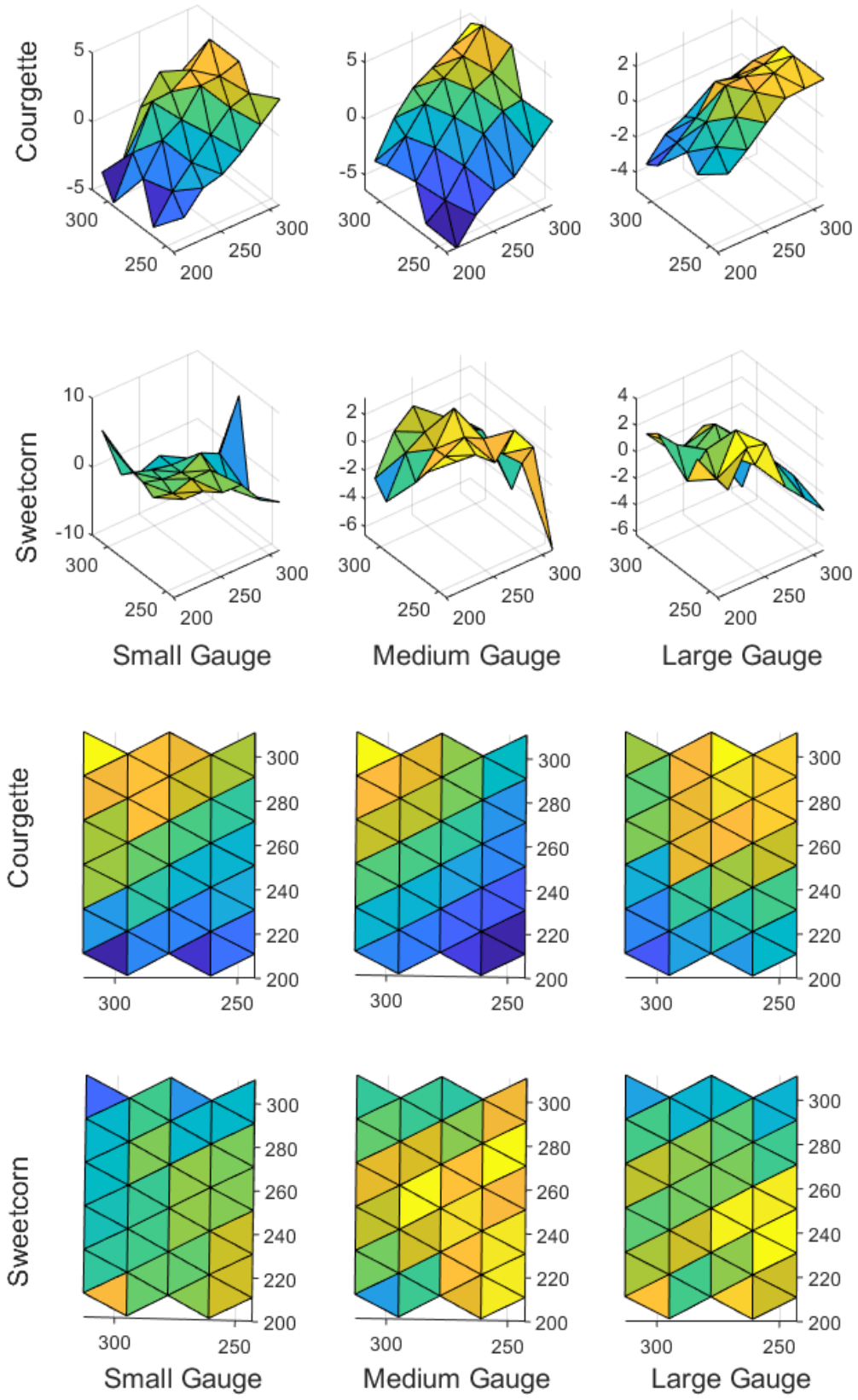


Figure 5.8: Reconstructed meshes Experiment 1. Plots showing examples of reconstructed meshes for the three gauge size conditions and both texture objects for one participant in the global conditions. Top shows meshes from the side, and bottom shows top-down view of the same data.

5.4.2 Experiment 1

5.4.2.1 Gauge size and texture

Here, the effect of differing sizes of gauge on the shape captured of the different textured objects is explored. Figure 5.8 shows an example of reconstructed meshes for the three gauge size conditions (small, medium and large) for both the sweetcorn and courgette conditions for one participant. The courgette conditions can be seen to have captured the rounded cylindrical shape of the courgette, whereas the sweetcorn conditions have far less of a cohesive global shape due to the increased surface texture.

To begin with, the depth range of each model was compared. This looked at the difference between the closest and furthest pixels in the reconstructed mesh as a measure of the global depth captured. Means (SE) are shown in Figure 5.9.

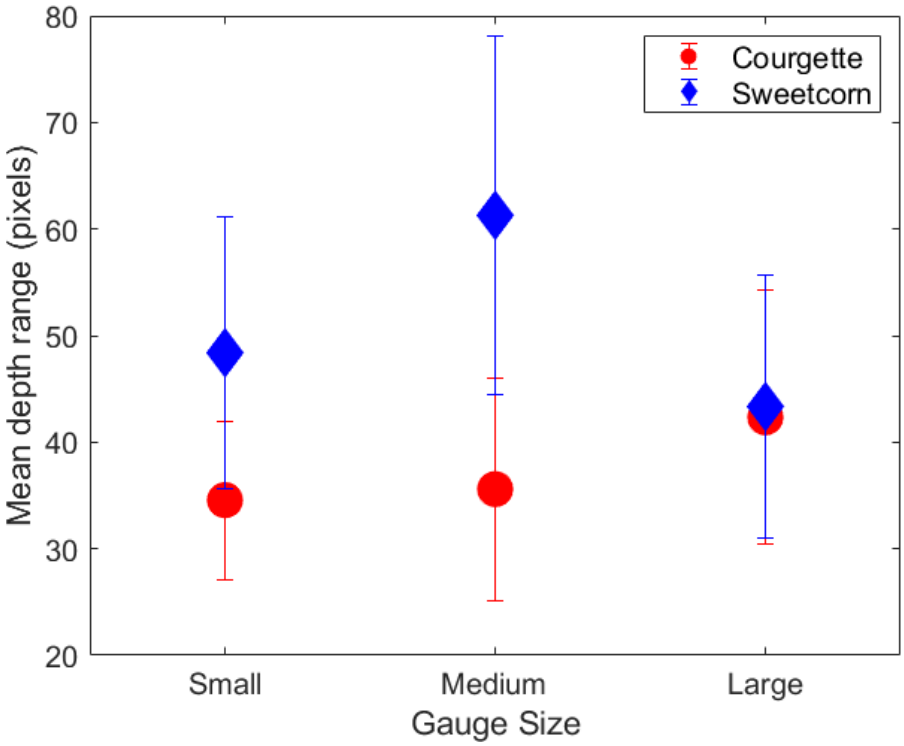


Figure 5.9: Depth range Experiment 1. Graph showing mean (SE) pixel depth range Experiment 1, calculated as the difference between the closest and furthest points in the reconstructed mesh.

A linear mixed effects model was conducted using Equation 5.2. This explored the effect of gauge size (g) and texture (t) on depth range (dr), grouped by observer (o). The model contains the maximal random slopes and intercepts, chosen as being the best fit for the data from AIC values.

Equation 5.2

$$dr \sim 1 + g * t + (1 + g * t|o)$$

Neither gauge size nor texture were found to significantly predict pixel depth range:

Table 5-1: Depth range ANOVA results Experiment 1.

ANOVA term	**F**	**DF**	***p* Value**
Main effect gauge size: g	0.535	2,90	.587
Main effect texture: t	1.68	1,90	.198
Interaction: g:t	1.32	2,90	.272

To explore the effect of gauge size on the level of texture captured, a measure of surface roughness was calculated for the reconstructed meshes which quantified the level of surface texture captured during the gauge task. This was achieved using the mesh simplification equation presented in the introduction (Wu, Hu, Tai, & Sun, 2001), which provided a measure of how much the surface changed across the object. The means (SE) for each condition are plotted in Figure 5.10:

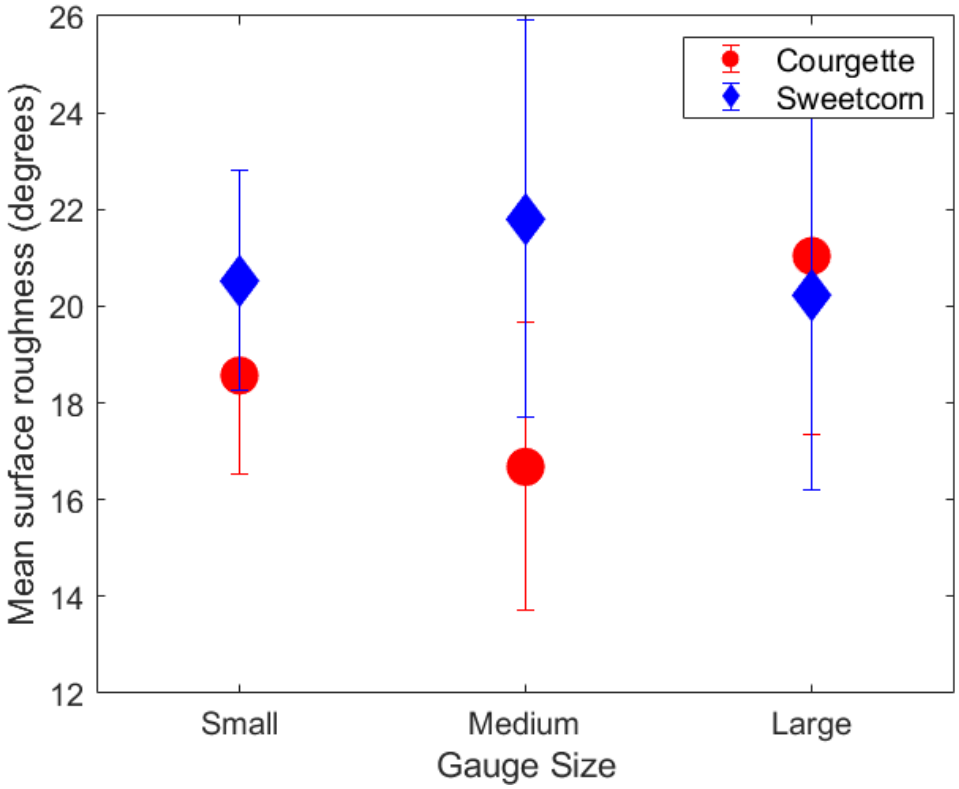


Figure 5.10: Surface roughness. Graph showing mean (SE) surface roughness in degrees for meshes in Experiment 1 for both objects at each gauge size.

At the large gauge size, the surface roughness is almost equal for the two differing texture objects. At the smaller and medium size gauge conditions, the surface roughness is higher on average for the sweetcorn than the courgette. However, these have large error bars denoting a large variance. An LME was conducted to see if surface roughness (r) was predicted by gauge size or texture:

Equation 5.3

$$r \sim 1 + g * t + (1 + g * t|o)$$

No main effects of gauge size or texture, or an interaction between them, were found for surface roughness:

Table 5-2: Surface roughness ANOVA results Experiment 1.

ANOVA term	F	DF	*p* Value
Main effect gauge size: g	0.535	2,90	.587
Main effect texture: t	1.68	1,90	.198
Interaction: g:t	1.32	2,90	.272

5.4.2.2 Observer preference

Observers were asked which of the gauge sizes they felt best captured the shape of both the courgette and the sweetcorn. A smaller gauge was chosen for the sweetcorn than for the courgette almost unanimously. Only one participant chose a larger gauge for the sweetcorn than the courgette and an explanation for this choice is covered more in the discussion. Table 5-3 shows the number of observers (percentages) who chose each gauge size for either object. The small gauge was the most preferred for capturing the shape of the sweetcorn, with observers preferring the medium size gauge best on average for the courgette.

Table 5-3: Gauge size preference. Chi-square contingency table showing number (and percentage) of observers who chose each size gauge for each object.

Object	Small	Medium	Large
Courgette	0 (0%)	9 (56%)	7 (44%)
Sweetcorn	12 (75%)	3 (19%)	1 (6%)

A chi square test of independence found that a smaller gauge size was chosen significantly more often for the sweetcorn than for the courgette, ($X^2(2, N=16) = 19.5$, $\mathbf{p<.001}$).

5.4.2.3 Global versus local viewing condition

Multiple regression slopes were calculated to see if depth reported in the local context conditions was predicted by the depth in the global context conditions, and if this varied with the texture of the object. This was achieved by comparing the mesh from settings made in the local context sweetcorn condition with the mesh in the global context sweetcorn condition through affine transformations. The results are plotted in Figure 5.11. Here the top plot shows all observers' data for the local condition, with the grey line showing the global condition. Observer 12 is a clear outlier, so the second plot shows the same data on a reduced scale, without removing observer 12, to show the pattern of results more clearly.

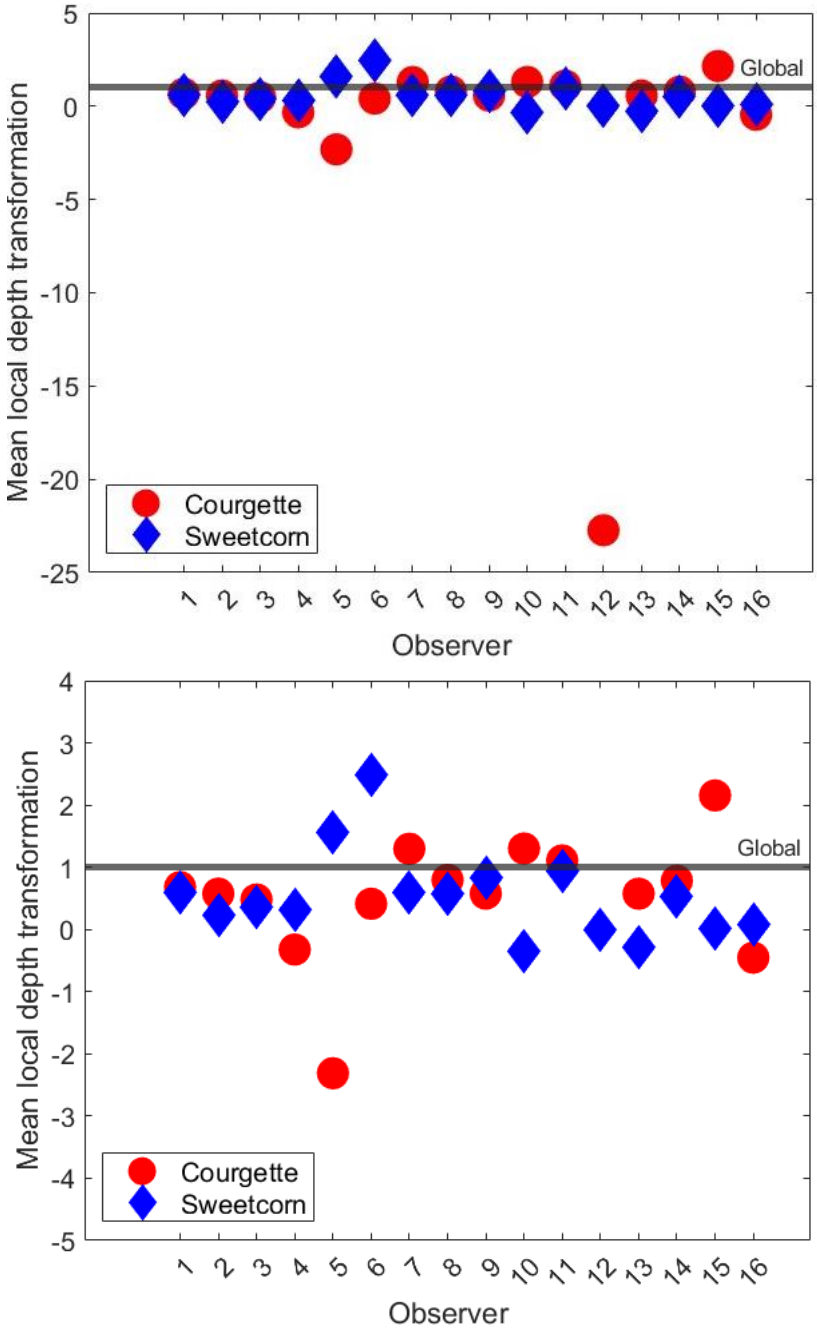


Figure 5.11: Local versus global viewing. Graph showing plot) of depth transformations for local versus global viewing conditions for all observers. Top plot shows all data, and bottom plot shows reduced scale without displaying outlier P12. The blue triangles show local depth transformation of the sweetcorn relative to global settings, and the red circles show courgette. The black line shows global setting and expected local settings if no depth transformation occurred.

This shows a difference in depth reported in the local conditions relative to their global viewing conditions for both the courgette and the sweetcorn texture

conditions for all observers. The mean depth transformation in the local sweetcorn condition of 0.53 shows that around half of the depth was perceived relative to the sweetcorn in the cluttered global context setting. A *t*-test found that the mean of depth transformations was significantly different from 1, meaning that on average less depth was perceived in the local condition compared to the global for the sweetcorn, (*t*(15)= -2.657, SD=0.707, ***p*=.018**, 95% CI [0.154 0.907]), with an effect size of -0.469 95% CI [-0.846 -0.929].

Interestingly, the mean depth transformation for the courgette conditions was seen to be a negative, with a value of -0.94 meaning that the meshes inverted between the local and global conditions. There is also a very high margin of error noted for this condition, and as such did not differ significantly from a slope of 1 due to the variation in scores, (*t*(15)= -1.32, SD=5.89, *p*=.207, 95% CI [-4.08 2.20]), with an effect size of -1.942, 95% CI [-5.081 1.198]. Figure 5.11 shows that this observed effect is due to the depth transformations of several observers, and this is explored in the next section.

5.4.2.4 Presentation order

Exploring this observed difference between the global and local conditions further, the inversion of meshes was found to be related to the presentation order of stimuli, in that those who viewed the local viewing conditions first were more likely to see this as a concave shape than those who had the benefit of the global condition first to provide context. Example of reconstructed meshes for an observer who saw the local condition first (left) and an observer who saw the global condition first (right) is shown below in Figure 5.12.

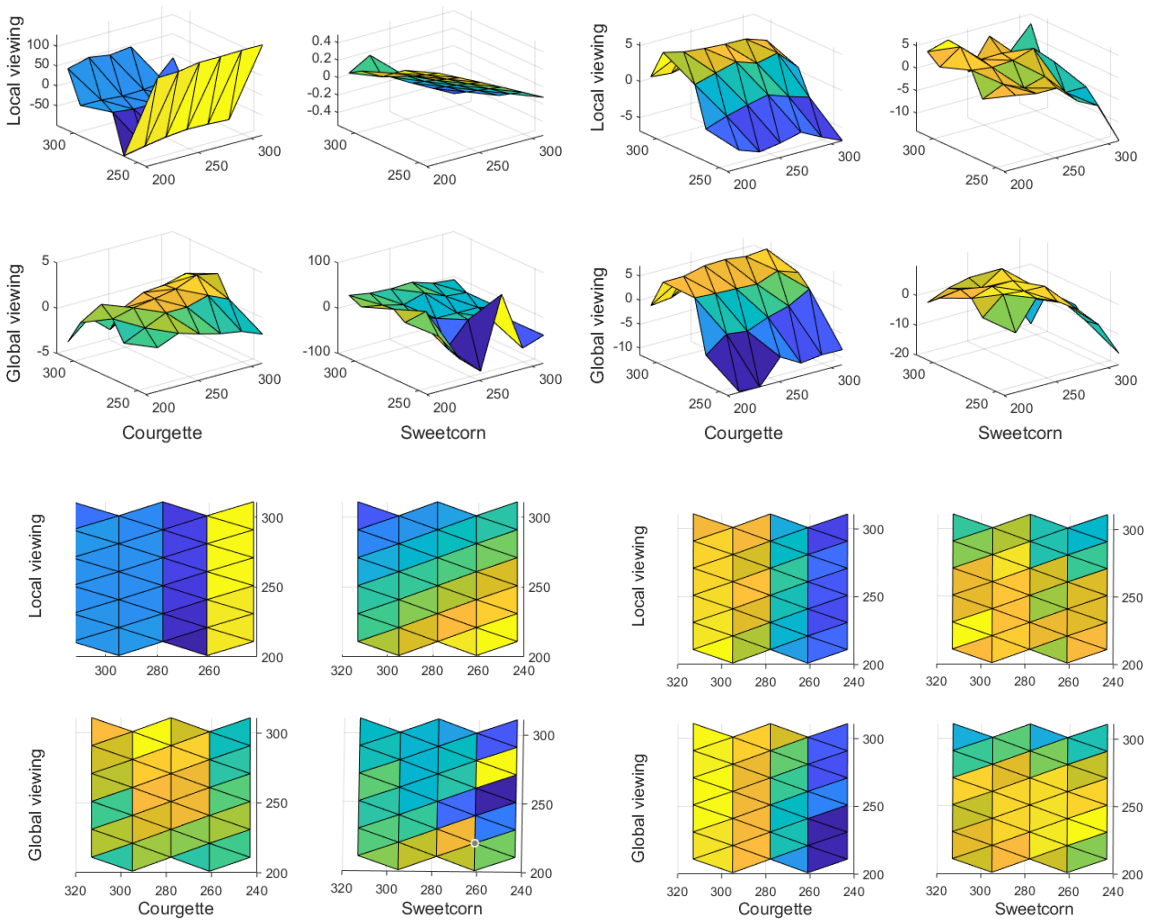


Figure 5.12: Presentation order. Example reconstructed meshes from an observer who saw the local conditions first (left) and an observer who saw the global conditions first (right). Top row shows a side view, and bottom row shows the top-down view of the same data.

As can be seen, one observer who received the local viewing context condition first (left) perceived the courgette as a concave surface when viewing it as if through the letterbox. They then viewed the courgette in the cluttered global scene in the next block of trials, and their perception changed to a convex surface. This effect was seen in several of the observers who viewed the local condition first. However, a *t*-test comparing the courgette local and global means did not reach significance, (*t*(14)= -1.65, SD=5.58, *p*=.121, 95% CI [-11.42 1.48]), effect size of -4.970, 95% CI [-11.424 1.484].

5.4.3 Experiment 2

As before, meshes were reconstructed using depth gradients from the slant and tilt settings captured in the experiment. A visualisation of one observer's data is shown in Figure 5.13:

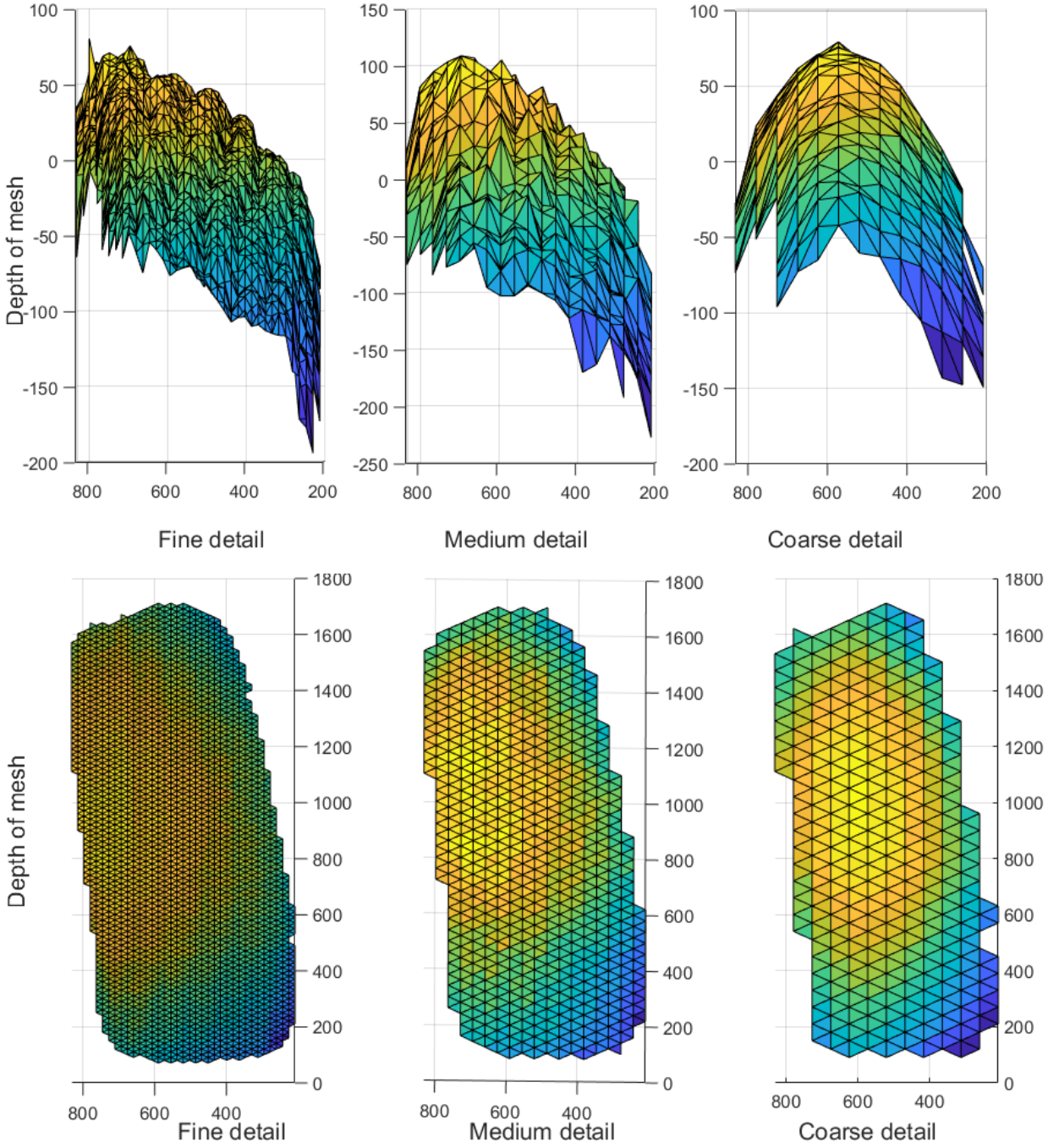


Figure 5.13: Reconstructed meshes Experiment 2. These show significantly more texture captured in the fine detail condition. Top row shows the side view, and the bottom row shows the top-down view of the same data.

It is clear to see that far more surface texture was captured in the fine detail condition than either the medium or coarse conditions. This observed effect was explored using the roughness measure outlined previously. The mean (SE) surface roughness in degrees for each condition are plotted in Figure 5.14:

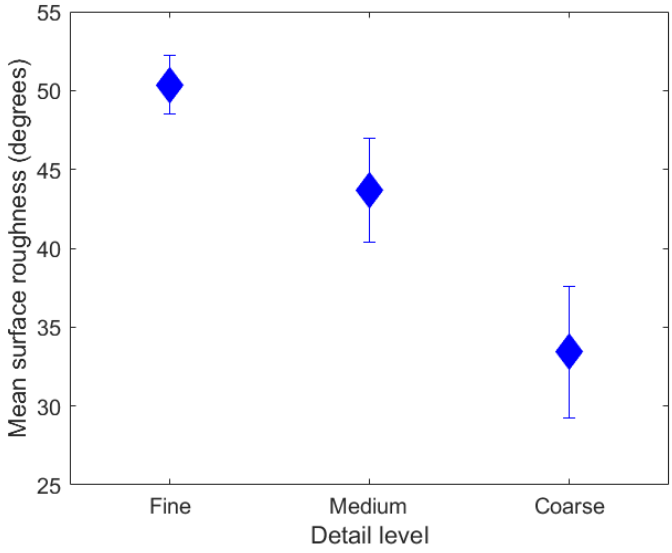


Figure 5.14: Surface roughness Experiment 2. Graph showing mean (SE) surface roughness in degrees for the three detail level conditions.

A linear mixed effects model was run to explore the finding, with the following equation:

Equation 5.4

$$r \sim 1 + g + (1 \mid o)$$

This looked to see if surface roughness (r) was predicted by gauge scaling (g), with random intercepts. This was found to be the best fit of the data. Here, a significant effect of gauge scaling was found, ($F(2,6)= 77.21$, $\boldsymbol{p<.001}$). Exploring the levels of main effects further, it was found that as gauge size and gauge location distance increased, less surface texture was captured as predicted, showing that both gauge size and distance between sampled points is important for surface detail

capture in a gauge figure task. This was found to be significant at both levels of comparison between medium and fine, and coarse and fine, as shown in Table 5-4.

Table 5-4: LME results for the impact of gauge scaling on surface roughness.

Fixed effect coefficient	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
Gauge scale (g) comparing medium (g _2) to fine detail (g _1)	-6.68	1.02	**6**	**<.001*** **	-9.17	-4.19
Gauge scale (g) comparing coarse (g _3) to fine detail (g _1)	-16.95	1.45	**6**	**<.001*** **	-20.51	-13.39

In addition to the local surface texture information, the impact of gauge scaling on global shape captured was also assessed. This looked to see if the depth range, as defined above as the difference between the closest and further pixels on the reconstructed mesh, could be predicted by gauge scaling (g). The mean (SE) depth range for each condition is shown in Figure 5.15.

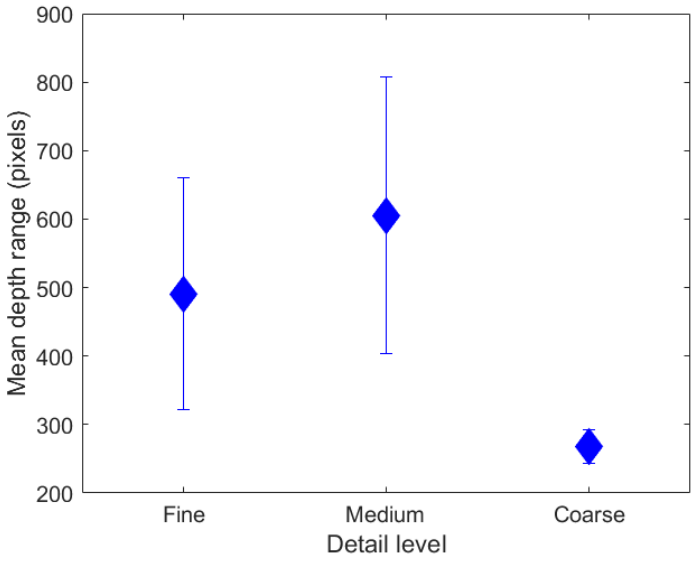


Figure 5.15: Depth range Experiment 2. Graph showing the mean (SE) depth range in pixels, calculated as the closest and furthest pixels in the reconstructed meshes for each detail condition.

An LME was fit to the data using the following equation:

Equation 5.5

$$dr \sim 1 + g + (1 \mid o)$$

This looked at the effect of gauge scaling (g) on depth range (dr). Here, the model with random intercepts was chosen as the best fit of the data. Gauge scaling was found to significantly affect the depth range captured, (F(2,6)= 42.25, ***p*<.001**). Exploring this main effect further, it was found that the coarse detail condition captured significantly less depth than the fine detail condition, with an estimate of 222 fewer pixels in depth. Interestingly, the medium detail condition was found to capture significantly more depth than the fine detail condition, with an estimated increase of 115 pixels in depth between the fine and medium detail conditions.

5.5 Discussion

5.5.1 Experiment 1

The focus of this chapter was exploring the importance of certain parameters of a gauge figure task. In Experiment 1, we manipulated gauge size, object texture, viewing context and presentation order. It was predicted that the size of the gauge would affect the depth and therefore the shape captured from reconstructed meshes. It was also predicted that this difference in captured shape across gauge sizes would differ for the high and low texture objects, with an increased effect of gauge size seen in the high texture sweetcorn condition. We also predicted that there would be less depth captured in the local than the global viewing context, and that the presentation order would also affect settings, with those who viewed the global context first would have the benefit of a Bayesian prior on which to base the local

context judgements, resulting in more perceived depth than those who viewed the local context first.

Here, we consider the findings of the gauge size and texture analysis. The global depth of models from the pixel depth range was not found to be predicted by either gauge size or texture of the objects. This suggests that differing gauge size did not affect the global depth of the shape captured in the gauge figure task, contrary to our predictions. The local detail of the shape captured in the gauge figure task was further explored for high texture and low texture objects with a measure of roughness. This was not found to be predicted by gauge size or object texture, again contrary to predictions. This suggests that merely changing the size of the gauge figure alone does not significantly impact the global depth or the local texture of the object shape captured with a gauge figure task.

When asked which gauge size best captured the shape of objects, the majority of observers preferred a smaller gauge size to capture the shape of the sweetcorn than the courgette. This follows the prediction of the work and shows that navigating the increased surface texture in the sweetcorn condition is better suited to a smaller gauge than is needed for the courgette. Interestingly, of those who preferred a smaller gauge to best capture the shape of the sweetcorn than for the courgette, three quarters of observers preferred the small gauge, with those remaining choosing the medium size. A small majority of observers preferred the medium gauge for the courgette, although responses were more split between that and the large size gauge. These results suggest that observers were conscious of the additional surface texture they would be able to capture with the smaller gauge

for the sweetcorn condition, seeing a perceptually better ‘fit’ of the smaller gauge to the surface.

One participant was found to have the opposite pattern of results for preference, instead choosing the large gauge for the sweetcorn condition. When asked why, they reported that the larger gauge size allowed for easier viewing of the gauge, and made it easier to navigate the more intricate detail in the model. They chose a medium and therefore smaller gauge for the courgette as this did not have as much surface detail to capture. These results inadvertently support our hypotheses by acknowledging the differing level of surface detail, with this participant interpreting the question as which made the task easier, rather than which gauge allowed for the best capture of the shape. Further work in this area could explore this by asking both questions, or allowing an open response for observers to explain their answer.

Affine transformations were conducted between meshes created using reported depth in the local context conditions against the global context viewing conditions. More depth was reported in the global condition than the local conditions as predicted, although interestingly only for the sweetcorn. These findings suggest that the additional depth cues present in the global condition, such other occluding objects and extra information about the light source in the scene contributed to increased accuracy of depth judgements, and that this effect interacted in some way with the increased surface texture of the sweetcorn.

When comparing depth transformations between local and global viewing conditions, a large error was found for the courgette condition under the local viewing context, with depth transformation indicating that the surface of the courgette

in the local condition was viewed as concave instead of convex. Upon further investigation, this effect was found to relate to the order in which observers were presented the stimuli, although this result did not reach significance with the low comparison numbers in this study. However, this finding preliminarily seems to supports that found by Hartle and Wilcox (2022) and suggests that viewing the global context first may provide a benefit of Bayesian a priori knowledge on which to anchor judgements when presented with the local context scene.

It has also been suggested that a perceptual advantage of identification of objects in a scene to be more accurate when there is a semantic relation between objects (Biederman, 1981). Here, this suggests that observers who were presented the local condition first, did not have the semantic context of a cluttered scene of fruits and vegetables to aid recognition of the ‘letterbox’ patch, particularly for the courgette given its lack of surface texture, hence leading to the perception that the surface was flatter, or indeed in some cases concave rather than convex. More work in this area is needed to explore this interesting finding further.

Koenderink (2012) defines the fit of a gauge figure to be pre-cognitive, in that the accuracy of its placement cannot be explained, it is confirmed through visual awareness. In presenting his gauge figure task, Nefs (2008) states that his participants found the task intuitive and had no difficulties in completing it. Egan and Todd (2015) report that all of their participants were able to perceive the ellipse shape gauge figure as a circle oriented in depth in the tangent plane. This is seemingly the case for many participants in this study, although not all. Indeed, one participant reported that despite having instructions, including images, and practice trials, they felt unable to complete the experiment due to not sufficiently

understanding the task. An additional four participants did not complete the trials, as shown by the slant and tilt not changing between successive trials. As these participants did not report back, it is unclear if they also did not sufficiently understand the task. However, due to this being an online study, and the four participants in question being students who had signed up via the SONA system for course credit, it also cannot be ruled out that they were simply moving through the task without completing it in order to obtain the credit without providing meaningful data. There was also one participant whose data was not able to be used due to an incorrect screen calibration due to the effects of stimuli scaling. Were this to have been completed in the laboratory, this error would not have occurred. These highlight some of the downsides of online data collection, and this point will be debated further in the general discussion in Chapter 9.

In summary, while observers showed a strong preference for a smaller gauge during the high texture conditions compared to the low texture conditions, no additional global depth or local surface roughness information was captured. However, Experiment 1 only changed the size of the gauge, not the distance between sample points. A natural next step for this research was to explore if changing the distance between gauge sampling locations, as well as changing the size of the gauge itself, would affect the depth captured. This gauge scaling question forms the basis of Experiment 2.

5.5.2 Experiment 2

This section expands on Experiment 1 by answering the question of what would happen to the level of surface detail and therefore the shape of the mesh

reconstructed from judgements if the possible gauge locations were scaled as well as the gauge sizes. In this experiment, three detail levels were created: fine detail, medium detail, and coarse detail. Both the size of the gauge and the distance between points scaled uniformly with condition.

From the reconstructed meshes, it is clear to see that reducing the gauge size and the distance between sampled points allowed for an increase in capture of surface detail, as predicted. The fine detail condition resulted in clearly rounded sweetcorn kernels, with the medium condition resulting in spiked peaks across the model, and only a vague cylindrical shape being captured in the coarse detail condition.

Gauge size in Experiment 2 was found to significantly negatively predict surface roughness, in that far more textured surface detail was captured in the fine detail condition compared to the medium or coarse detail conditions respectively. In addition, gauge scaling was found to significantly impact the depth range of the reconstructed mesh. Here, the coarse detail gauge scaling captured significantly less depth than the other conditions. Interestingly, the medium gauge scaling captured the most depth. This suggests that a balance must be made between the need to capture the global depth of the model and the local surface texture when designing a gauge figure task.

In summary, to combine this relationship between gauge scaling and both global shape and surface detail, with the lack of global depth or local surface roughness effect in Experiment 1, we conclude that both gauge size and sample locations are important considerations when setting up a gauge figure task. Our findings show that simply changing the size does little to affect either the global

shape from the depth range or the local surface texture from the surface roughness. Additionally, our results suggest that a finer gauge scaling is best for capturing local surface texture, and a slightly larger scaling is best for capturing global shape, leaving future researchers to decide which is more pertinent to their work.

5.5.3 Future work

Several areas for future study are highlighted above, relating to the relevant findings of the work. Here, other general areas are discussed for this work to be expanded on.

A possible expansion on Experiment 2 would be the inclusion of other texture objects to explore any possible interaction between gauge size and texture when gauge location distance has also been scaled. From the findings here, we would predict an interaction between gauge size and texture, in that objects such as the courgette would not benefit from the increased number of points sampled due to the smooth surface texture.

In addition, both main objects in this work, despite having differing surface textures, had a similar underlying global geometry, in that they were both cylindrical in nature. Testing other objects in the scene with differing underlying geometry could explore how global shape capture differs when the underlying geometry of objects also differs.

For practical reasons of testing online, the experiments were conducted assuming binocular viewing of the stimuli by participants. From the literature (Koenderink, van Doorn, & Kappers, 1995; Koenderink, 2012) it would be expected

that more pictorial depth would be observed with monocular viewing. Further work could explore this effect by presenting the stimuli both monocularly and binocularly, which presents an area of further exploration that can be continued beyond the scope of this research. However, if the experiment were to be conducted online again, issues around how to ensure monocular viewing would need to considered carefully.

Additionally, these findings are important in the context of viewing natural objects. As highlighted in the introduction, the logic of fractal geometry was key to the design of this work. The experiments here contained cluttered scenes, although only a single object was measured in each scene. This could be expanded to probe points on multiple objects within the cluttered scene. In addition, the stimuli here were all scanned objects of natural fruit and vegetable objects. Further work on this could expand the range of natural objects used as stimuli. Chapter 8 addresses this by using the methods outlined above for judgements of faces.

Finally, whilst this work provides valuable insight into the parameters and limits of the gauge figure task, and our scaling of the gauge figures and distances has been reported to provide a reference for future studies, we have scaled in relative terms rather than focus on absolute values. A useful metric for future study would be a measurement of the stimuli or surface texture in comparison to the gauge figure size and sampling distance. As such, an important next step in this research would be to provide a comprehensive set of guidelines for how to choose the correct parameters for gauge figure output relevant for the desired level of perception capture.

6 Binocular cues provide a modest contribution to depth judgements at close distances in naturalistic scenes

6.1 Abstract

3D shape perception depends on multiple cues, which can be weighted according to their reliability in order to provide statistically optimal estimates of depth. Typical psychophysical tests of optimal cue combination seek to isolate two cues and vary their relative reliabilities. This allows us to measure their relative weightings and how these vary with changes in the reliability of each cue. In typical natural scenes, many cues will be available, and the weighting of each cue will depend on both the availability and reliability of all other cues. The current study measured the weighting of binocular cues in naturalistic scenes in which multiple cues, such as perspective, texture and shading, also contributed to the perception of depth. In Experiment 1, observers were presented with a single object which had been 3D scanned and rendered at two distances of 50 and 96cm. We manipulated the effective depth from binocular cues by varying the simulated interocular distance between 0 and 2 times that of each observer. A ‘pushpin’ gauge figure was presented monocularly at multiple points on the object, and its slant and tilt were adjusted so that it appeared to lie flat on the surface. The best-fitting surface mesh was calculated. Slant and tilt settings were highly correlated with the ground-truth values calculated from the object scan. At the near distance, the depth range specified by observers’ slant and tilt settings increased with the simulated IOD with a gain of 9%. At the far distance, there was no change in the perceived depth range with simulated IOD. Experiment 2 followed up on these findings with a cluttered scene. Here, a 10% weighting of

binocular cues was found at the closer distance, reducing to 9% at the further distance, suggesting that increased cue information resulted in a stronger binocular weighting than the single object condition. Interestingly, binocular perturbation was found to be more successful with IOD gains smaller than observers' own, which we suggest is due to accommodation and convergence interference. These results show a modest but reliable contribution of binocular cues to the perception of surface shape for near objects. As predicted from geometrical considerations, this contribution reduced as the distance to the object increased.

6.2 Introduction

Depth perception is often a vital skill for a person when it comes to navigating their world. As outlined in previous chapters, many factors go into making a judgement of depth, with information being available from a wide variety of cues. The depth cue of particular interest in this chapter is derived from the fact that humans have two forward-facing eyes. This binocular vision provides additional information about the location and shape of objects in the visual field. As the eyes are set apart by an average of 63mm for adult humans (Dodgson, 2004), this creates slightly different images on the two retinas (Nefs, 2008), and this difference creates the cue of binocular disparity.

Binocular disparity is often referred to as one of the most reliable cues to depth (Harris, 2004; Keefe, Hibbard, & Watt, 2011). Julesz (1971) showed that depth can be perceived when only the cue of binocular disparity is available, through the presentation of random-dot stereograms (RDS), where a pair of random dot images are presented one to each eye using a stereoscope, and small horizontal differences

between the dots‘ locations in the two images create the appearance of depth. When binocular disparity is present, estimates of depth can be very precise, with stereoacuity of just a few seconds of arc (Viguier, Clément, & Trotter, 2001). This chapter explores the importance and measures the contribution of binocular disparity in combination with other cues when perceiving 3D object shape.

6.2.1 Calculating binocular disparity

To begin with, the mechanics of binocular disparity will be explained in more detail. As this cue utilises information from the two eyes together, early work in the field focused on triangulation using the principles of geometry, as described by Linton (2022). This was presented in Chapter 3 to explain the cue of vergence, and is presented here and shown in Figure 6.1 below in the context of this chapter for ease of reference to explain the mechanisms behind the cue of disparity. This shows how distance (D) can be calculated using a known interocular distance or IOD (I) and an estimate of the given vergence angle (a), the angle at which the two eyes meet on an object (Cutting & Vishton, 1995).

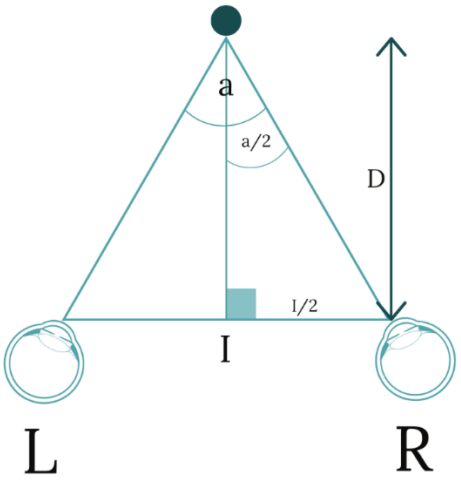


Figure 6.1: Geometry of triangulation. Showing the geometry of triangulation between the left and right eyes (L and R) and a fixated point, and how a known

interocular distance (I) and an estimation of a given vergence angle (a) can be used to calculate distance (D).

Figure 6.1 shows the triangle created between the two eyes (L and R), and the observed object, and how this can be split into two to create two right-angle triangles. From here, basic principles of trigonometry can be applied, by calculating the tangent of an angle as before, recapped here for ease of reference:

Equation 6.1

$$\tan\left({}^{a}/_{2}\right) = \frac{{}^{I}/_{2}}{D}$$

If we assume that an estimate of the vergence angle is provided by the brain from the condition of the extraocular muscles (Tresilian, Mon-Williams, & Kelly, 1999), the above equation can be rearranged to calculate an estimate of distance (D) as such:

Equation 6.2

$$D = \frac{{}^{I}/_{2}}{\tan\left({}^{a}/_{2}\right)}$$

This process shows how distance (D) may be estimated from a known IOD and an estimate of the vergence angle, such that distance (D) is equal to half the IOD (I) divided by the tangent of half the vergence angle (a). It should be noted however that this equation is based on the assumption that the point being fixated falls along the midsagittal plane, i.e., that it is in the middle of the visual field, which is not always the case (Cormack & Fox, 1985) and will be covered more in the equations below.

This principle of triangulation can be extended to show binocular parallax through the difference between two points, as shown in Figure 6.2, where an observed point extends a wider vergence angle (a) than a point further away in distance (D), which extends a narrower vergence angle (b). It is this difference between these angles (a and b) that creates horizontal retinal binocular disparity (Stidwell & Fletcher, 2011), which highlights how disparity is determined by distance.

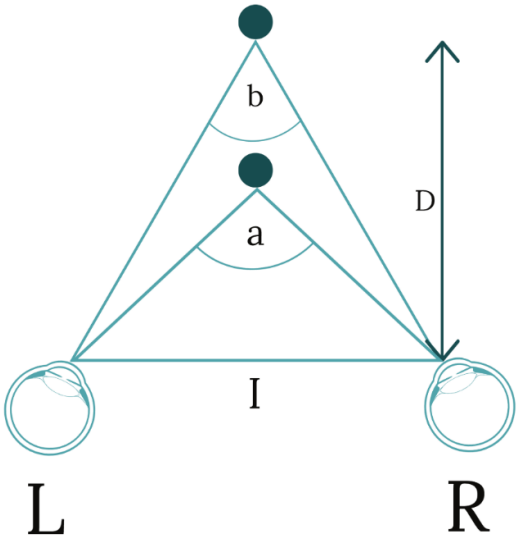


Figure 6.2: Vergence angle comparison. Diagram to show how the two eyes joining on a closer point creates a wider vergence angle (a) than when they converge on a further point which creates a narrower vergence angle (b), by distance (D) for a known interocular distance (I).

Figure 6.3 presents an image that shows points along the horopter, a theoretical curve that defines points in space that fall on corresponding locations on the retina, such as the blue dot in the image below, the basic form of which is known as the Vieth-Müller circle or VMC (Stidwell & Fletcher, 2011). Points in space that do not fall on this curve project onto differing points on each retina, and therefore have disparity, as can be seen for the purple dot.

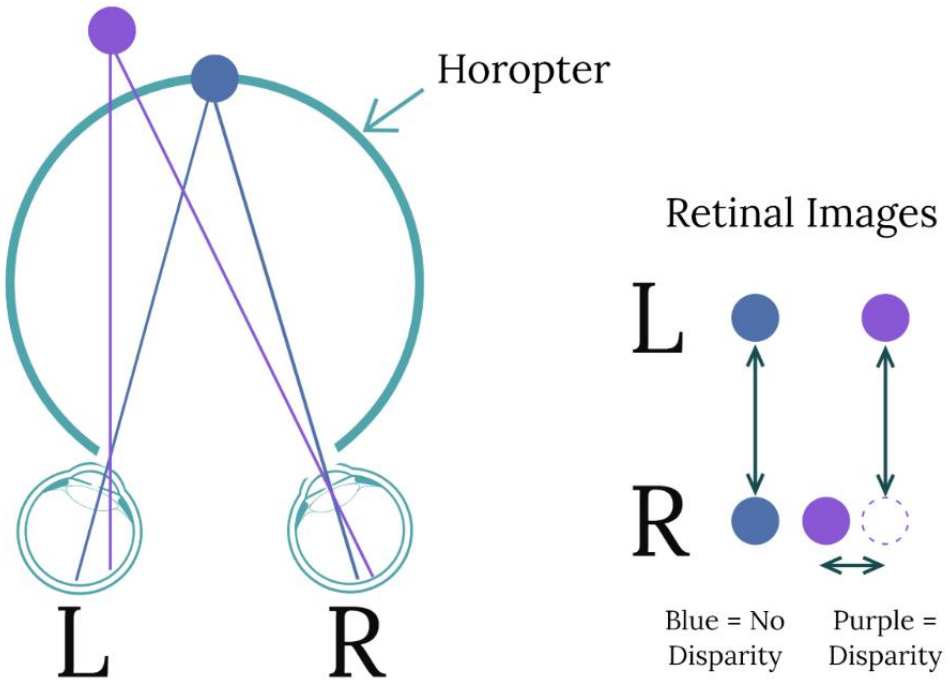


Figure 6.3: The horopter and disparity. Image showing how viewing a point on the horopter such as the blue point creates corresponding points in the retinal images with no disparity, and how viewing a point elsewhere in the scene creates disparity between the two retinal images.

Figure 6.3above shows how different points in the scene fall on different points of the retina, and how these non-corresponding retinal points create disparity when images are fused for stereopsis (Stidwell & Fletcher, 2011). Figure 6.4 illustrates crossed and uncrossed disparity, relative to the point of fixation.

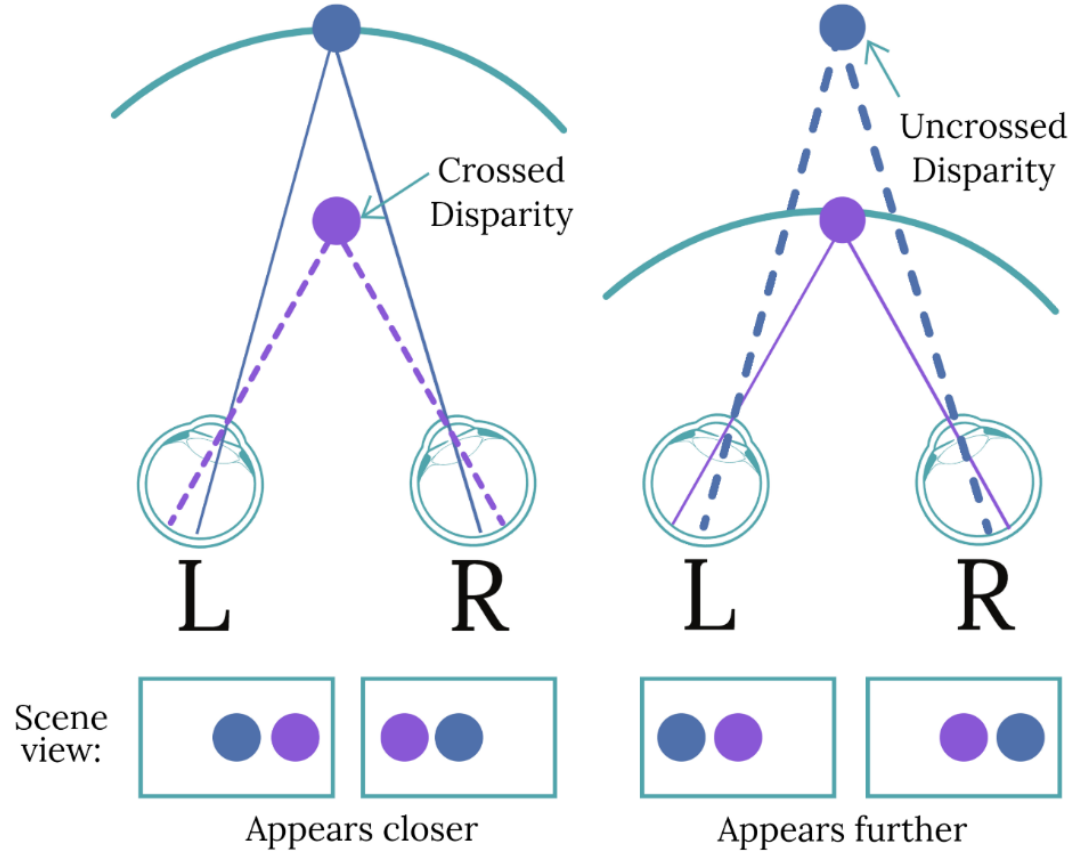


Figure 6.4: Crossed and uncrossed disparity. Showing on the left how points appear to have crossed disparity when a point (purple) is closer to the observer in the scene than the point of fixation (blue), and on the right how a point (blue) has unfixed disparity when it is further away than the point of fixation. The boxes at the bottom show how the scene would appear to the observer for left and right eye views, with the point of fixation in the middle.

In the image on the left, the observer is viewing the blue dot, and the purple dot which is closer to the observer has intersecting secondary lines of sight closer than the point of fixation, which gives it crossed disparity and it appears closer in the scene, which can be seen in the 'scene view' boxes. The image on the right shows that when an observer views a point closer in the scene (purple), points further away in depth (blue) have secondary lines of sight that intersect beyond the point of fixation and therefore have an uncrossed disparity and appear further away in the scene (Stidwell & Fletcher, 2011).

Taken together, the figures above show that the cue of binocular disparity depends on distance and IOD (Figure 6.1 and Figure 6.2), as well as depth (Figure 6.3 and Figure 6.4). Cormack and Fox (1985) provide a comprehensive explanation of how to compute retinal disparity using these parameters through a series of equations which will be covered here:

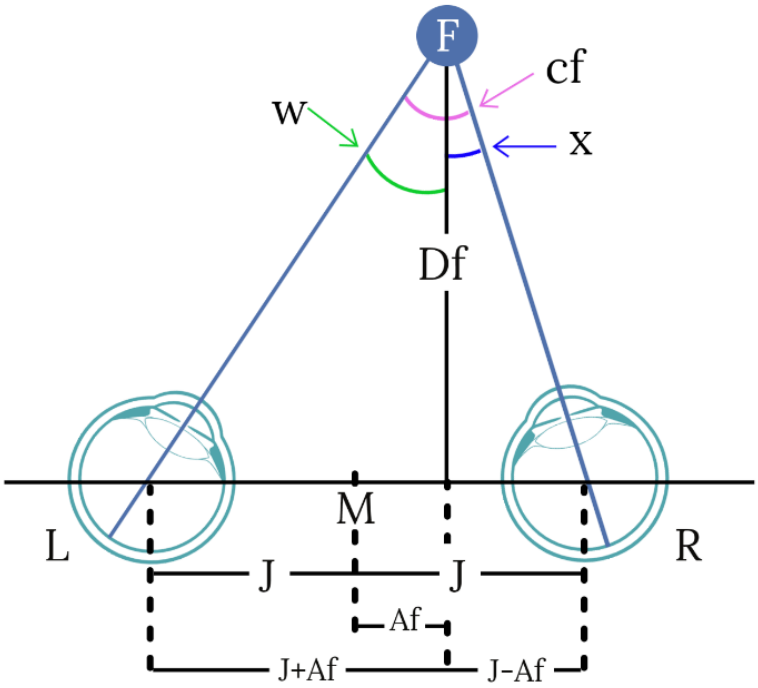


Figure 6.5: Geometric relationship between the eyes and the point of fixation. Showing the geometric relationship between the eyes (L and R) and the point of fixation (F) at a set distance (Df).

Figure 6.5 defines the geometric relationship between the two eyes and the point of fixation as described above. Here, cf (shown in pink) defines the convergence angle of the eyes (L and R) joining on the point of fixation (*F*), and is equal to the sum of angles w (green) and x (blue). *M* is the midpoint of the line between the centre of rotation of each eye and Df defines the distance between this line and the point of fixation, mirroring the triangulation shown in Figure 6.1. There, *I*

denotes the IOD, and here, J is equal to half the IOD. Af is defined as the distance of the point of fixation from the midsagittal plane, anything to the left of which would be negative and so $J + Af$ would be less than J. The equation for the tangent of angle w, calculated by dividing $J + Af$ by Df, is therefore:

Equation 6.3

$$w = atan\left(\frac{(J + Af)}{Df}\right)$$

Angle x is calculated in much the same way, here using $J - Af$:

Equation 6.4

$$x = atan\left(\frac{(J - Af)}{Df}\right)$$

The convergence angle cf is therefore:

Equation 6.5

$$cf = w + x$$

Cormack and Fox (1985) highlight problems with previous simplification of the equations for calculating binocular disparity which made assumptions that angles w and x were symmetric, meaning that the point of fixation falls along the midsagittal plane. However, as in the above diagram, and with some of the stimuli presented in this work, the angles are not equal, thus the importance of defining both angle w and x in these equations.

Once the calculations for these are established, they can be used to calculate a difference between the disparity of two points. Figure 6.6 outlines the geometry for a target point in the scene (*T*), here beyond the point which is being fixated upon (*F*).

This shows how this point stimulates the retina relative to the point that is being fixated.

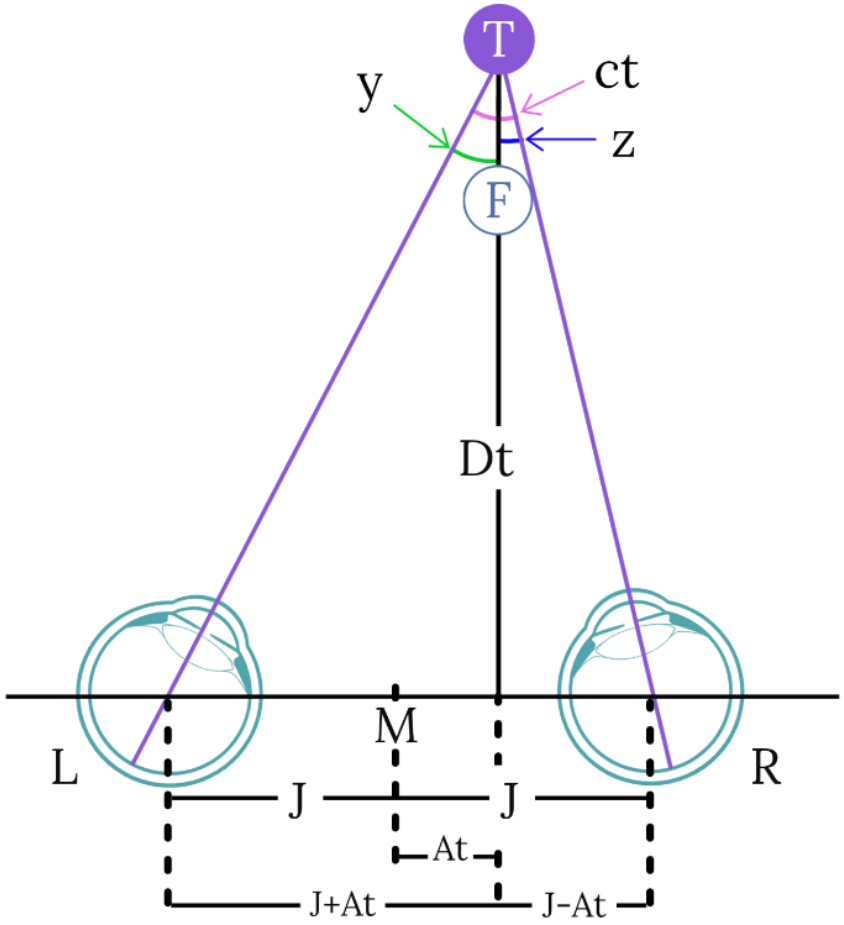


Figure 6.6: Geometric relationship between the eyes and a target point. Illustrating the geometric relationship between the eyes (L and R) and the target in the scene (T) relative to the point of fixation (F) at a set distance to the target (Dt).

Using similar notation as above, Dt here is the distance between the line between the centre of rotation of both eyes and the target (*T*), with At representing the difference between Dt and the midsagittal plane. Despite this point not being fixated, the theoretical convergence angle can be calculated as ct, which as outlined above will be narrower than the convergence angle of cf due to it being further away

in distance. The equations for angle y, angle z and angle ct mirror those above and are as follows:

Equation 6.6

$$y = atan\left(\frac{(J + At)}{Dt}\right)$$

Equation 6.7

$$z = atan\left(\frac{(J - At)}{Dt}\right)$$

Equation 6.8

$$ct = y + z$$

The difference between Equation 6.5 and Equation 6.8 is used to calculate disparity as the difference between the vergence angles:

Equation 6.9

$$disparity = cf - ct$$

Combining the above gives an equation that can be used to calculate disparity between the point of fixation (*F*) and any given target (*T*) in the scene:

Equation 6.10

$$disparity = \left\{atan\left[\frac{(J + Af)}{Df}\right] + atan\left[\frac{(J - Af)}{Df}\right]\right\} - \left\{atan\left[\frac{(J + At)}{Dt}\right] - atan\left[\frac{(J - At)}{Dt}\right]\right\}$$

Therefore, an object with a fixed depth, when viewed with a fixed IOD, will as discussed extend a wider vergence angle at a closer distance and therefore have more disparity, than when viewed at a further distance. This non-linear relationship of distance being inversely proportional to disparity is plotted in Figure 6.7.

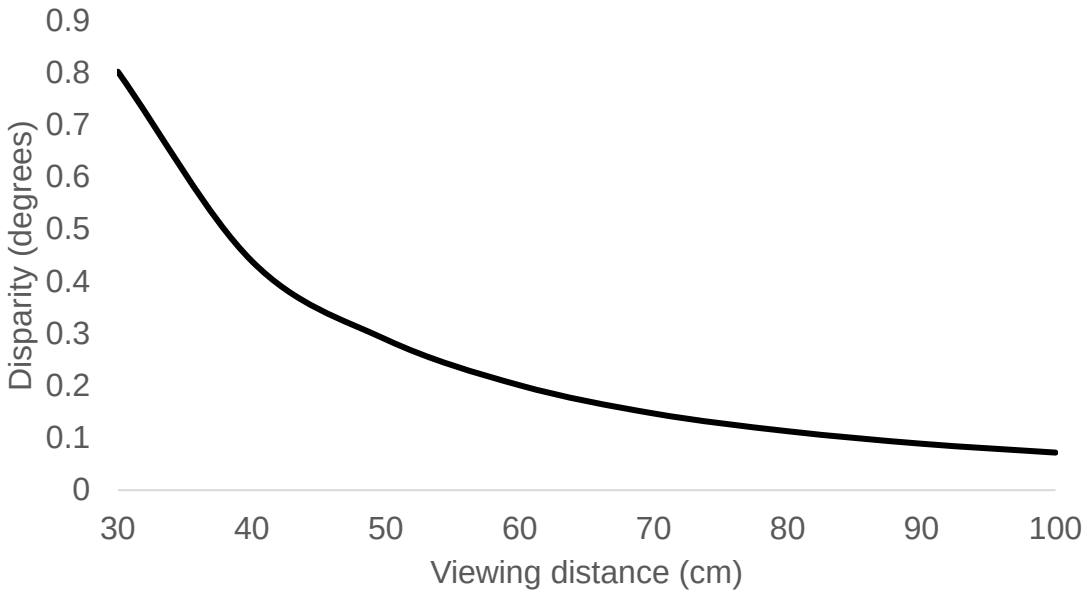


Figure 6.7: Change in disparity with viewing distance. Graph plots the expected change in disparity (in degrees) when viewing an object by viewing distance (cm) for a fixed IOD of 6.3cm.

6.2.2 Gain on binocular disparity

This can be built on by altering Equation 6.10 and, rather than having a fixed IOD, instead holding distance constant by multiplying J by a gain (g) on IOD to simulate a different IOD, which creates an effective gain on binocular disparity:

Equation 6.11

$$disparity = \left\{atan\left[\frac{(gJ + Af)}{Df}\right] + atan\left[\frac{(gJ - Af)}{Df}\right]\right\} - \left\{atan\left[\frac{(gJ + At)}{Dt}\right] - atan\left[\frac{(gJ - At)}{Dt}\right]\right\}$$

With a gain of 1, disparity would remain the same as the original calculation for *F*, a gain of 0.5 would halve disparity, and a gain of 2 would double disparity. Therefore, this shows that the relative disparity will vary with changing IOD and a fixed depth, as in Figure 6.8.

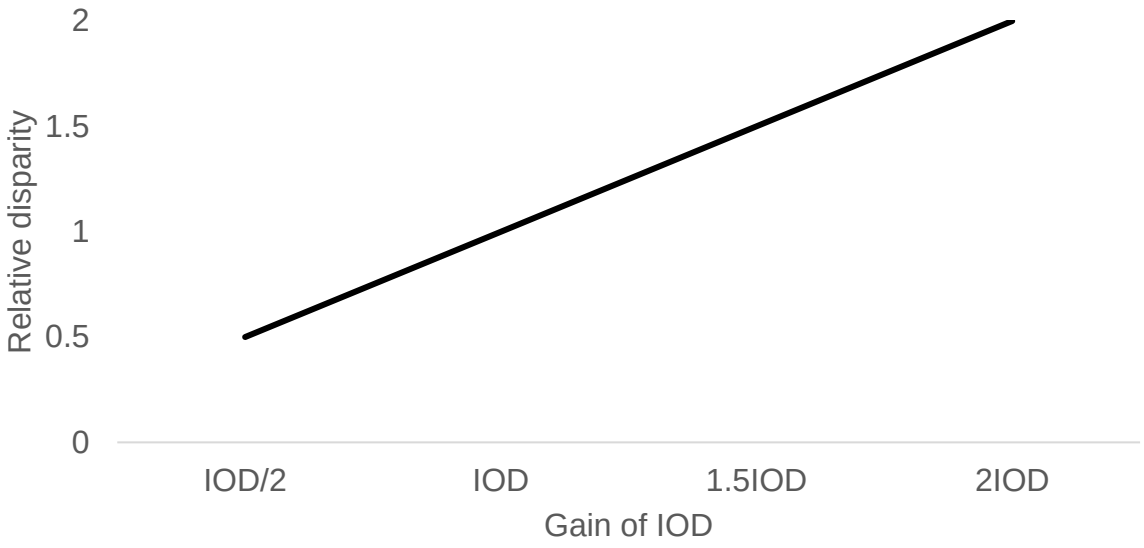


Figure 6.8: Predicted change in disparity with varying IOD. Graph plots the expected linear change in relative disparity with increasing IOD gain.

Given the complexity of the information processed in the depth estimate, and the complexity of scenes in natural viewing, the stimuli used to test these hypotheses need to accurately represent the job the visual system is doing. In order to maintain sufficient experimental control, some experiments in this field choose to present simplified stimuli. However, it is possible to achieve the desired experimental control whilst still presenting naturalistic stimuli, by presenting rendered scenes of 3D models, created using meshes made from scans of real objects. In addition, by using technology to present the images, rather than natural viewing conditions, it allows for the manipulation of variables such as the observer's effective IOD, as presented in Figure 6.8.

When the stimulus is, for instance, an image from a photograph such as in Koenderink, van Doorn and Kappers' (1992) study, this can be used to investigate the change in depth across the image, but it is not possible to calculate a true ground truth that includes depth magnitude. However, as this experiment used 3D models,

an objective ground truth of the depth of each point did exist, allowing for comparison with perceived depth measured in the experiment.

As the aim of this body of work is to explore naturalistic stimuli in natural viewing conditions, the present study presented 3D models both as a single object, and as part of a cluttered scene. Increased clutter in the scene provides additional depth cue information such as occlusion and additional information about the light source in the scene. In addition to providing several 3D objects in the scene, this work expands on previous chapters by using natural objects that do not necessarily have a simple geometric shape that might be known to the viewer in advance. For instance, when using a recognisable natural object such as a tomato or an apple, certain dimensions will be assumed due to prior knowledge of these fruits. Both of these two examples are approximately spherical, and while individual apples or tomatoes may vary in dimensions of width or height, someone with prior knowledge of these items will have a schema that dictates sphericity. The observer knows with a fair amount of certainty that these objects will not, for instance, have any angular vertices. As such, the present work provided the observer with some natural objects with a predetermined geometric shape, here a cylindrical ear of corn, as well as other natural objects in the scene that do not necessarily follow these geometric principles. Here, this was in the form of a scone, where the observer was expected to have some familiarity of the object in a general sense, but not the specific dimensions presented to them.

6.2.3 Binocular disparity in cue combination

Despite binocular disparity being one the most reliable cues to depth (Harris, 2004; Keefe, Hibbard, & Watt, 2011), as mentioned previously, the visual system relies on the information from many different cues at once. Cutting and Vishton (1995) outline the two binocular cues mentioned here of binocular disparity and convergence, as well as a number of monocular cues, such as pictorial cues which are derived from the retinal image, which include occlusion and shape from shading. These cues are averaged, weighted by their respective reliabilities and noise, to provide statistically optimal estimates of depth (Trommershäuser, Körding, & Landy, 2011).

There are two competing theories for how these cues are combined using the Bayesian approach, which uses previously known information, or a priori knowledge, to make decisions, known as the strong and weak fusion models (Tyler, 2019). The strong fusion model states that the brain compares two-dimensional information reported by the image against all possible 3D scene structures it has encountered and chooses the best match using the maximum a priori (MAP). However, Tyler (2019) criticises this theory in that the sheer number of possible scenes to compare the image against is far too many for the brain to compute quickly enough to make this theory pragmatically viable, estimating the number of possible scenes to be in the order of quintillions, or 10^{18}.

In comparison, the weak fusion model provides a more realistic theory of how the brain might handle the weighting of depth estimates to make a judgement (Tyler, 2019). This theory follows a mid-level Bayesian approach that assumes the scene is made up of a continuous depth map of different surfaces, beginning at the local

starting point. This local starting point ‘anchors’ the choices of possible slant values, and enables rapid sampling around this point. From here, the MAP is limited to a much smaller, more manageable number of options, making this a more plausible approach given the resources available for this task, and how quickly the brain can make depth judgements.

The knowledge of how the brain combines weighted depth estimates into a judgement can be used to investigate the individual weightings of a cue by manipulating only that cue and holding the information reported by all other cues constant. Therefore, if the depth reported changes across conditions, it can be seen to be as a result of the individual cue of interest, and the weighting of this cue to the final depth judgement can be calculated. If the depth reported does not change across conditions, it would show that the cue of interest is not highly weighted compared to the other cues available, and therefore the changes in depth reported by that cue were ignored in the final depth estimate. Bayes formula as cited by Knill and Richards (2008) may be used to account for how the weighting of cues may be calculated to also take into account prior knowledge, such as previous experience of cue reliability, known as a priori information. Bayes formula then computes a posterior from the information after combining cues and prior information, from which a depth estimate can be obtained. The equation to combine this prior information (p), into the depth estimate for an image ($\mathbf{I}$) with certain scene properties ($\mathbf{S}$), including depth, is as follows:

Equation 6.12

$$p(\mathbf{S}|\mathbf{I}) = \frac{p(\mathbf{I}|\mathbf{S})p(\mathbf{S})}{p(\mathbf{I})}$$

The above can be simplified by treating $p(\mathbf{I})$ as the probability of the occurrence of an image as a normalising constant (Knill, Kersten, & Yuille, 2008), to give:

Equation 6.13

$$p(\mathbf{S}|\mathbf{I}) \propto p(\mathbf{I}|\mathbf{S})p(\mathbf{S})$$

However, this equation offers a calculation for the single depth estimate, not a measure of the cues individually. The next section explores methods that can measure the relative reliabilities of individual cues for their contribution to the overall depth estimate.

6.2.4 Measuring the contribution of individual cues

The varying weights applied to the cues through their differing reliabilities poses a challenge as to how to measure the contribution of individual cues. In order to measure just one cue, typically previous work has chosen to strip away everything except the cue of interest in order to measure it in isolation, or reduce the information down to two cues and vary their relative reliabilities. For example, Hillis, Watt, Landy and Banks (2004) tested a maximum-likelihood estimation (MLE) model of depth cue combination for the cues of perspective from texture and binocular disparity for estimates of surface slant. They presented four participants with stimuli with either disparity information, through random dot stereograms for binocular viewing, or a texture patch viewed monocularly where one side appeared to be slanted further back in depth for the texture only condition, as part of a slant discrimination task.

These stimuli were presented either as a single-cue condition by presenting just the texture-based stimuli or the disparity-based stimuli in isolation in order to

estimate the individual reliabilities, or through a two-cue condition in which some stimuli were presented with a conflict of slant between the texture and disparity cues, in order to measure their relative weights. In doing so, the weighting of one cue over the other becomes apparent, as the estimate skews towards the cue with the highest reliability with increasing cue perturbation as described above.

In the single-cue conditions, they found that the cue of texture increased in reliability with increasing slant, and was consistent across viewing distance. The reliability of disparity was found to decrease with increasing viewing distance, and interacted with varying degree of slant across viewing distance. This finding has been replicated by other work in this area (Keefe, Hibbard, & Watt, 2011), and as such the present work presented stimuli at both a close and far distance to examine the contribution of binocular disparity across viewing distance, with the prediction being that the binocular weight would reduce when going from close to far viewing as the reliability of this cue reduces with distance.

When both binocular disparity and texture cues were available, observers were found to make slant judgements with higher precision, implying statistically-optimal combination of these cues to maximise the precision of the depth estimate (Hillis, Watt, Landy, & Banks, 2004). Other work has also found this statistically-optimal combining of cues not just for depth estimates, but between other sensory modalities, such as between visual and haptic information (Ernst & Banks, 2002). As with the studies presented here, the benefit of isolating a single cue, or a pair of cues in order to perturb them in this way is the high levels of experimental control it affords, allowing the researchers to be sure that the observed effect comes from the manipulation of the cue of interest. However, there is little ecological validity in the

viewing conditions and stimuli used in these experiments, and the results are harder to extrapolate to natural viewing conditions, which is a main consideration of this work.

A solution to this issue was outlined by Koenderink (1998) whereby stimuli are presented in full-cue viewing conditions, and all cues are held constant, except the cue of interest. Any results can then be linked to the manipulated cue. This creates an ‘operating point’ around which the contribution of the cue may vary, as introduced in Chapter 5.

6.2.5 Measuring surface orientation

Using this full-cue operating point methodology, the present work investigated the weighting given to binocular disparity using a 3D orientation gauge figure task. Experiments of these kind present participants with an object, either using a photograph or, as is the case here, a rendered scene of a 3D model. On this image, an item resembling a pushpin is superimposed, consisting of a circular element, with a pin which sticks out perpendicularly from it. By rotating the gauge figure until the circle appears to be painted onto the surface of the object (Koenderink, van Doorn, & Kappers, 1995) information about the perceived shape of the object can be obtained. The rotation of the pin gives an indication of the slant and tilt of the object‘s surface at each location sampled.

Nefs (2008) presents a method, introduced by Koenderink, van Doorn and Kappers (1992) for turning these slant and tilt values gained in a gauge figure task into depth gradients, which describe the orientation and magnitude of surface locations, and using these to reconstruct a best fitting 3D mesh of the perceived

global surface of the scene. This 3D mesh can then be used, such as in this work, to compare perception against the objective ground truth of the original 3D model of the stimuli. The reconstructed mesh was used to provide two measures of shape and depth change between conditions. Firstly, we used the overall depth range between the closest and furthest pixels in the reconstructed mesh to give an overall measure of depth of the perceived object surface. These were normalised relative to a gain of 1, which is covered in detail below. The second measure we took from the reconstructed meshes was to use affine transformations. These have been used previously in other chapters of this work using the methods outlined by Nefs (2008). Here, the reconstructed mesh surfaces were compared with a reference, in this case the gain of 1 to represent the original stimulus presented with no depth manipulation. Comparing the meshes in this way gave us a measure of how the shape changes between conditions. This is presented in detail in Chapter 2: Methods, along with definitions for slant and tilt used in this work.

6.2.6 Simulating IOD gain

One way to simulate a difference in binocular disparity is to render with a different simulated IOD. This is the premise behind the creation of 3D films, whereby the placing of the cameras and the subsequent disparity between the two images creates the appearance of depth protruding from or receding into the visual scene. Simulating a gain on disparity in this way negates any additional concerns for scene lighting that would be apparent if the 3D models themselves were altered. Hibbard, van Dam and Scarfe (2020) present the implications of varying IOD for virtual reality. They argue that when the two dynamic retinal displays inside the virtual reality

headset match the optic array experienced by an observer, this can create veridical viewing conditions, as if the observer were viewing the scene in real life. However, they point out that if the visual input from the display does not match the observer's parameters, such as their IOD, this mismatch creates conflicts which can lead to unnatural fixations and ocular discomfort (Hibbard, van Dam, & Scarfe, 2020).

Whilst these conflicts can lead to unpleasant experiences when using virtual reality, they can be exploited in visual research to manipulate the effective IOD of the observer. This approach was taken in the present study. We simulated a larger or smaller IOD than the observer's own when rendering 3D stimuli to be presented using 3D glasses and the VIEWPixx system, details of which will be provided in the methods, or alternatively in Chapter 2: Methods.

Figure 6.9 builds on the triangulation theory presented earlier in the chapter (a), by showing how images can be presented to the left and right eyes in virtual environments to simulate an object at a distance, D, for an observer with an IOD, I (b).

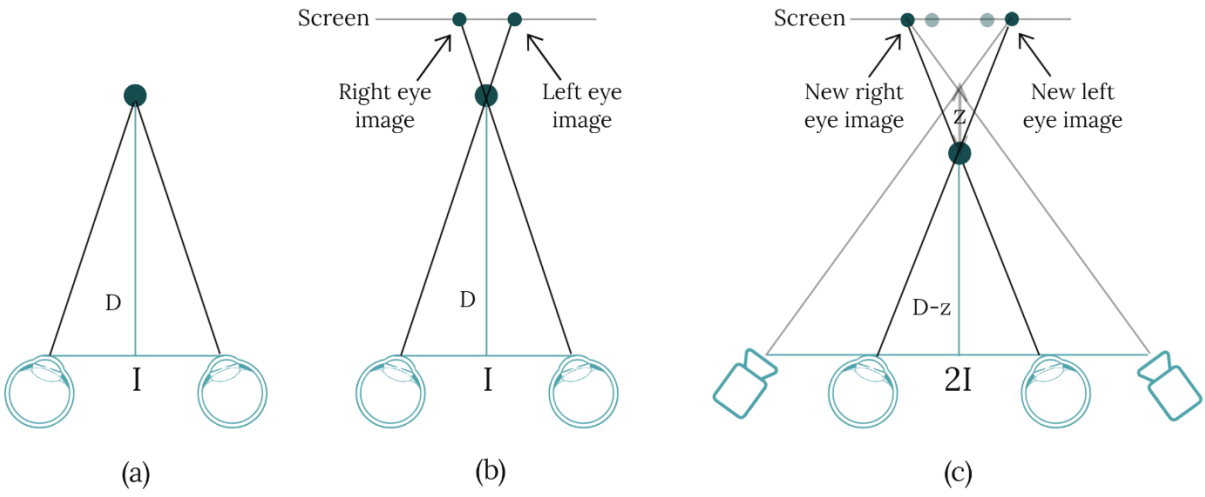


Figure 6.9: Simulating IOD gain. Representation of how images of objects for the left and right eyes can be positioned virtually to project an image in front of the 3D screen and simulate an increased IOD, here doubled from I (b) to 2I (c).

As shown in Figure 6.9, images are projected going to the left and right eye respectively, which render the object as if it were in front of the 3D screen. This causes the eyes to converge on a point closer than the screen to fixate on the object, which creates the 3D effect. The distance between where the left and right eye images are placed is calculated to create a veridical 3D object for the observer's IOD, as seen in Figure 6.9b. However, by presenting the left and right objects at varying locations horizontally, it is possible to simulate any IOD, in order to manipulate the depth perceived. By using cameras at twice the observer's IOD, 2I, as shown in Figure 6.9c, a larger disparity between images can be created, and projecting these back to the eyes creates the appearance of increased depth of points, z. As some points would have zero disparity between images, these would not increase in depth with this process. The combination of these points that remain the same in both images and the points that increase in depth creates the appearance of an overall scaling of the object in depth.

This methodology relies on observers knowing and using their own IOD to judge distance, which Taya (2023) has recently presented evidence for. Observers made depth judgements of stereograms and results showed a strong correlation between IOD and perceived depth, consistent with the geometric relationships outlined in Figure 6.5 between IOD, distance and disparity. This shows that it is possible to vary binocular cues while keeping the information from all other cues in the depth estimate constant.

This increase in perceived depth with increasing binocular gain is not expected to have a slope of 1. If people were to solely rely on binocular cues, a slope of 1 between depth and perceived depth would be expected, as the increases in

binocular gain should evenly increase the depth perceived by participants, as shown by the line with a slope of 1 in Figure 6.10 below.

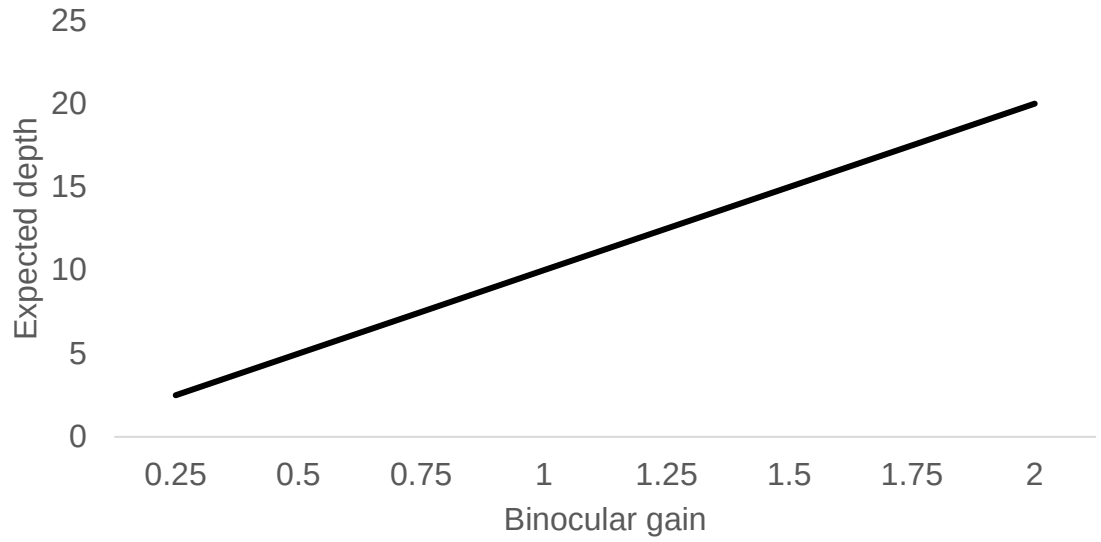


Figure 6.10: Predicted effect of gain for binocular cues only. Showing a representation of how changes in binocular gain might affect the perceived depth of objects reported by binocular cues if solely relied upon to make depth judgements.

However, observers do not solely rely on information from binocular disparity to form depth judgements; many other cues report their own estimate of depth. The information reported by the group of cues is not always the same, as some cues are inherently more reliable than others, and some are more prone to bias and noise or vary with factors such as distance and slant (Cutting & Vishton, 1995; Hillis, Watt, Landy, & Banks, 2004; Keefe, Hibbard, & Watt, 2011). This creates a set of cues reporting slightly different estimates, which presents a problem of which estimate is to be believed and actioned upon. One way for the visual system to solve this would be to simply combine these into an average. This method, however, would not account for the previously mentioned varying levels of inherent reliabilities within the cues. As such, the brain weights the individual cues based on their reliability, and these weighted estimates are combined in a statistically optimal way. Hibbard, van

Dam and Scarfe (2020) present equations to calculate the weighting of binocular cues in the context of other cues available in the scene:

Equation 6.14

$$D_P = W_B D_B + (1 - W_B) D_O$$

Here, the perceived distance (D_P) is predicted by a combination of the distance estimated by binocular cues (D_B) according to their weighting (W_B) and all other available cues (D_O). This equation assumes that cues are weighted despite conflicting estimates (Muller, Brenner, & Smeets, 2009), although it has been suggested that the cue that is in conflict with all the others may be weighted less accordingly (Landy, Maloney, Johnston, & Young, 1995).

For the depth cue of interest in the present work, the more the perceived depth increases as a result of binocular gain, the more this information will not match the perceived depth being reported by other cues. Using this information, it is possible to calculate the weighting given to binocular disparity by comparing how reported depth varies with the effective IOD gain used in rendering. Were the cue to be highly weighted, the perceived depth would be expected to increase across the binocular gain conditions. This can be seen in Figure 6.11, where, as binocular gain increases, the depth reported by binocular cues increases, as shown by the black line with a slope of 1, but the depth reported by other cues does not increase, as shown by the blue line with a slope of 0. This is because increasing binocular gain should not affect depth from pictorial cues such as shading and texture, and therefore these cues should not report an increase in depth for the ideal observer.

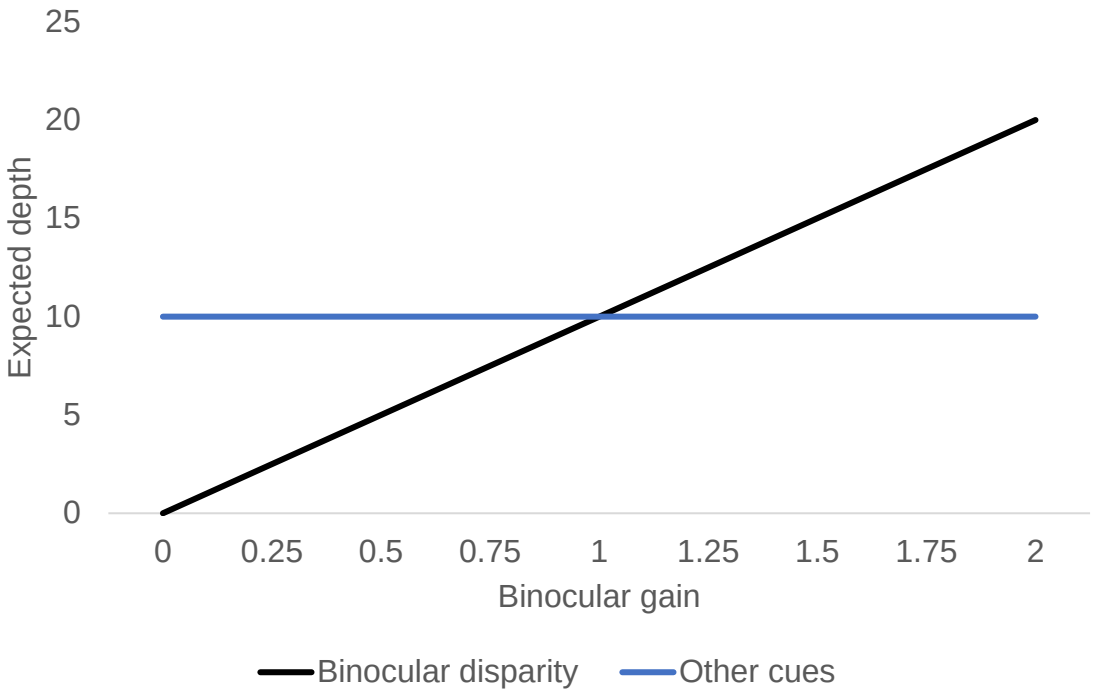


Figure 6.11: Predicted effect of gain for binocular disparity and other cues. Showing a representation of how depth may be reported by binocular disparity vs other cues for increasing binocular gain.

Manipulating the simulated IOD in this way allows for increasing binocular gain. The contribution made by binocular disparity to the final depth estimate can be calculated by manipulating the binocular gain and comparing the depth reported by observers across various gain conditions. A binocular gain of 1 represents regular vision, so that for the ideal observer there would be a one-to-one ratio of displayed depth to perceived depth. By changing the binocular gain, the perceived depth of objects can be manipulated. For instance, a binocular gain of 2 should make stimuli appear twice as deep, and a binocular gain of 0 should make objects appear shallower with effectively no disparity. The present work displayed stimuli with a varying binocular gain in increments between 0 and 2. From the literature it is expected that as binocular gain increases, so too will perceived depth (Hibbard, van

Dam, & Scarfe, 2020; Taya, 2023). The slope value can then be taken as a direct weighting of binocular disparity to the overall depth estimate.

This work presents a cue perturbation method where binocular disparity is varied within the combined depth estimate and provides evidence of using this method to measure the contribution and weighting of individual cues to the overall depth estimate in this way. It presented 3D representations of fruits and vegetables, both as a single object and as part of a scene, and recorded estimates of depth from slant and tilt measures using the gauge figure method. This allowed for exploration of the effect of increasing binocular gain on the depth reported by participants, with the change in depth range used as a metric, and the slope of which giving a direct measure of the relative weighting of binocular disparity to the overall depth estimate.

It was predicted that binocular disparity would produce a smaller weighting at a far distance due to reduced reliability. This tests a key component of cue fusion theory, and provides evidence of how this operates in natural images. The study also explored the effect of clutter by presenting objects both alone and as part of a crowded scene. Here, the prediction was the slope of gain on binocular disparity against perceived depth, and therefore the relative weighting and contribution of binocular disparity to the overall depth estimate, would be higher in the cluttered scene than for the single object. This is because the cluttered scene contained more vertical disparity information with which to scale disparity (O'Kane & Hibbard, 2007) and therefore, estimates of depth were predicted to be more veridical.

6.3 Methods

6.3.1 Participants

6.3.1.1 Experiment 1

Nine participants between the ages of 18 and 30 with normal or corrected-to-normal vision were recruited. 44% identified as female and 56% as male. The participants included one researcher, as well as eight people naïve to the purpose of the experiment.

6.3.1.2 Experiment 2

Five participants with normal or corrected-to-normal vision between the ages of 25 and 32 took part in this experiment, including one researcher and four naïve participants. 60% identified as female and 40% male. Testing was begun on a further five participants, although these datasets were not completed, as due to the length of the experiment it was completed in several sessions, and testing coincided with the beginning of the Covid19 lockdown, when physical access to the laboratory was prohibited.

6.3.1.3 Recruitment

Participants for both experiments were recruited through the University of Essex's online SONA system, as well as through word of mouth, from the University of Essex staff and student population. Some participants who were enrolled as Psychology students received course credit for their participation, while others were compensated financially.

6.3.1.4 Screening and set up

All participants were screened prior to beginning the experiment using two vision tests for normal vision and stereo acuity. Stereo Optical's Butterfly random dot depth test (Stereo Optical, 2020) was administered to screen for sufficient gross stereopsis, with the cut-off point for participation being if participants could view the entire 3D butterfly, which equated to 700 seconds of arc. This was viewed through polarised glasses at a distance of 41cm (16 inches), as per the instructions. Participants were screened for normal or corrected-to-normal vision using the Lighthouse Distance Visual Acuity Test. The cut-off point for participation was receiving a Snellen score of 32 or better, as this gave participants a visual acuity score of 90, with the Snellen ratio 20:20 being considered normal vision, which indicates a visual acuity score of 100.

Participants who did not meet the screening criteria were thanked for their time and did not participate. Those who did pass the screening then underwent set up tasks. Participants' IOD was measured as the distance between the two eyes using a standard ruler. This was measured three times and average taken to ensure accuracy. Participants' dominant eye was assessed by holding up a pen at arm's length with both eyes open and aligning it with a mark on the far wall, then alternately closing each eye to see which remained aligned (Porac & Coren, 1976).

6.3.2 Apparatus

All 3D models used in this work were created by scanning fruit and vegetables with the NextEngine 3D scanner and Multidrive turntable using eight divisions at 360-degree rotation, two tilt settings of ± 20 degrees with a resolution of 62 points per

mm^2, and the macro distance range. This captured the shape, colour and texture information of the objects with an accuracy of 0.1mm. Scans were then processed in the NextEngine Scan Studio HD software, fusing meshes with a tolerance of 0.064mm to create watertight meshes. For more details on this process please refer to Chapter 2: Methods.

The stimuli were rendered and presented using MATLAB with OpenGL and the Psychophysics Toolbox extension (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) . Stereoscopic pairs of images for the left and right eye views were rendered to cover a range of adult IODs (Dodgson, 2004) between 0x and 2x each observer's IOD by varying the inter-camera distance to simulate different IODs and accurately present disparity defined by the 3D structure in the scene.

The experiments were viewed on a 52 by 29cm monitor with a resolution of 1920 by 1080 pixels. VIEWPixx 3D synchronisation LCD shutter goggles synchronised to the 120Hz refresh rate of the screen, along with a 3DPixx IR emitter, presented a different image to each eye individually, giving a total of 60 frames per second to each eye. Participants sat with their head on a chin rest, adjusted so that the middle of the screen was at eye level for each participant, to minimise head movements during trials to eliminate additional depth cue information. Responses were recorded using a standard computer mouse, where participants would rotate the gauge figure by moving the mouse, clicking the left mouse button to confirm their judgement and move onto the next trial.

6.3.3 Stimuli

6.3.3.1 Experiment 1

The stimulus used in this experiment explored perception of a single object, which was a 3D rendered model of a pomegranate against a grey background measuring 1920 pixels high by 1080 pixels wide, shown in Figure 6.12. The scene was presented at two different physical distances of 50cm and 96cm using the track upon which the monitor was sat, with stimuli rendered for each distance, with the centre of the object at the screen distance. The apparent depth of the pomegranate was manipulated by altering the IOD gain. There were nine conditions of varying IOD gain: 0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75 and 2, with 1 being veridical binocular depth. Also pictured is the 4 by 8 grid used to outline the 32 possible positions of the gauge figures spaced 30 pixels apart, although this grid was not visible during the experiment. The gauge figures used in this experiment were white with a base diameter of 26 pixels and locations were sampled three times each to get an average.

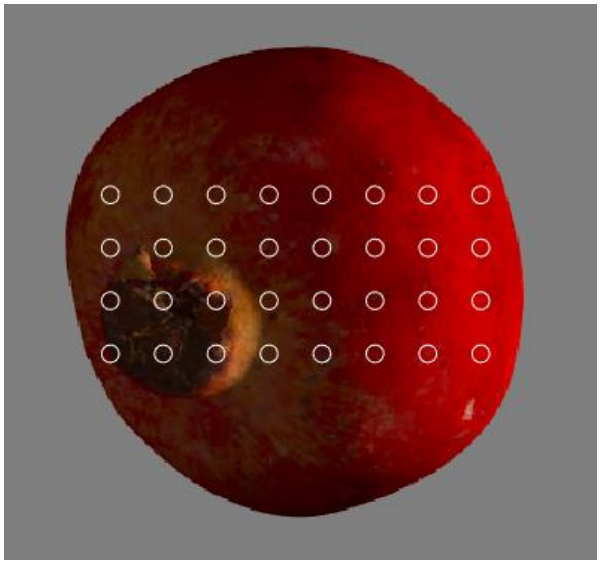


Figure 6.12: Pomegranate with gauge locations. Showing the pomegranate stimulus with point grid, not pictured in the experiment, used to outline the possible locations of the gauge figures presented to participants.

Lighting in the scene was applied with OpenGL in MATLAB, with ambient and diffuse lighting components each with a magnitude of (0.7 0.7 0.7). The light source was a spotlight directed at the centre of the scene and offset 50cm to the right.

6.3.3.2 Experiment 2

The stimuli used in this experiment was a cluttered scene of five rendered 3D models of fruits and vegetables against a grey background (see Figure 6.13). The scene was presented at two physical distances of 50cm and 96cm as before by moving the monitor on the track. Also pictured is the possible location of gauge figures for Experiment 2 at 96cm which correspond to points on either the sweetcorn cob or the scone. This grid was not visible during the actual experiment. Each gauge figure location was tested three times during each block and trial locations were randomised. Due to scaling of the objects with distance, blocks at the closer distance of 50cm had a total of 65 gauge figure locations, giving 195 trials per block, and the blocks at the further distance of 96cm had 57 gauge locations giving 171 trials.

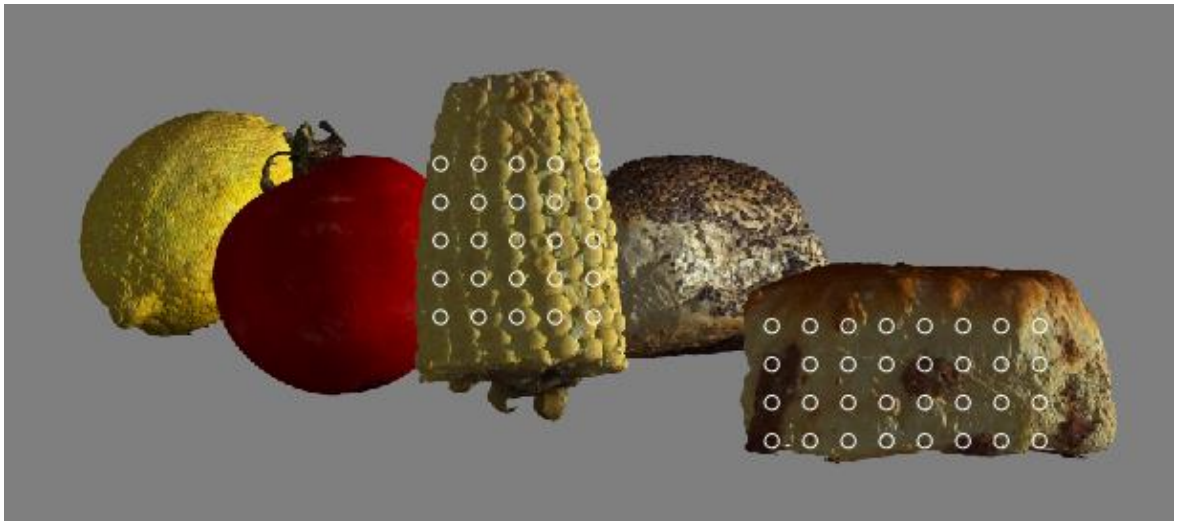


Figure 6.13: Cluttered scene with gauge locations. Showing the cluttered scene of five 3D models of food from scanned objects, including gauge figures locations for the two objects at 96cm.

Again, the apparent depth of the objects was manipulated by altering the IOD gain, with five conditions of varying IOD gain: 0, 0.5, 1, 1.5, and 2, with 1 being veridical binocular depth. As before, the gauge figures used in this experiment were white with a base diameter of 26 pixels.

6.3.4 Procedure

6.3.4.1 Experiment 1

After the screening and set up phase, participants were seated with their head in a headrest and the lab was darkened meaning the only light source in the room came from the monitor. They were presented with a 3D model of a pomegranate and were asked to define the orientation of the surface of the object using a gauge figure. The pomegranate was presented at two different physical distances of 50cm and 96cm by way of a moveable monitor. Participants were instructed to use the mouse to rotate the gauge figure such that the circular element appeared to be painted onto the surface of the object (Koenderink, van Doorn, & Kappers, 1995), with the stick portion of the gauge figure pointing outwards, perpendicular to the surface. Once participants were happy with the placement of the gauge figure, they clicked the left mouse button to place it and the next trial began. Having completed all trials, they moved on to the next block. There were nine differing IOD gain conditions and across the two distances, giving a total of 18 blocks. Participants had breaks in between each block, during which the room was once again illuminated to reduce extra cue information from darkness adaptation. All trials and blocks were randomised by participant. Once all blocks had been completed, participants were debriefed.

6.3.4.2 Experiment 2

The procedure was much the same as in Experiment 1, but with a few adjustments. Gauge locations were randomised between the two objects within blocks. There were five binocular gain conditions as well as a monocular condition. These were all tested at both distances of 50cm and 96cm, giving a total of 12 blocks.

6.4 Results

6.4.1 Data treatment

Raw data for both experiments were the slant (phi) and tilt (theta) values from the settings made by observers with the gauge figure, with settings related to their Cartesian x and y positions in the scene. These were first reordered having been presented to observers in a random order. As each gauge figure location was tested three times within a block, a mean of these was taken. The gradient of the angle between the average perceived setting and the line of sight was calculated as per the methods outlined by Nefs (2008) discussed previously. These provided the magnitude and orientation of the surfaces perceived from local settings. Following on from Nefs' method of creating depth gradients, these local settings were converted into a mesh grid of the global scene percept as a best fit 3D mesh model. These 3D meshes were visualised to ensure sensible data and a good fit from mesh reconstruction, an example of which can be seen in Figure 6.14, which shows plots ranging from a gain of 0 in 0.25 increments through to a gain of 2 for one participant in Experiment 1.

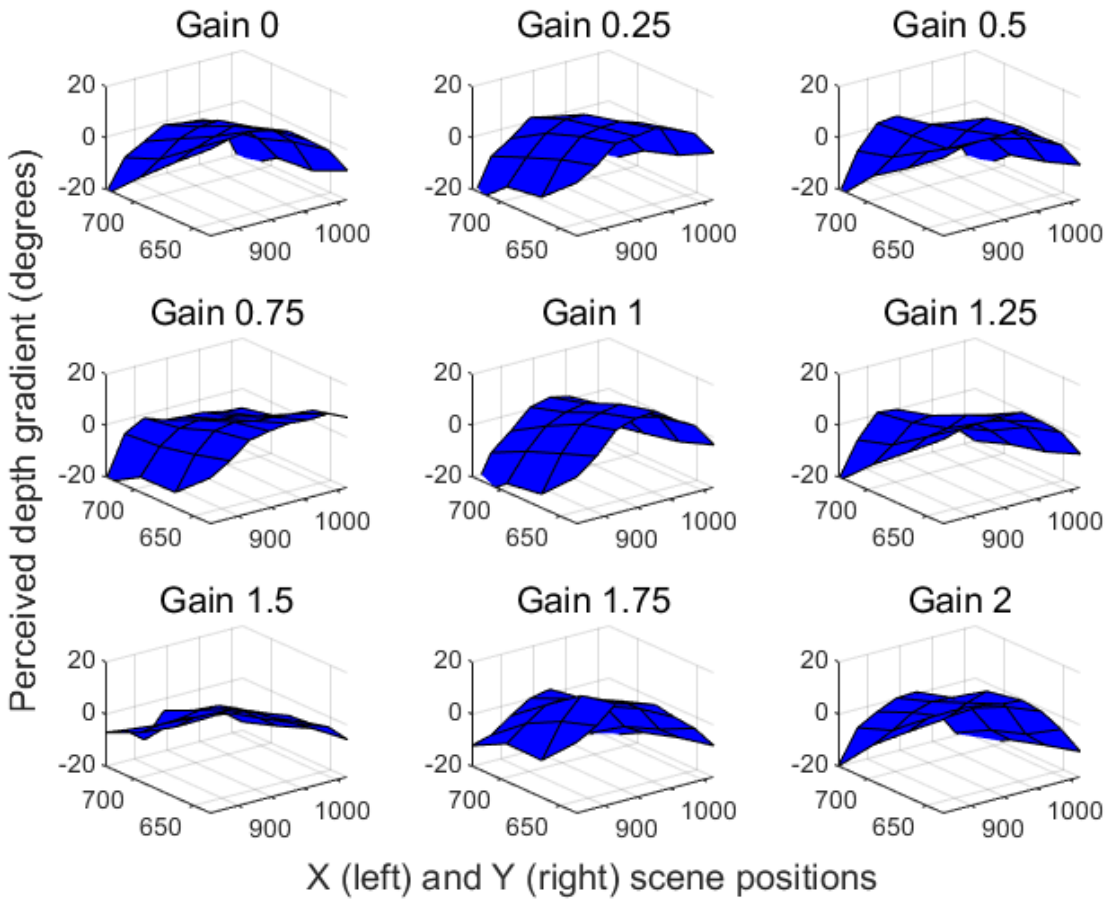


Figure 6.14: Reconstructed meshes Experiment 1. Example reconstructed mesh surfaces for Experiment 1 for one participant, showing the perceived depth gradient (in degrees) of the mesh surface for each gain.

6.4.2 Experiment 1

To begin with, observers' perception from the reconstructed meshes was compared to the objective ground truth from the original 3D model used to render the stimuli scenes to see if observers made sensible judgements The relationship between observers' settings and the ground truth was explored with multiple linear regression as outlined in Chapter 2: Methods. Here, the gradients set by observers were compared to the gradients from the original 3D model used to render the scene. This found that observers' settings for depth gradients were strongly predicted by the depth gradients of the original 3D model at both distances, showing

highly accurate judgements, with an average regression slope value of 0.920 at the closer distance and a slope of 0.979 at the further distance, where a slope of 1 would indicate perfect performance.

Continuing on from this, Figure 6.15 shows the mean gradients set by all observers and highlights gauge sample points with a bigger range of slopes on average compared to the congruent ground truth. From this, several outlier points can be clearly seen, each marked with their corresponding point sampling number. The increased slope estimates compared to the ground truth at these points indicates that on average observers overestimated the gradients of these points compared to the ground truth.

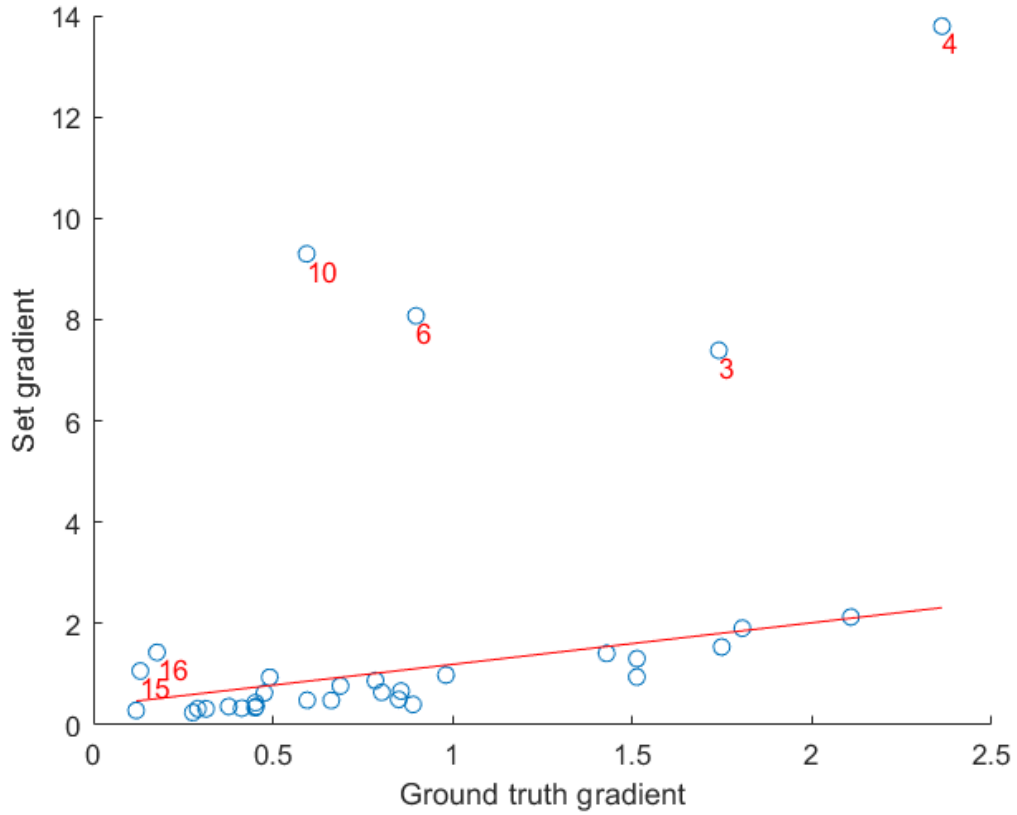


Figure 6.15: Gradient point comparison. Mean of gradients set by all observers are plotted against the ground truth gradients. Points with a high set gradient to ground truth gradient ratio are shown with their corresponding number which identifies them

in the stimuli image below. The red line shows the regression slope of ground truth against set gradient, inclusive of outliers.

Figure 6.16 shows the pomegranate with the sampling point numbers overlaid. Comparing this with the points highlighted above, it is clear to see that the overestimated points at 3, 4, 6. 10, 15 and 16 surround the stalk of the pomegranate. These points suggest relatively large gradient errors over a small area of the model, focused around the stalk of the pomegranate. This is further considered in the discussion.

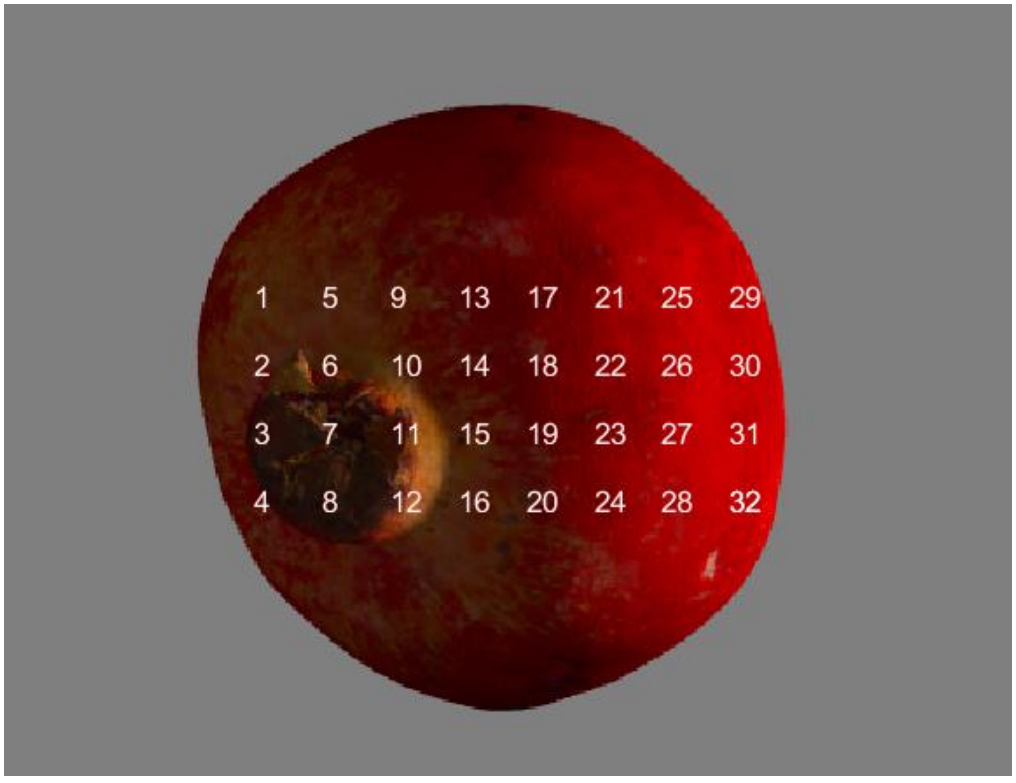


Figure 6.16: Sampling points Experiment 1. These show which point numbers correspond to which part of the stimulus surface.

Here the analysis begins to focus on the main prediction that binocular disparity will have a smaller weighting at the further distance. As observers completed multiple blocks of varying IOD gain at two distances, a linear mixed effects model

(LME) was used to account for repeated measures design. Here, the model tried to predict the set gradient (s) from the physical gradient (p), the binocular gain (g) and the viewing distances (d), and interactions between these, grouped by observer (o):

Equation 6.15

$$s \sim 1 + p * g * d + (1 + p | o)$$

Physical gradient from the ground truth was found to significantly predict the set gradient, mirroring the above findings that observers made sensible judgements. Binocular gain was also found to significantly affect the gradients set by observers, with a positive estimate here showing that as binocular gain increased so did the gradients set, which shows increased depth with gain as predicted. Viewing distance was also found to significantly affect the set gradients. A setting gain for each condition was calculated as the ratio of set gradients to physical gradients, as shown in Figure 6.17.

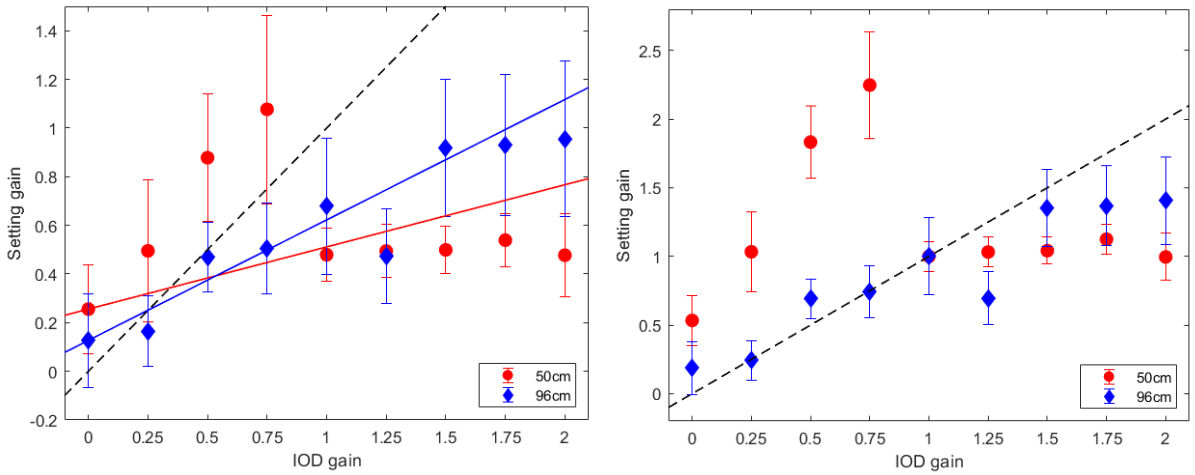


Figure 6.17: Setting gain against IOD gain. Plot on the left shows mean (SE) setting gain, which here shows the ratio of the set gradient to the physical gradient, for both distances. Lines show the respective slopes for distance, and the black dashed line shows predicted performance for binocular cues only. Plot on the right is normalised to a gain of 1.

A significant negative interaction was found between physical gradient and binocular gain, meaning that the effect of binocular gain differed between physical gradients. Additionally, a negative interaction between physical gradient and viewing distance was observed. Binocular gain was also found to significantly interact with viewing distance, which can be seen in Figure 6.17, where the effect of gain differs between the 50cm and 96cm viewing conditions. Finally, a three-way interaction between physical gradient, binocular gain and viewing distance was found, meaning that the pattern of results differed between all conditions. Full LME results are shown in Table 6-1.

Table 6-1: LME results for predicting set gradient Experiment 1.

Predictor	**Estimate**	**SE**	**DF**	***p* Value**	**Lower CI**	**Upper CI**
Physical gradient (p)	3.33	0.989	**5176**	**<.001*** **	1.40	5.27
Binocular gain (g)	2.00	0.741	**5176**	**.007****	0.55	3.45
Viewing distance (d)	2.26	0.558	**5176**	**<.001*** **	1.17	3.35
Interaction: physical gradient and binocular gain ($p * g$)	-2.47	0.713	**5176**	**<.001*** **	-3.87	-1.08
Interaction: physical gradient and viewing distance ($p * d$)	-1.93	0.536	**5176**	**<.001*** **	-2.98	-0.88
Interaction: binocular gain and viewing distance ($g * d$)	-1.74	0.469	**5176**	**<.001*** **	-2.66	-0.82
Interaction: physical gradient and binocular gain and viewing distance ($p * g * d$)	1.99	0.451	**5176**	**<.001*** **	1.10	2.87

An interesting effect can be seen for the difference between the effect of gain on the close and far distances. The ratio between the set gradients and the physical gradients can be seen to steadily increase at 96cm. As these are all below 1, they represent a consistent underestimation of depth at the further distance, consistent with the findings of Chapter 3 and 4 where depth is underestimated at far distances. However, the same pattern is not observed for the closer distance. At 50cm, reducing gains below a gain of 1 reduces the set gradient to physical gradient ratio from an average close to 1 at a gain of 1 showing accurate depth perception for this gain at this distance, although no additional depth is set with increasing IOD beyond the observer's own.

Returning to the reconstructed meshes, the ranges of depths, calculated as the difference in depth between the closest and furthest points in the reconstructed meshes. These depth ranges were then normalised relative to a binocular gain of 1, representing veridical depth where binocular cues should be consistent with pictorial cues to create a relative depth range. This shows the change in perceived depth relative to veridical binocular viewing. If observers were to only use binocular cues, depth range scores should match their corresponding gains, meaning for example that the gain of 0.5 should have a range of 0.5 compared to a gain of 1. Regression analysis was run to estimate slope values which were used as a direct measure of the weighting of binocular cues. As this work is interested in calculating the contribution of binocular cues at varying distances, we calculated relative depth ranges at both 50cm and 96cm, with results shown below.

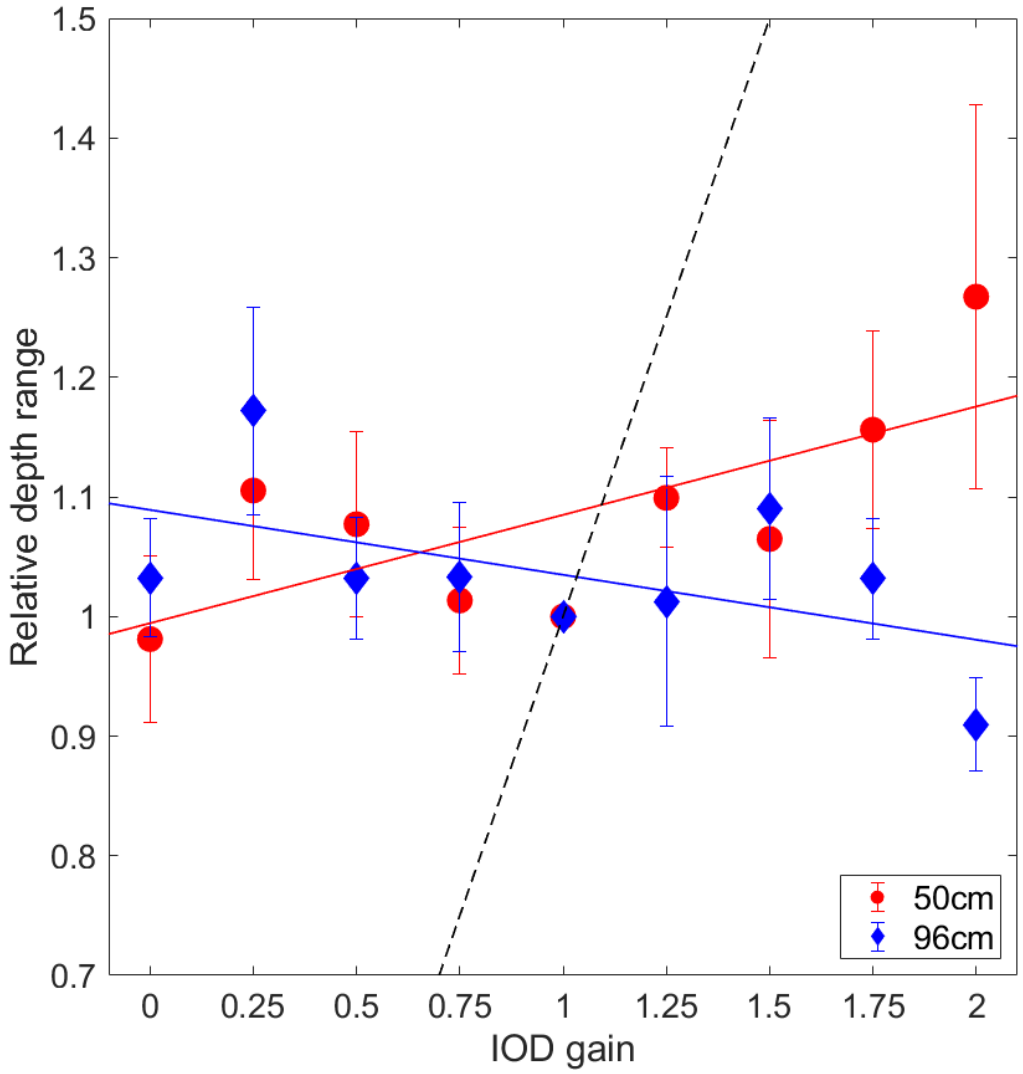


Figure 6.18: Relative depth range Experiment 1. Showing mean (SE) relative depth ranges for each condition. The red line shows regression slope of perceived depth range at 50cm, and the blue line shows regression slope for 96cm, with the dotted line showing expected slope for the ideal observer if binocular cues were weighted 100%.

Figure 6.18 shows the change in relative depth range across IOD gain conditions for both distances. The dotted black line shows a slope of 1, which shows expected performance if binocular cues were to be weighted 100% in the overall depth estimate. The red line shows a positive slope of 0.09, showing that at 50cm, the relative depth range did increase with increasing binocular gain as predicted (p=.010, 95% CI [0.02, 0.16]). As the slope value can be taken as a direct measure

of weighting, we can see that binocular cues account for around 9% of the overall depth estimate in naturalistic scenes at closer distances.

At the further distance of 96cm there was no significant effect of increasing IOD gain, although it does trend negatively with a slope of -0.05 (p=.161, 95% CI [-0.13, 0.02]). Taken with the result above, this shows that binocular disparity is weighted more in the weighted average and therefore more reliable at closer distances compared to further distances as predicted. This result also shows that people rely on pictorial cues more heavily than binocular disparity at further distances.

The final analysis looked at if the global shape perceived by observers changed with gain and viewing distance. The reconstructed meshes were compared against the gain of 1 representing the original presented stimuli through the process of affine transformation, further details of which are covered in Chapter 2: Methods. Regression analysis was used to see if the x, y and z positions as well as the intercept from the reference mesh could predict the depth in the target mesh. Transformations for the four parameters are shown in Figure 6.19.

The black dashed line shows predicted performance if only binocular cues are considered in the depth estimate. For depth, this creates a slope of 1. It can be seen that the effect of gain does not reach this slope value. Interestingly, the pattern of results mirrors what was found above for the setting gain, in that there is a clear disparity between the effect of squashing IOD from 1 down to 0, and stretching IOD from 1 to 2.

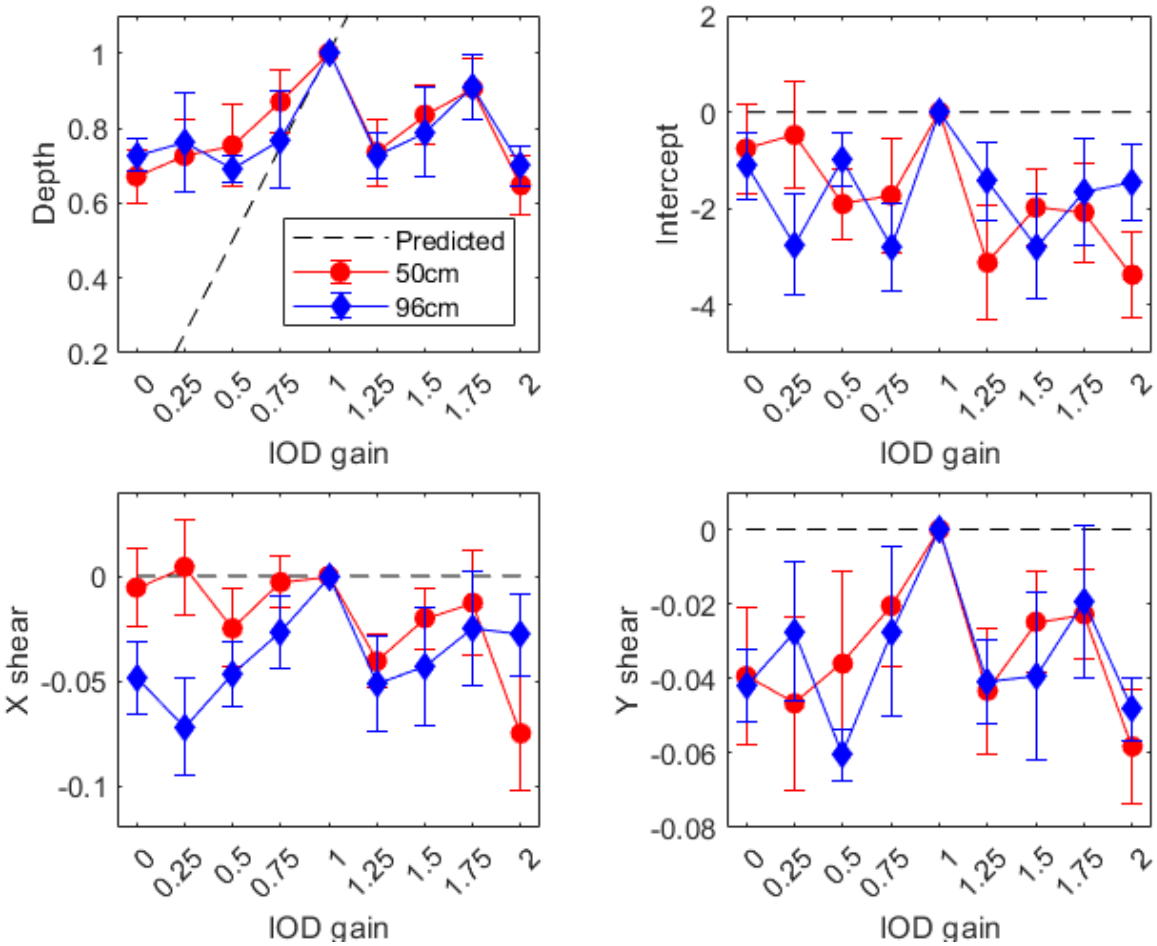


Figure 6.19: Mean affine transformations Experiment 1. Plots show the mean (SE) change in depth, intercept term, and x and y shearing between mesh comparisons. Black dashed line shows predicted performance if only binocular cues are considered in the estimate.

Next, an LME explored if these four parameters varied significantly with binocular gain, distance, or an interaction of the two using the following equation. The model selected as the best fit for all four parameters was the random intercepts and slopes of distance model:

Equation 6.16

$$b \sim 1 + d * g + (1 + d|o)$$

Here, the model was run for each of the four parameters to see if these were predicted by the manipulation of gain (g), viewing distance (d), or an interaction of the two ($d * g$), grouped by observer (o). LME results are shown in Table 6-2.

Table 6-2: LME results for affine transformation Experiment 1.

Coeff - icient	Model	Predict or	Estimat e	DF	*p* Value	Lower CI	Uppe r CI
bz	$bz \sim 1 + d * g + (1 + d \mid o)$	d	-0.009	158	.920	-0.178	0.160
		g	0.032	158	.439	-0.049	0.112
		$d * g$	0.001	158	.993	-0.114	0.115
b0	$b0 \sim 1 + d * g + (1 + d \mid o)$	d	-1.056	158	.255	-2.882	0.769
		g	-1.121	158	**.002****	-1.817	-0.425
		$d * g$	1.104	158	**.028***	0.120	2.088
bx	$bx \sim 1 + d * g + (1 + d \mid o)$	d	-0.056	158	**.010***	-0.098	-0.013
		g	-0.024	158	**.006****	-0.041	-0.007
		$d * g$	0.038	158	**.002****	0.014	0.062
by	$by \sim 1 + d * g + (1 + d \mid o)$	d	-0.004	158	.821	-0.036	0.028
		g	-0.0003	158	.968	-0.014	0.013
		$d * g$	0.002	158	.823	-0.017	0.022

We found evidence of a negative effect on shearing in the x dimension for both distance and gain, as well as an effect of gain on the intercept, which reflects the starting depth of the reconstructed models differing with gain, interacting with distance. However, no main effect of gain or distance was found for the depth

parameter. Given the pattern of findings from Figure 6.19, post hoc analysis was conducted to explore this pattern further.

The affine transformations were split into two categories; gains 0 to 1 were collated representing where IOD had been squashed (sqz), and gains 1 to 2 were collated representing where IOD had been stretched (stz). The same LME model was used as above for consistency. Results are shown in Table 6-3:

Table 6-3: LME results for squashed and stretched IOD affine transformation Experiment 1.

IODs (gains)	Model	Predictor	Estimate	DF	*p* Value	Lower CI	Upper CI
Squashed (0 – 1)	$sqz \sim 1 + d * g + (1 + d \mid o)$	d	0.036	86	.687	-0.142	0.214
		g	0.321	86	**.003****	0.115	0.527
		$d * g$	-0.102	86	.490	-0.393	0.189
Stretched (1 – 2)	$stz \sim 1 + d * g + (1 + d \mid o)$	d	-0.070	86	.744	-0.494	0.354
		g	-0.214	86	**.032***	-0.409	-0.019
		$d * g$	0.047	86	.738	-0.229	0.322

From this, we find that reducing gains below a gain of 1, that is squashing an observer's IOD, results in a consistent reduction of perceived depth. This follows the predictions of this work. However, we also found that increasing IOD gain from 1 to 2, that is stretching the IOD of observers, also resulted in a reduction in perceived depth from the reconstructed meshes. This is contrary to our predictions and the potential reasons for this will be covered in the discussion.

6.4.3 Experiment 2

Meshes were reconstructed as before and the data visualised. Figure 6.20 shows examples from one participant which show sensible settings as reconstructed meshes have a smooth cohesive surface.

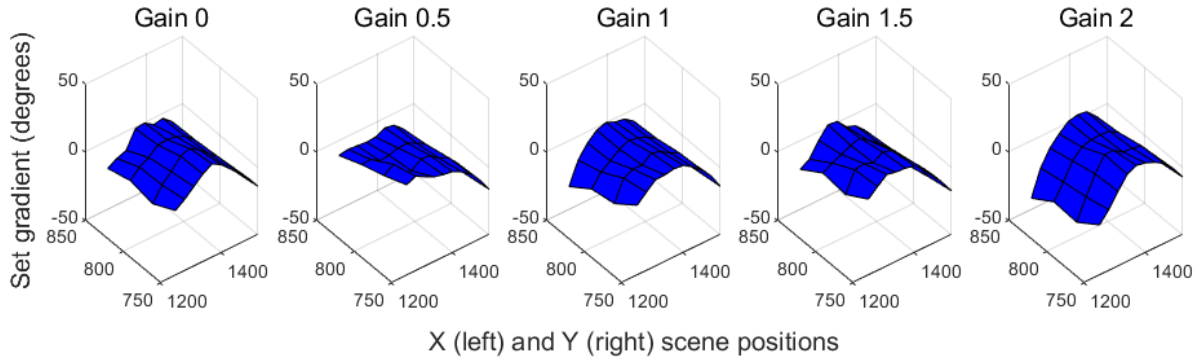


Figure 6.20: Reconstructed meshes Experiment 2. Plots show example reconstructed mesh data per gain for one participant at one distance for one object.

As before, depth gradient ranges were normalised relative to a gain of 1 to evaluate change in the depth range with increasing gain on binocular disparity. A linear mixed-effects model was fit to these ranges to test the relationship between the reported surface orientation (r) and the viewing distance (d), binocular gain (g), and an interaction of the two ($d * g$), with the grouping variable of observer (o) as seen in Equation 6.17:

Equation 6.17

$$r \sim 1 + d * g + (1 + d + g \mid o)$$

This model contains random intercepts and slopes for distance and gain which was the best fit to the data as seen by the lowest AIC value. The LME results are shown in Table 6-4.

Table 6-4: LME results for predicting reported depth Experiment 2.

Predictor	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
Viewing distance (d)	-0.117	0.077	96	.133	-0.269	0.036
Binocular gain (g)	0.100	0.047	96	**.034***	0.008	0.193
Interaction ($d * g$)	-0.007	0.044	96	.874	-0.094	0.080

A significant overall effect of gain on relative depth was found, showing that depth increased by 10% with increasing binocular gain, which supports the findings of Experiment 1. Estimates show binocular cues account for 10% of the weighting at closer distances (p=.048, 95% CI [0.001, 0.200]), and 9% at further distances (p=.023, 95% CI [0.010, 0.177]), showing a slight reduction in weighting with increasing viewing distance as predicted, although overall, results show no significant effect of viewing distance on the relative depth, and no interaction with gain, which goes against findings from Experiment 1 and our predictions that increasing viewing distance should reduce the weight of binocular cues and therefore reduce the effect of gain. Means (SE) are plotted in Figure 6.21.

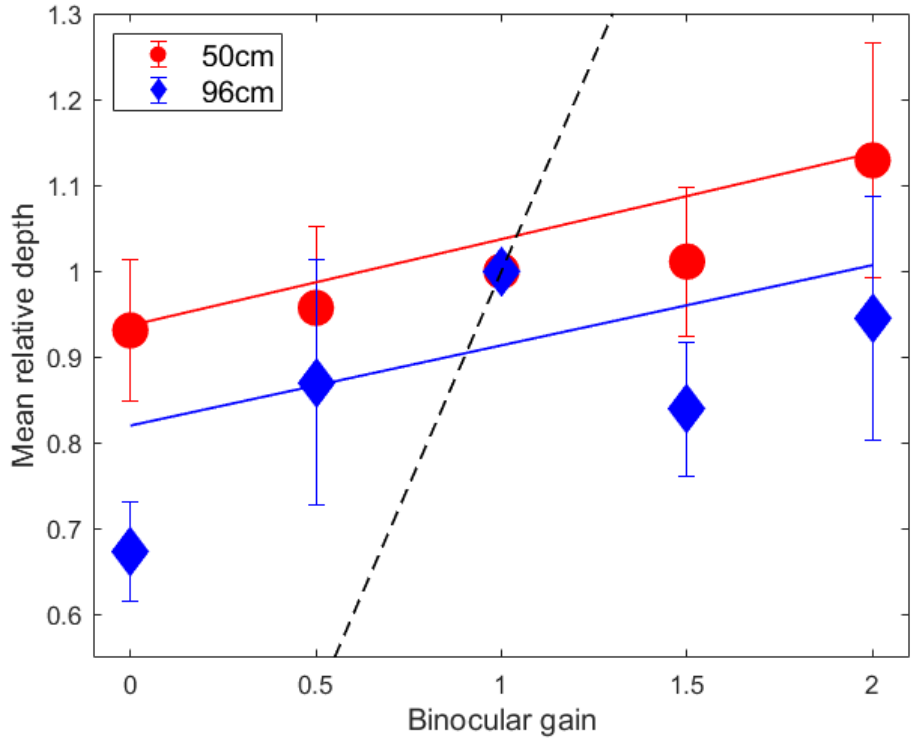


Figure 6.21: Mean relative depth Experiment 2. Plot shows mean (SE) relative depth slope values Experiment 2 for both distances. Black dashed line shows predicted effect of gain for binocular cues only.

However, as observed in Experiment 1, there is a clear difference in the effect of gain for squashed and stretched IODs. As before, the gains were split to analyse the effect of squashing and stretching the gain on binocular cues. The mean (SE) relative depths split by squashed or stretched gains are shown in Figure 6.22.

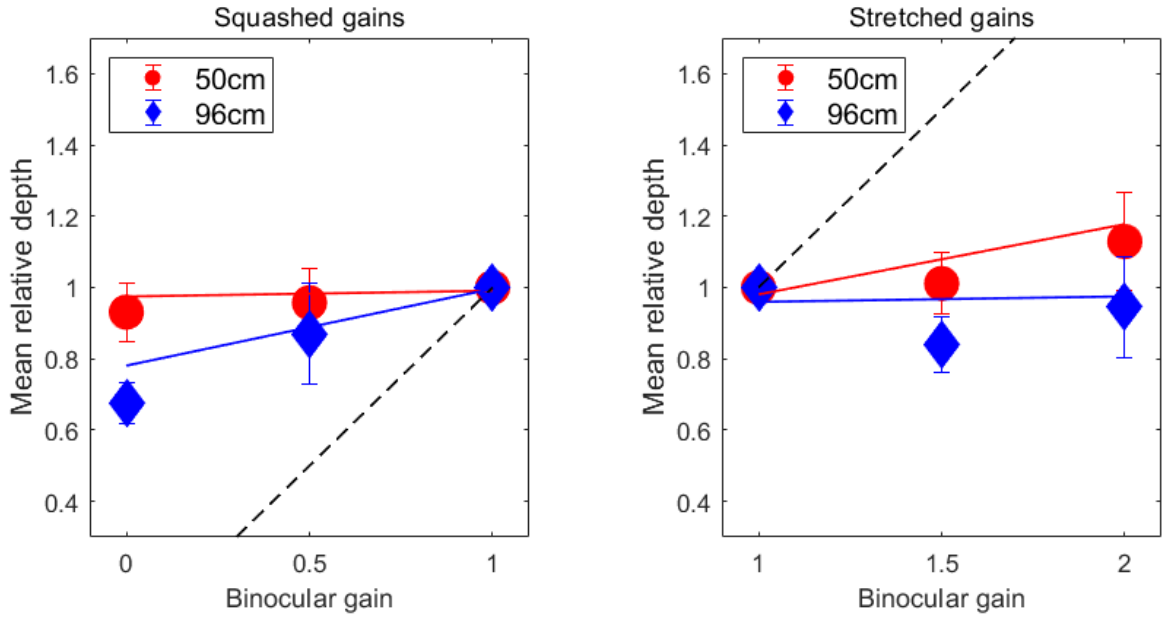


Figure 6.22: Squashed and stretched gain comparison Experiment 2. Showing mean (SE) relative depth for each gain, split into squashed or stretched. Black dashed line shows predicted performance for only binocular cues. Red and blue lines show regression slope values for 50cm and 96cm distances respectively.

The effect of the squashed and stretched gains was analysed using the same LME as in Equation 6.17. Results are shown in Table 6-5:

Table 6-5: LME results for squashed and stretched gains Experiment 2.

IODs (gains)	Model	Predictor	Estimate	DF	*p* Value	Lower CI	Upper CI
Squashed (0 – 1)	$sqr \sim 1 + d * g + (1 + d + g\|c$	d	-0.194	56	**.020***	-0.357	-0.031
		g	0.016	56	.822	-0.127	0.159
		$d * g$	0.200	56	**.023***	0.029	0.370
Stretched (1 – 2)	$str \sim 1 + d * g + (1 + d + g\|c$	d	0.161	56	.208	-0.092	0.413
		g	0.197	56	.086	-0.029	0.423
		$d * g$	-0.182	56	**.025***	-0.340	-0.024

Here, a significant effect of distance on relative depth was found, as well as a significant interaction with gain. This shows that the effect of gain is dependent on the viewing distance, with a 20% increased effect found at the far distance, which is contrary to predictions and the findings in Experiment 1. A significant interaction between distance and gain on relative depth was also found for the stretched gains. Here, the effect of gain is 18% stronger at the closer distance, in keeping with predictions and findings from Experiment 1. However, the confidence intervals denote a wide range, likely due to small sample sizes, and therefore these results warrant further investigation with a larger sample.

Next, the change in global shape was assessed with affine transformations as before. The mean (SE) transformations are shown in Figure 6.23.

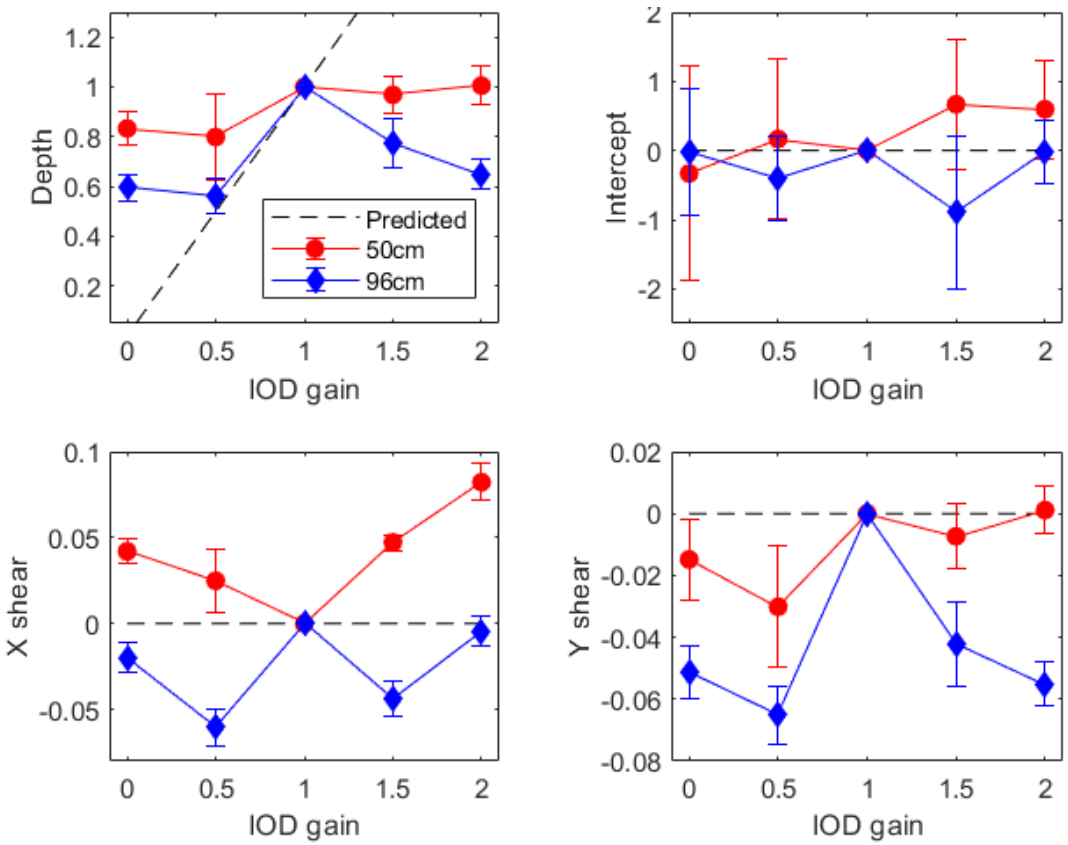


Figure 6.23: Affine transformations Experiment 2. Plots compare mean (SE) depth, x and y shearing and intercept transformations from affine regression for both distances. Black dashed line shows predicted performance.

An LME looked at the effect of gain, distance and the interaction of the two on the four parameters of depth, x and y shearing and the intercept. The model chosen as the best fit of the data was the random intercepts model, as this was the best fit for the depth parameter as the main interest of this work. LME results are shown in Table 6-6.

Table 6-6: LME results for affine transformations Experiment 2.

Coeff - icient	Model	Predic tor	Estimat e	SE	DF	*p* Value	Lower CI	Upper CI
bz	$bz \sim 1 + d * g + (1\|o)$	d	-0.167	0.163	96	.095	-0.364	0.029
		g	0.141	0.099	96	.272	-0.112	0.395
		$d * g$	-0.039	0.081	96	.633	-0.199	0.122
b0	$b0 \sim 1 + d * g + (1\|o)$	d	0.079	0.795	96	.921	-1.500	1.657
		g	1.042	1.027	96	.312	-0.995	3.080
		$d * g$	-0.570	0.649	96	.382	-1.859	0.719
bx	$bx \sim 1 + d * g + (1\|o)$	d	-0.054	0.022	96	**.015***	-0.097	-0.011
		g	0.032	0.028	96	.261	-0.024	0.088
		$d * g$	-0.011	0.018	96	.530	-0.046	0.024
by	$by \sim 1 + d * g + (1\|o)$	d	-0.025	0.018	96	.168	-0.060	0.011
		g	0.019	0.023	96	.413	-0.027	0.064
		$d * g$	-0.007	0.014	96	.588	-0.037	0.021

Once again there was evidence of shearing in the x dimension, which reduced with increasing viewing distance. As before, no overall effect of gain on depth transformation was observed, but the graphs show the same pattern as before, so the gains were once again split into squashed and stretched gains for analysis. The

effect of the squashed and stretched gains was analysed using the same LME as before. Results are shown in Table 6-7:

Table 6-7: LME results for squashed and stretched gains Experiment 2.

IODs (gains)	Model	Predictor	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
Squashed (0 – 1)	$sqr \sim 1 + d * g + (1\|o)$	d	-0.279	0.125	56	**.029***	-0.529	-0.029
		g	0.167	0.137	56	.228	-0.107	0.440
		$d * g$	0.239	0.193	56	.221	-0.148	0.626
Stretched (1 – 2)	$str \sim 1 + d * g + (1\|o)$	d	0.352	0.212	56	.103	-0.073	0.774
		g	0.005	0.097	56	.956	-0.188	0.199
		$d * g$	-0.357	0.137	56	**.011***	-0.631	-0.084

As before, a negative effect of distance was observed for the squashed gains, meaning that depth transformations were greater for the further distance relative to a gain of 1. Likewise, a significant negative interaction between distance and gain was observed for the stretched gains. This suggests that the effect of gain is reduced with increasing viewing distance, which supports predictions that binocular cues are weighted less and therefore the effect of gain is reduced at further distances.

Finally, the results of the monocular and binocular conditions were compared using linear regression. The reported depth in the gain 1 binocular condition was compared against the reported depth in the monocular condition. No significant difference was found between them (R^2= 0.043 *p*= .395, 95% CI [-0.058, 0.145]) and there was no significant effect of distance on relative depth either (R^2= -0.029 *p*= .563, 95% CI [-0.131, 0.072]), which shows that the cues were consistent for the amount of perceived depth relative to a gain of 1.

6.5 Discussion

6.5.1 Experiment 1

As participants were highly accurate overall in this task as shown by the set gradients largely matching with the objective ground truth gradients, and physical gradient significantly predicting set gradient in the LME, we can be confident that observers were successfully perceiving the binocular IOD manipulation as a cohesive surface, and therefore the gain on binocular cues worked as expected.

While ground truth gradients and those set by observers was largely congruent, showing accurate judgements, there were several specific points where the gradient was consistently overestimated between conditions and observers. Further investigation revealed these points to be all those surrounding the stalk of the pomegranate but not directly on it. The extreme gradients here are inflated by a ceiling effect of the maximum gradient angle to which a restriction was set during analysis. These consistent overestimations at this landmark on the object suggest that observers were perceiving the gauge as resting on the stalk itself, and therefore setting the gauge to an angle closer to 90 degrees, whereas the point itself in actual fact resided on the surface of fruit, requiring an angle closer to 0 degrees to create the ‘bullseye’ gauge effect.

Additionally, the size of the gauge figure may have impeded accurate surface capture. In Figure 6.24 the experiment is pictured as if from a bird’s eye view to illustrate the effect.

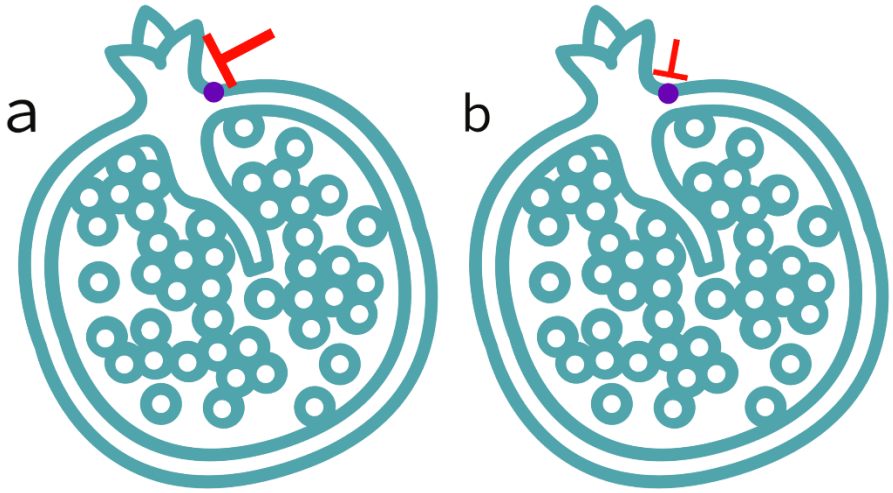


Figure 6.24: Gauge size consideration. Experiment pictured as if from a bird's eye view to illustrate. Here, the potential setting from a bigger gauge (a) is compared to the potential gradient setting from a smaller gauge (b).

In Figure 6.24a, a larger size gauge on a point so close to the stalk may have caused observers to set the gauge to avoid 'clipping' the gauge figure into the stalk, that is, that the diameter of the circle element caused observers to set great gradients to avoid touching this object landmark. This is because observers are asked to set the gauge so that it appears flat to the object. The larger the surface area of the gauge circle, the larger the area observers will be required to estimate across. Here, the sharp gradients of the stalk protruding out towards the observer may have affected the set gradients. This finding mirrors the importance of gauge size and setting distance found in Chapter 5. Future studies with protruding object landmarks such as the pomegranate stalk should consider size of gauge figure and the location of points carefully to avoid this effect, as shown in Figure 6.24b, although this does not affect the overall variables of interest in this work.

Binocular gain was also found to increase set gradients, showing increased depth, as predicted. Viewing distance was found to increase the gradient set, consistent with findings from Chapters 3 and 4 that show that space at far distances

is underestimated and therefore depths are set too big to compensate. This is also seen by the setting gains, which compare the ratio of set gradients to physical gradients, being consistently below 1 for the further viewing distance.

Considering the three-way interaction between physical gradient, viewing distance and binocular gain, results are consistent with our predictions that perceived depth will change to differing extents between conditions. The observed negative relationship between the effect of physical gradient and binocular gain is consistent with the gain manipulation, in that physical gradients became less able to predict set gradients with increasing gain, which is due to the increased perceived depth. Likewise, the negative interaction between the effect of physical gradients and viewing distance can be explained by the increasing misestimation of depth with viewing distance, meaning a bigger difference between physical and set gradient with increasing viewing distance. The interaction between the effects of gain and viewing distance is consistent with the prediction that binocular cues are weighted more heavily at closer distances, as these results show less effect of gain with increasing viewing distance, showing that binocular cues are weighted less at further distances due to them becoming less reliable.

When comparing overall global depth ranges of the reconstructed meshes, increasing binocular gain was found to have a significant effect on the depth reported, accounting for around 9% of the weighted depth estimate in the natural scene at the closer distance as predicted. This shows that binocular disparity is not the only cue that is being listened to in the overall weighted average, but that it is a reliable cue at this distance as manipulating it with binocular gain did produce a significant difference. The lack of significant difference between binocular gain

conditions found at the further distance displays that binocular disparity is weighted more heavily and is therefore more reliable at closer distances.

Finally, the affine transformations explored the global shape change between conditions. Significant negative shearing in the x dimension was observed, as well as a significant shift in the intercept, denoting the starting point of the reconstructed meshes, indicating that objects were perceived as being smaller and further away with increasing gain, showing a lack of shape constancy which supports findings from Chapter 3.

Affine transformations found no overall effect of gain on the perceived depth. However, when gains were analysed separately, an effect of gain on increased depth was observed for gains from 0 to 1, meaning that 'squashing' observers' IODs resulted in less depth perceived as predicted. Additionally, a negative effect of gain on depth was found for gains from 1 to 2, meaning that 'stretching' IOD beyond the observer's own resulted in consistently less depth, contrary to predictions. This is likely due to issues surrounding observers ability to effectively fixate on the stretched IOD stimulus. The nearest point that people can comfortably converge the eyes differs between people but is considered to be around 6 to 10cm, and is known as the near point of convergence or NPC (Stidwell & Fletcher, 2011). This would be most relevant at the closer distance due to the decreased distance between observer and stimulus. While we screened observers for sufficient gross stereopsis, we did not measure the NPC, and therefore cannot rule out that some observers were unable to successfully converge on stimuli at the higher gains. Additionally, prolonged close convergence can be difficult for those with convergence infacility, which describes the lack of ability to repeatedly converge, such as here making

repeated settings for the larger gains. Future work could address this by measuring observers' NPC as part of the screening process.

6.5.2 Experiment 2

Experiment 2 builds on the findings here by addressing two key areas. A natural expansion for this work was the addition of more 3D models of natural objects. Experiment 1 provides a good account of the contribution of binocular disparity on depth judgements of natural objects, and Experiment 2 builds on this by including additional models in the scene to add additional depth information and provide more true-to-life viewing conditions. Another improvement on Experiment 1 was to provide natural objects that do not typically conform to a geometric shape. The pomegranate, as a recognisable fruit, has a predictable roundness, given that it is typically spherical. As such, this may have encouraged participants to rely on prior knowledge of the shape of a pomegranate, rather than purely setting the surface as perceived during the task. Experiment 2 follows up on this idea by utilising natural objects that do not have to conform to these geometric conventions, in this case baked goods.

As predicted, analysis shows that there was a significant effect of increasing binocular gain on depth estimates, which accounted for around 10% of the weighted depth estimate at the closer distance, and around 9% at the further distance. This supports the findings of the previous experiment which measured the binocular gain contribution to be around 9% also. The addition of other cues as part of the cluttered scene increased binocular weighting slightly between the single object and scene experiments, consistent with findings of previous chapters where additional cluttered

scene cue information provides more accurate depth estimates. As predicted, binocular weighting was reduced at the further distance, in line with the findings of Experiment 1, although this time the effect of gain was found to be significant.

Interestingly, we observed the same polarised pattern of results for both experiments in that the 'squashed' gains had consistently different effects of gain and distance than the 'stretched' gains. In Experiment 1, a positive effect of squashed gains was found, with a negative effect of stretched gains, meaning that increase IOD from 0 to 1 resulted in increasing relative depth, but beyond observers' own IOD, increasing gain resulted in a reduction of relative depth. For Experiment 2, a positive interaction for squashed gains was found, indicating a stronger effect with increasing viewing distance and a negative interaction for stretched gains suggesting a reduction in effect with increasing viewing distance. Gains from 0 to 1 provide clear evidence of binocular cue weighting manipulation which differs with distance as predicted. However, gains beyond the observers' own produce a less clear pattern of results between the two experiments and represent a reduction in relative depth rather than the predicted increase. One possible explanation for this is the 'cardboard cut-out' effect, whereby images viewed as stereograms appear flatter, like they are a cardboard cut-out, due to their vertical disparities (Howard & Rogers, 1995; Yamanoue, Okui, & Yuyama, 2000; Benzeroual, Allison, & Wilcox, 2011). Howard and Rogers (1995) explain that disparity and the size in the visual field of an object inversely vary with distance, and that the visual system uses an estimate of this distance by which to scale depth and disparity information. Therefore, an underestimation of distance would lead to the depth of the object from disparity scaling being too small compared with the perceived size from size scaling, which results in the object looking too flat. This effect was found in Chapter 3 where an

increasing underestimation of depth lead to participants overestimating the depth of the triangle needed due to it looking too flat. Here, it is possible that the gain manipulation combined with the misestimation of information by which to scale disparities has led to the objects appearing flatter than they should.

However, the findings of this work represent an initial investigation into these effects due to the low observer numbers and wide confidence intervals. Further experimentation is warranted to explore this difference in detail.

Experiment 2 also looked at separate monocular and binocular conditions and found no significant difference between the depth reported in each. This shows that a consistent amount of depth was perceived and reported when the cues were relative to a gain of 1. When taken together these points show evidence for cue combination theory for complex natural scenes, as no systematic bias was found at a gain of 1, with the linear weighted averaging model accounting for the effect of binocular gain.

6.5.3 Future work

Experiment 2 expanded on Experiment 1 by providing additional information in the form of other 3D models of fruits and vegetables to create a scene, whereby providing contextual reference points. A natural progression of this work would be to provide a further level of context by including background information by which to reference objects in their environment. Gibson (1950) states that without a ground surface, such as when looking at a cloudless sky, or in the case of the present work, looking at objects on a plain grey background, the impression of distance is hindered.

Additional scene context as described above could be done at varying levels, including the current example of a plain grey background, including a horizon as a sole reference point, up to a full cohesive scene in which the fruit and vegetables would naturally occur. This would provide an interesting avenue of research to measure the contribution of binocular disparity in these varying contexts.

One thing to note for Experiment 2 in particular is that testing was undertaken in March 2020 and was subsequently cut short due to the Covid19 pandemic. This left five full data sets for analysis, with several other planned participants uncollected due to lack of in-person activity at that time. Were this experiment to be conducted under different circumstances, higher numbers of participants, at least in line with Experiment 1, would have been tested to ensure sufficient analysis power.

Similarly to Chapter 5, possible gauge setting locations for both Experiments 1 and 2 did not go right to the contour of the object, which is where most extreme depth measurements are seen. The following chapters present work that addresses this by testing gauge settings at the extreme of objects as well as the centre.

A final consideration is that this work simulated an increase in the binocular disparity of objects by increasing the IOD gain. An alternative to this would be to increase the disparity by stretching or squashing the 3D models used for rendering. This approach forms the basis of the next chapter.

7 Occluding contour and shape from shading alone do not convey 3D metric depth

7.1 Abstract

Pictorial cues are derived entirely from the retinal image, and provide many sources of depth information in natural scenes. Here a cue perturbation method is presented to measure the contribution of the pictorial cues of shape from shading and the occluding contour to the overall depth estimate for 3D object shape in naturalistic scenes. A gain on pictorial cues was created by manipulating the rendered depth of points on naturalistic objects, here 3D rendered models of fruits and vegetables, to effectively stretch or squash them. Objects were presented in isolation or within a cluttered scene, and perceived depth was measured using a surface normal gauge figure task. It was predicted that correct scaling of depth estimates should match the scaling of the gains for the ideal observer. However, results of regression analysis through a linear mixed effects model show minimal effects of the gain on pictorial cues. These findings suggest that the occluding contour and shape from shading do not provide sufficient contributions to the depth estimate when other cues are not present for scaled metric depth.

7.2 Introduction

The previous chapter explored measuring the contribution of binocular disparity to the overall depth estimate through a method of cue perturbation, by introducing a conflict between binocular and pictorial cues, to measure its weighting. This chapter

employs a similar approach, this time to measure the contribution of the pictorial cues of shape from shading and the occluding contour.

7.2.1 Pictorial cues

An estimate of depth is made up of information from a wide variety of cues. Cutting and Vishton (1995) provide a comprehensive overview of these. Previous chapters have explored binocular cues such as disparity and vergence. The focus of this chapter is pictorial cues, which are monocular retinal cues, entirely derived from information from the image presented on the retina, and do not require binocular viewing to perceive depth.

There are a number of pictorial cues. For instance, Chapter 1 introduced the pictorial cues such as relative size and texture gradient. This work focuses on two particular pictorial cues: shape from shading, and occluding contours, both of which are illustrated in Figure 7.1.

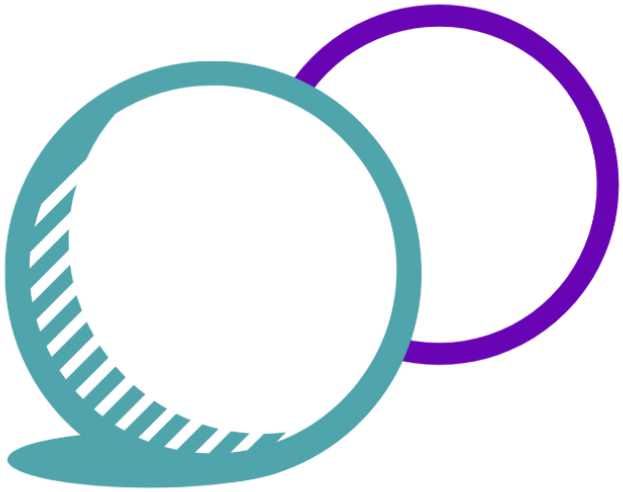


Figure 7.1: Showing shape from shading and occlusion.

The hatch markings on the turquoise ball show shape from shading, where shadows are created relative to the light source, top right in this image, which gives information about object shape (Cutting & Vishton, 1995), here giving the circle the three-dimensional appearance of sphericity. Points on the surface receive differing levels of light depending on their surface geometry, that is if they are orientated towards the light source or not, the type of illumination and how the material of the surface interacts with the light source (Egan & Todd, 2015; Anderson & Marlow, 2023).

Chapter 4 presented information from two sources of shading, from directional and diffuse illumination as presented by Hibbard, Goutcher, Hornsey, Hunter and Scarfe (2023). To summarise, directional lighting creates shadows that are proportional to the orientation of the surface relative to the light source, with surfaces normal to the light source appearing brightest (Hibbard, Hornsey, & Asher, 2022). Secondly, diffuse lighting creates shadows depending on the distance to the light source, with points closest to the light source appearing brightest, with those further away exhibiting the ‘Dark is Deep’ rule explored in Chapter 4.

The turquoise ball in Figure 7.1 is also seen to be closer to the observer than the purple one, due to the pictorial cue of occlusion, where objects closer to the observer naturally blocks those further away (Cutting & Vishton, 1995). Objects are not only susceptible to occlusion from other objects in the scene, but also from their own geometry. The boundary contour, or the self-occluding contour, defines which parts of the object are not visible to the observer due to being beyond the point at which the line of sight meets the object, here shown in Figure 7.2 with a cross. Parts of the object beyond this point are blocked from the observer and therefore the

retinal image by the rest of the object itself (Koenderink & van Doorn, 1982), a point which Cutting and Vishton (1995) argue is overlooked in some definitions given its obvious nature. Figure 7.2 shows how this point of self-occlusion is closer to the observer for an object stretched in depth.

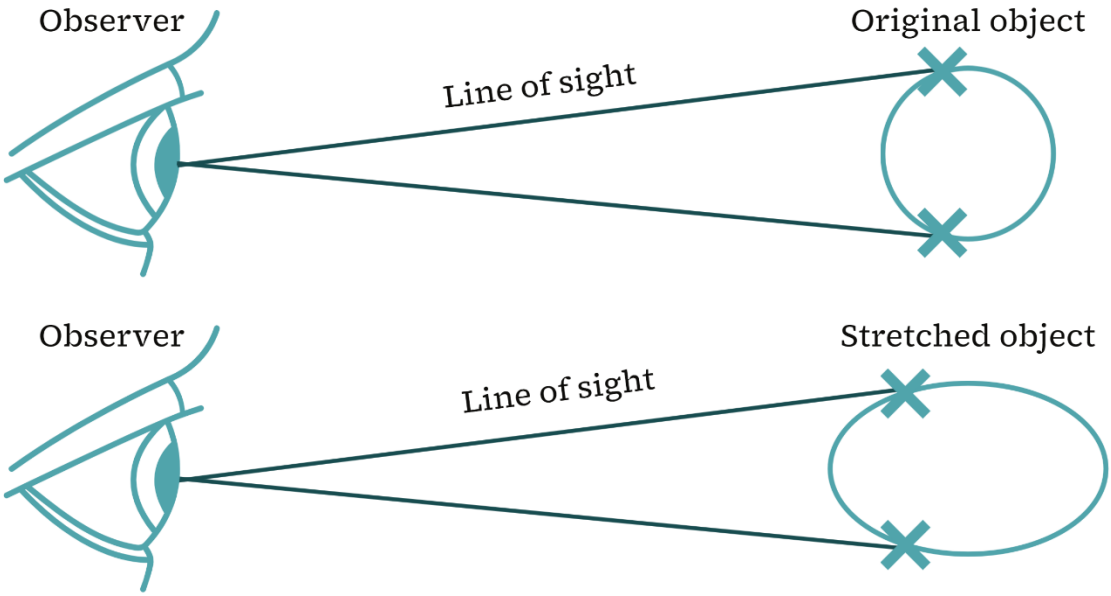


Figure 7.2: Self occlusion. Image showing the difference in self-occlusion where part of the object is not visible due to where the line of sight falls on an object compared to when that object is stretched.

Koenderink (1984) discusses how all points on the surface of a smooth solid object define the tangent plane, and how the rays from the vantage point tangential to the surface can be constructed into a conical surface or contour. He describes the contour as a spherical curve, with convexities and concavities, and argues against previous work that has suggested that the inflection of the contour has no significance for the surface (Marr, 1982). Instead, he says that inflection of the contour can indeed provide inflection of the surface and that the radius of the curvature of contours can be measured to understand surface shape (Koenderink, 1984). In this way, the occluding contour may provide information about surface

orientation, and therefore the occluding contour can itself be a potential monocular pictorial cue. Also, Marr (1977) points out that while very little information about the 3D surface shape can be gleaned from the occluding contour alone, the visual system is able to easily and accurately make use of occlusion information in judgements of surface shape, and other work has shown that smooth occluding contours can help infer shape from shading (Egan & Todd, 2015).

To explore the contribution of the occluding contour and shape from shading to the overall depth estimate, the work in this chapter manipulated the depth of naturalistic stimuli. Relevant to both cues of interest in this work is surface luminance. This is determined by both the reflectance properties of the surface and the effects of shading. This means that variations in luminance are determined by both of these factors. For example, if one part of the surface is darker than another, this may be attributed to differences in reflectance (the pigmentation of that part of the surface is simply darker than other parts) or shading (that the darker part of the surface is facing away from a directional light source more, or is partially occluded from a diffuse light source). A complex challenge for the visual system is to disentangle these contributions, so that both the surface structure and reflectance can be estimated (Kingdom, 2003; Hibbard, Goutcher, Hornsey, Hunter, & Scarfe, 2023). In the current study, we focused on shading cues only by ‘painting’ the objects with a uniform mid grey reflectance, as shape from shading information is conveyed by the luminance channel, and interference from surface colour information can reduce the ability to perceive surface shape (Ware, 2021).

7.2.2 Measuring depth estimates for pictorial cues

The measure of 3D object shape used in this chapter was a gauge figure task, where observers manipulated a figure shaped like a pushpin until it was perceived as being painted on the surface of the object, with the pin element sticking up normal to the surface (Koenderink, van Doorn, & Kappers, 1995). The orientation of the gauge figure that perceptually 'fits' the apparent surface (Koenderink, van Doorn, & Kappers, 1992) provides a measure of slant and tilt. When taken together, these provide information on the perceived local surface attitude (Koenderink, van Doorn, & Kappers, 1996). Nefs (2008) outlines the method to convert the slant and tilt settings into depth gradients, which he defines as the change in depth with a change in horizonal or vertical direction, to reconstruct a mesh surface of the percept, which has been used successfully previously in other works (Koenderink, van Doorn, & Kappers, 1992), within the previous chapters 5 and 6, and was used in the present study. Full details of this are outlined in Chapter 2: Methods.

These gauge figure gradients were used in a regression analysis using an affine transformation, which allowed for comparison between the reconstructed meshes, by measuring the change in one mesh when stretching and shearing it to fit a reference (Nefs, 2008). This method of affine transformation was employed in this work, with the models compared representing differing levels of the cues of interest through creation of a gain on pictorial cues.

7.2.3 Creating a gain on pictorial cues

The present work manipulated shape from shading, occluding contour and the effect of visual clutter, to measure the weighting of these pictorial cues in the overall

depth estimate. Depth from these cues was changed using graphic rendering to simulate a different depth from shading and boundary contour, effectively ‘stretching or squashing’ the mesh of the object. By doing so, an effective gain of pictorial cues was created. This was achieved during scene rendering, whereby the depth gradient of each vertex as described above was multiplied by the gain to alter the depth of the surface and therefore the overall scaling of the object. The experiments presented objects rendered without their scanned colour, ‘painting’ them grey by applying a consistent colour over the surface of the object, to reduce the effect of any colour information captured during the scanning process and isolate the cue of shape from shading.

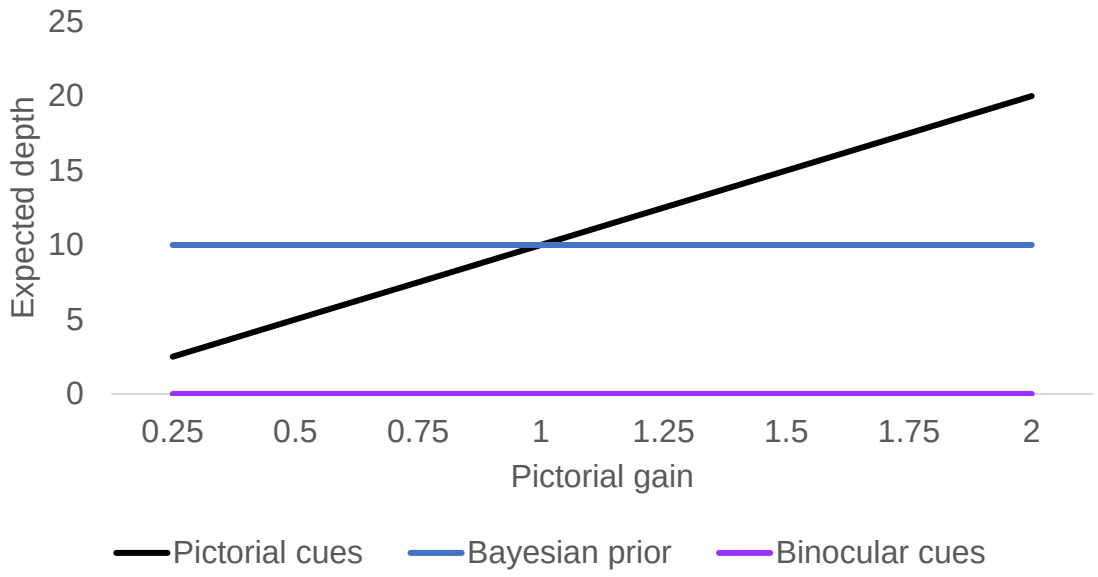


Figure 7.3: Pictorial cues, Bayesian prior and binocular cues slope comparison. Showing the predicted slopes for pictorial cues, Bayesian prior information and binocular cues for the ideal observer.

Figure 7.3 shows the predicted change in the depth reported if depth were to scale perfectly with pictorial gain, Bayesian prior information, and binocular cues. A gain of 1 here indicates the original depth of the 3D scanned object, for example

10cm. A gain of 2 would represent an object stretched in the z dimension to twice the size of the original, and therefore twice the depth, 20cm, would be expected to be reported by pictorial cues for the ideal observer. A gain of 0.5 represents half the original depth of the object, and therefore half the depth, 5cm, would be expected from pictorial cues alone.

Comparing differing conditions against the gain of 1 in this way probes the internal consistency of data structures as defined by Koenderink, van Doorn and Kappers (1992), and how this is affected by the stretching of the depth of the object. They presented photographs of a sculpture via a monitor using a CRT. Observers rotated an overlaid gauge figure until it appeared painted flat on the surface, giving an indication of surface orientation through slant and tilt settings. While the sculpture in the photographs was a real object, they did not compare observer's settings to the ground truth, instead they were interested in the pattern of results showing an internal consistency of the structure of local settings. As such, the present work used a gain of 1, representing veridical model depth, as a representation of the internal structure of the stimuli against which to compare the manipulated gain conditions for internal consistency.

7.2.4 Cue combination considerations

Many cues contribute to the overall depth estimate. As such, a measure of the contribution of pictorial cues to the overall depth estimate can be calculated. The work here presents two experiments exploring the contribution of pictorial cues using a cue perturbation method of introducing depth gain. Experiment 1 presented a single object and measured the depth estimated by participants across a series of

gains. Experiment 2 built on this by presenting both a single object, and that same object within a cluttered scene, to explore the effect of pictorial gain and how this interacted with the context of other information provided by other objects in the scene. Notably in this case, cue perturbation did not introduce any conflict across pictorial cues since all of these are consistent in each case with the same surface structure.

Koenderink, van Doorn and Kappers (1995) presented participants with a sphere and asked them to set a gauge figure flat to the surface. Their results found that whilst no settings were entirely veridical to the true depth of the sphere, they could not rule out that participants were being influenced by the prior knowledge of the circular contour of the sphere. Here, the Bayesian prior would have an effective slope of 0, given that this information would always report a consistent depth due to the spherical nature of the stimuli. That is, even without committing to a particular shape of prior, from our perspective the important consideration is that this would be constant across the variations in depth gain as shown in Figure 7.3.

Contrary to the experiments in the previous chapter, no additional information was present in the stimuli from binocular disparity, creating a slope of 0 across all gain conditions, as per Figure 7.3. This cue was not manipulated in this design, and indeed, the disparity cue should report that the surface is flat, given that stimuli were being viewed on a screen. In the weighted depth estimate this would mean that binocular disparity would be weighted so as to reduce depth (Hibbard, Hornsey, & Asher, 2022), although as it is not a cue in pictorial space, it may be ignored entirely in these judgements, such as how the surface of the canvas is ignored when perceiving depth in pictorial space within a painting (Koenderink, 2012).

This can be seen in experiments where, when viewed monocularly, more depth is reported when viewing pictures (Hibbard, Hornsey, & Asher, 2022). This is because, when present, binocular disparity is a strongly reliable cue, and therefore weighted highly in the depth estimate, leading to less depth being reported. When this cue is removed, by making observers view stimuli monocularly, other cues such as pictorial cues are weighted more in the depth estimate, leading to more depth being reported. While there was no additional information from binocular disparity in these experiments, the cue itself was still present given that observers completed the experiment with both eyes open. This is due to the online nature of data collection for the experiments in this chapter.

In summary, this work explored the contribution of pictorial cues to the overall weighted depth estimate for the weak fusion model of Bayesian cue combination. Specifically, this work was interested in shape from shading, the boundary contour and the impact of clutter in the scene. The aim was not to manipulate their contributions directly, but to measure how sensitive observers were to changes in depth gain for stimuli defined only by these pictorial cues. It was predicted that manipulating stimuli to create a gain on these pictorial cues would affect the perception of depth. Measuring the change in perception across gain conditions gave a direct measure of the contribution of these pictorial cues via their weighting in the overall depth estimate.

7.3 Methods

7.3.1 Participants

7.3.1.1 Experiment 1

28 participants took part in this study. The data for one participant were excluded after data visualisation as they had not completed the task as instructed, leaving the data from 27 observers for analysis. Full details of this are included in the Appendix. Participants were aged between 18 and 61. 19 (70%) identified as female, 8 (26%) male and 1 (4%) as trans male/ trans man. Participants included two researchers and 25 people naïve to the purpose of the experiment. All were required to have normal or corrected-to-normal vision (glasses or lenses).

7.3.1.2 Experiment 2

19 participants aged between 23 and 61 took part in this study. 12 (63%) identified as female, 5 (26%) male, 1 (5%) trans male/ trans man and 1 (5%) non-binary. Participants included two researchers and 17 people naïve to the purpose of the experiment. All were required to have normal or corrected-to-normal vision (glasses or lenses) as before.

7.3.1.3 Recruitment

Participants were recruited online through the University of Essex's SONA system, as well as through word of mouth. Some participants who were enrolled as Psychology students received course credit for their participation, while others were compensated financially. All participants signing up for this study were asked to wear glasses or lenses if they typically would do so to use a computer.

7.3.2 Apparatus and materials

The 3D models used in both Experiment 1 and 2 were created by scanning fruit and vegetables with a NextEngine 3D Scanner and Multidrive turntable. Objects were scanned on a 360 degree rotation with eight divisions. Settings were taken at two different tilts of ± 20 degrees with a resolution of 62 points per mm^2, and the macro distance range. Scanned objects were edited in the NextEngine Scan Studio HD software, with meshes fused with a 0.064mm tolerance to ensure a water-tight model. Further information on the scanner can be found in Chapter 2: Methods.

Images of the objects were rendered using OpenGL in MATLAB using the Psychophysics Toolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Scenes were rendered as if presented to the 'cyclopean eye' (Stidwell & Fletcher, 2011), in that the left and right eye positions were identical and rendered as if placed centrally in the scene, as per Figure 7.4.

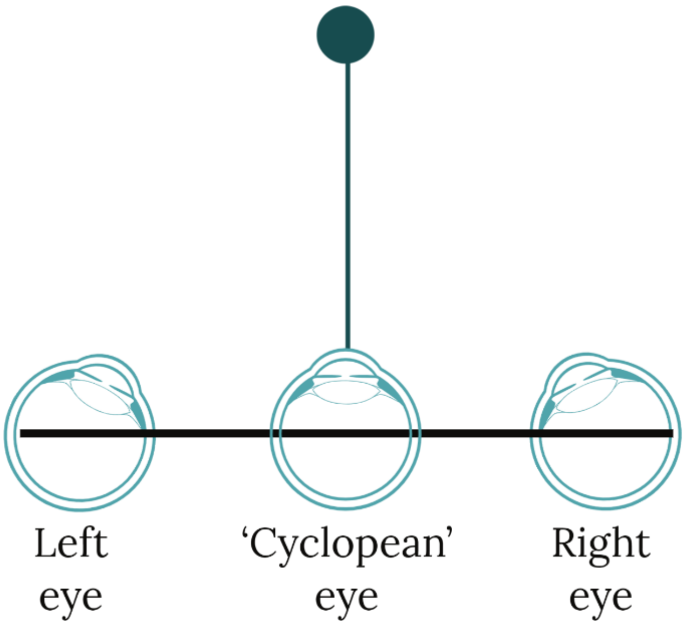


Figure 7.4: The cyclopean eye. Image showing how the 'cyclopean' eye view was created for rendering scenes.

Ambient and diffuse OpenGL lighting components were used, each with a magnitude of (0.7, 0.7, 0.7). The diffuse lighting defines the orientation of the surface relative to the directional light source, here located centrally and directed at the middle of the scene. In both Experiments 1 and 2, the main object was placed centrally in the scene, with objects in the cluttered scene of Experiment 2 placed at varying distances behind the main object and spaced horizontally.

Depth of the objects was manipulated by multiplying the depth of each pixel by the gain, such that a gain of 2 created a surface with twice the original depth, and a gain of 0.5 created a surface with half the original depth. Gains for Experiment 1 were 0.25, 0.5, 1, 1.5 and 2, and 0.5, 1, 1.5 and 2 for Experiment 2.

JavaScript code to render the gauge figures on the scene was adapted from work by Wijntjes and Van Zuijlen (2004). The experiment was created using the P5.js library (https://p5js.org/) in Qualtrics (https://qualtrics.com) and completed online using the participants' own desktop or laptop computer. Responses were recorded using the participants' desktop or laptop computer mouse, where participants would rotate the gauge figure by moving the mouse, clicking the left mouse button to confirm their judgement and move onto the next trial.

7.3.3 Stimuli

7.3.3.1 Experiment 1

The stimuli used in the first experiment were rendered from a 3D model of a scanned red onion. The surface of the onion was painted a uniform mid grey colour using MATLAB to remove any colour and reflectance information whilst retaining surface texture. As the object was painted grey, it was presented against a

background of squares of various randomly generated colours to make the object stand out, see Figure 7.5.

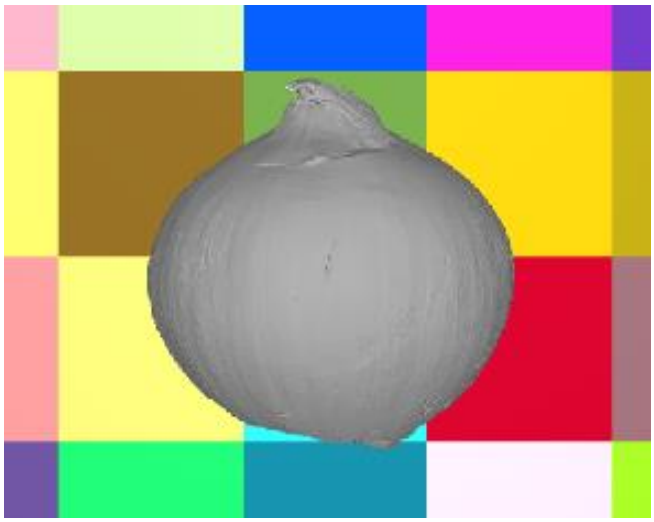


Figure 7.5: Stimuli Experiment 1. Image showing example image of the red onion presented to participants, here shown at a gain of 1 representing the original model.

The gauge figures superimposed onto the objects used in this experiment were white with a base diameter of 10 pixels. The possible gauge figure locations differed for each condition as the physical size of the models differed. A mesh of each 3D model was created using MATLAB whereby a triangular mesh was fit over the entire stimuli image, with points spaced 20 pixels apart. This created vertices across the scene, with three vertices representing a face, the centre of which, or barycentre, was sampled using the gauge figure, as illustrated in Chapter 2: Methods.

Vertices that fell within the occluding contour of the object were included in the object mesh. As the occluding contour changed with pictorial gain manipulation, some conditions had a different number of vertices, ranging from 52 points sampled at the smallest gain of 0.25, to 63 at the largest gain of 2. However, the location of the points that were present did not differ between conditions, allowing for direct comparison of these vertices between conditions. Figure 7.6 shows the best fitting

mesh for possible gauge figure locations at a gain of 1, with gauge figures sampling the middle of each triangle.

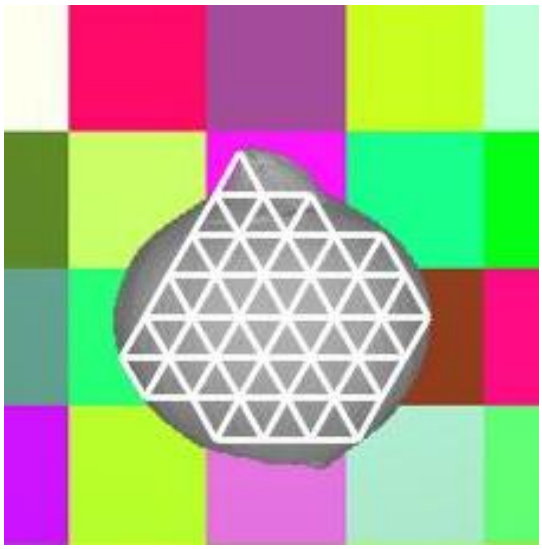


Figure 7.6: Sample points Experiment 1. Image showing a grid, not shown in the experiment, defining possible gauge figure locations at a gain of 1. Gauge figures were presented at the centre of mass, or barycentre, of the triangular faces.

Figure 7.7 plots the depth of points on the five gain stimuli presented via the z position, showing a gain of 0.25, 0.5, 1, which is the original object, 1.5 and 2 respectively. These are derived from the 3D model manipulated in scene rendering, and are presented here to illustrate the veridical change in depth with gain.

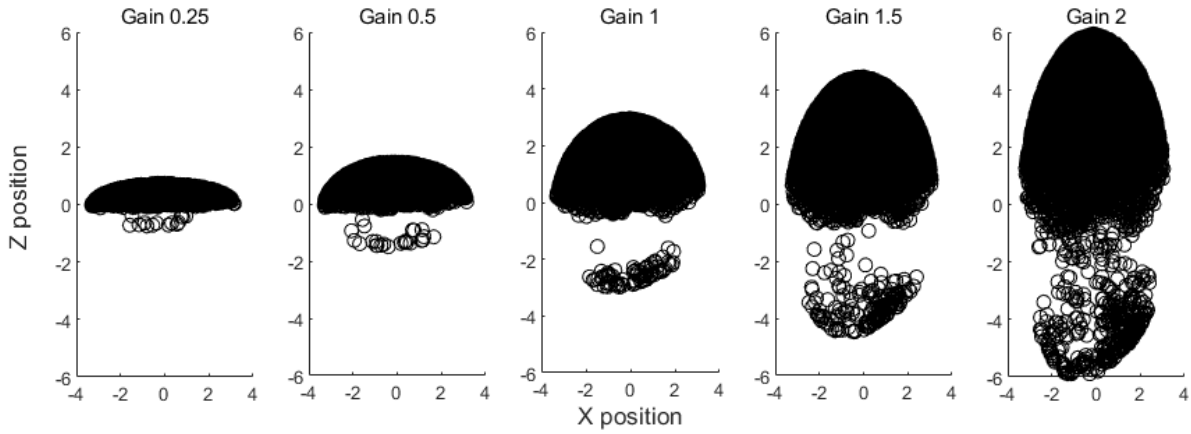


Figure 7.7: Depth of the 3D model presented in Experiment 1. Plots showing depth of the 3D model stimuli plotted for each gain: 0.25, 0.5, 1, 1.5 and 2 respectively. This shows increasing depth with pictorial gain manipulation.

From these manipulated 3D models, the average distance from the screen of the visible points was plotted, shown in Figure 7.8. This shows decreasing frequency of points close to the screen with increasing pictorial gain, with a positive number on the x axis representing points closer to the observer.

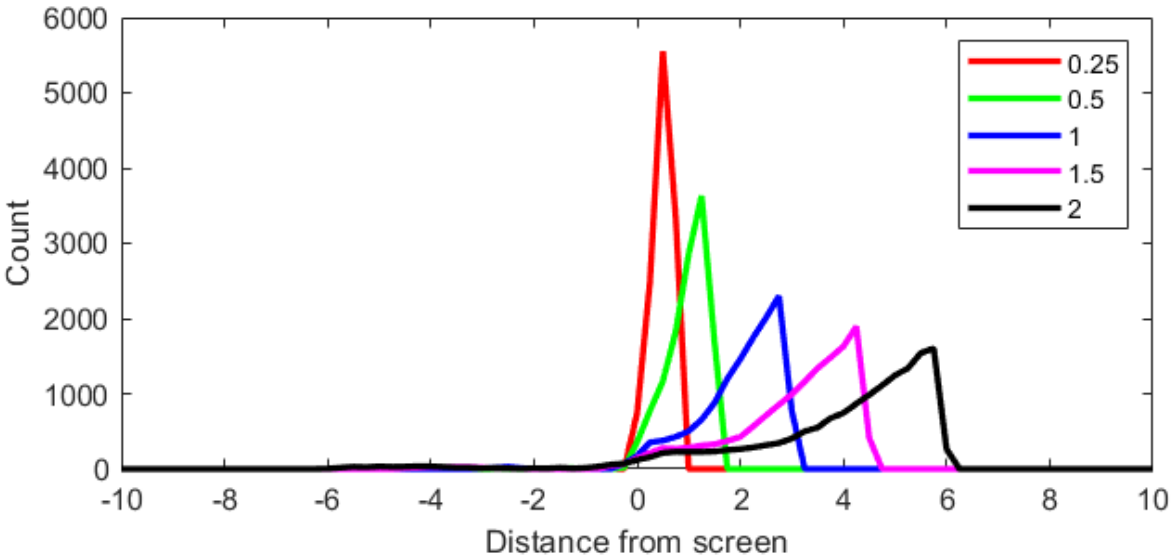


Figure 7.8: Point distances. Graph showing distance from screen of all visible points on each of the five gain objects in Experiment 1.

7.3.3.2 Experiment 2

For the second experiment, a scene of 3D models was used in addition to a single object condition. Similar to the first experiment, these were painted grey to remove reflectance information and presented in an arrangement where some naturally occluded others against a colourful block background - see Figure 7.9.

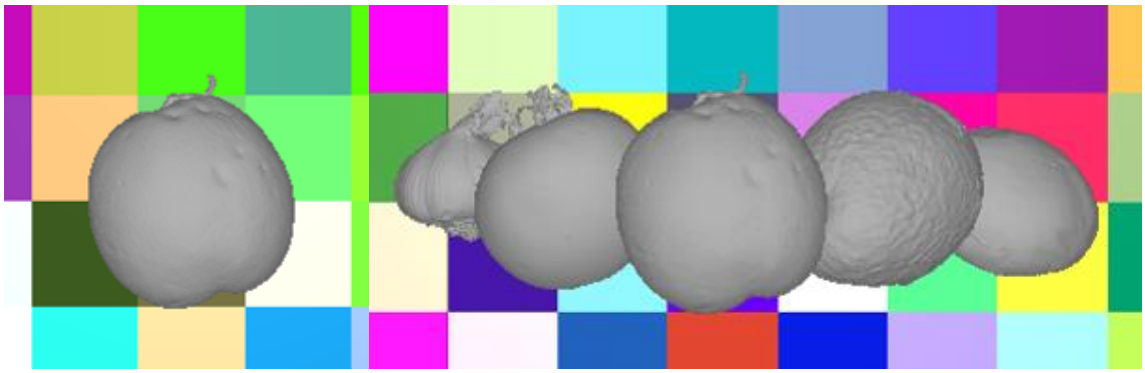


Figure 7.9: Stimuli Experiment 2. Showing stimuli used in the second experiment showing the single object condition on the left, and the cluttered scene of fruits and vegetables on the right.

For consistency the same object at the same rotation was used in the single object condition as within the cluttered scene. In the cluttered scene, the object from the single object condition was sampled using the gauge figure at the same gauge locations, allowing for comparisons for gain, as well as the context provided by the cluttered scene.

7.3.4 Procedure

7.3.4.1 Experiment 1

The experiment began with a few practice trials to ensure participants understood the task. This was completed with a different 3D model at a gain of 1 to ensure participants were not primed for the main task. Participants were asked to capture the shape of the practice object by rotating the gauge figure so that the circle element appeared to be painted flat onto the surface of the object (Koenderink, van Doorn, & Kappers, 1995), with the pin element normal to the surface. This was achieved by rotating the mouse, and once happy with the placement, participants clicked the left mouse button to record their answer and move onto the next trial.

Once participants felt happy they understood the task, they moved on to the experimental trials, where an estimate of surface slant was taken at each of the possible gauge figure locations in a randomised order. When all locations had been sampled the block ended and participants began a new block. In total there were 5 blocks, as the gain of the object was manipulated with 5 levels of 0.25, 0.5, 1, equivalent to normal viewing, 1.5 and 2, which corresponded to the object being stretched to twice its natural depth. Blocks were presented in a random order, and participants were offered breaks in between. Once all blocks had been completed, participants were debriefed.

7.3.4.2 Experiment 2

The procedure was much the same as above, but in this second experiment there were two different scenes, one with a single object and one with that same object presented in a cluttered scene. This time, there were four gains of 0.5, 1, 1.5 and 2 which were applied only to the target (central) object in the scene, with the other objects being rendered with a gain of 1 throughout. These were in addition to four gain conditions for the single object scene. This gave a total of eight blocks which were presented to participants in a random order.

7.4 Results

7.4.1 Data treatment

Raw scores were extracted from Qualtrics in the form of phi (slant) and theta (tilt) values relating to x and y positions in the scene. As these had been presented in a random order for each participant, these were reordered before analysis through

regression of x and y locations. These were then converted into depth gradients, calculated as the gradient of the angle between the perceived surface orientation and the line of sight, to give a measure of the magnitude and direction of the object surface using the methods outlined by Nefs (2008), details of which can be found in Chapter 2: Methods.

A mesh was then fit to the results using the face and vertices information, along with the observed orientations from the depth gradients, using a best fit from least squares deviation approach (Koenderink, van Doorn, & Kappers, 1992; Nefs, 2008). As the meshes included more vertices with increasing pictorial gain, some conditions had extra points during triangulation. In order to compare results across meshes, points not common to all meshes were excluded from analysis, leaving 39 common points for analysis. Example meshes for each of the five gains are shown for one participant in Figure 7.10.

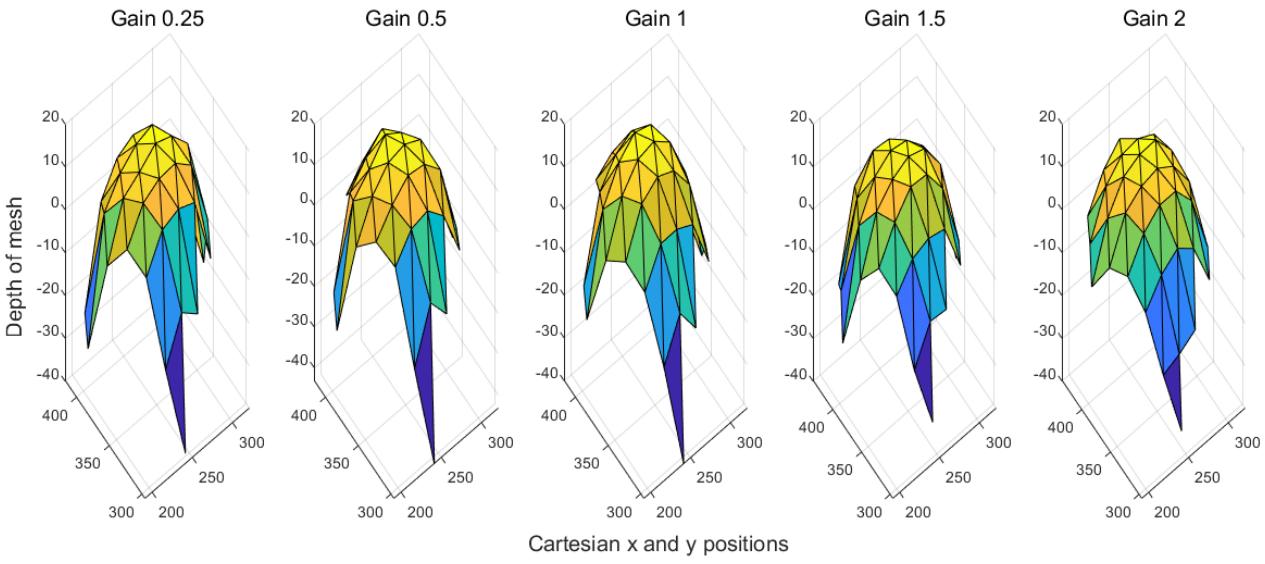


Figure 7.10: Reconstructed meshes Experiment 1. Plots showing example reconstructed mesh triangulations for one participant showing the five gain conditions. Closest to the observer in depth is shown in yellow, with furthest depth shown in blue, and colour gradients in between representing the depth gradients.

Reconstruction of the meshes allowed for visual verification of the data. Upon checking, one participant was found to have made settings not congruent with sensible settings for the task, in that the meshes were entirely inverted. This may be interpreted as them setting the normals pointing into, rather than out of, the object. As such, this data was excluded from analysis. However, this was not found to have a significant effect on the perceived depth, details of which can be found in the Appendix. Data from Experiment 2 were treated in the same way, converting slant and tilt settings into depth gradients and first verifying the data visually by rendering the meshes of the perceived surfaces. No participants needed to be excluded in Experiment 2.

As with the other mesh reconstruction experiments in this body of work, a restriction was added to the maximum angle for gauge settings of 85 degrees, in order to reduce the exponential nature of extreme gradients, more details of which can be found in Chapter 2: Methods. Figure 7.11 shows an example of mesh reconstruction without a maximum angle constraint, which shows how angles beyond 85 degrees result in extreme gradients which skew the overall scaling of the mesh beyond interpretation, compared with meshes fit to the same data with an 85-degree constraint. Note the extreme values on the y depth axis for the top set and the overall visual flattening of the mesh without the constraint.

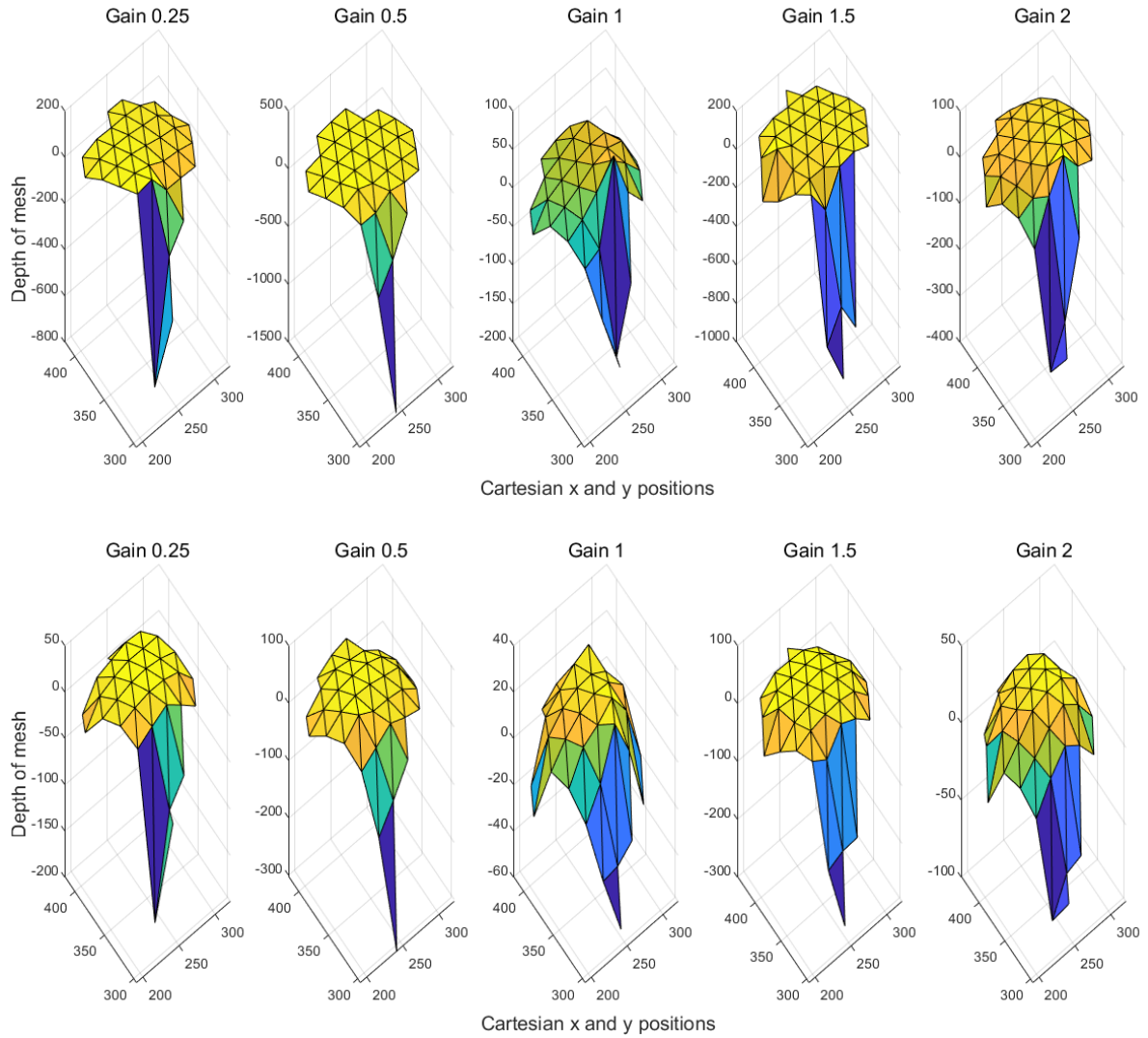


Figure 7.11: Maximum angle constraint. Plots showing example of mesh reconstruction without maximum angle constraint (top), and with an 85 degree constraint (bottom) for the same data from one participant.

During data visualisation, one point repeatedly resulted in a very large gradient, as shown by the set of meshes at the top of Figure 7.12, even with the 85 degree angle constraint. The large gradient of this single point resulted in skewing of the reconstructed meshes as described above, with very little differentiation across the model. When this point was removed as shown in the bottom set of meshes, the reduced depth scale allows for proper visualisation of the surface orientation changes across the model.

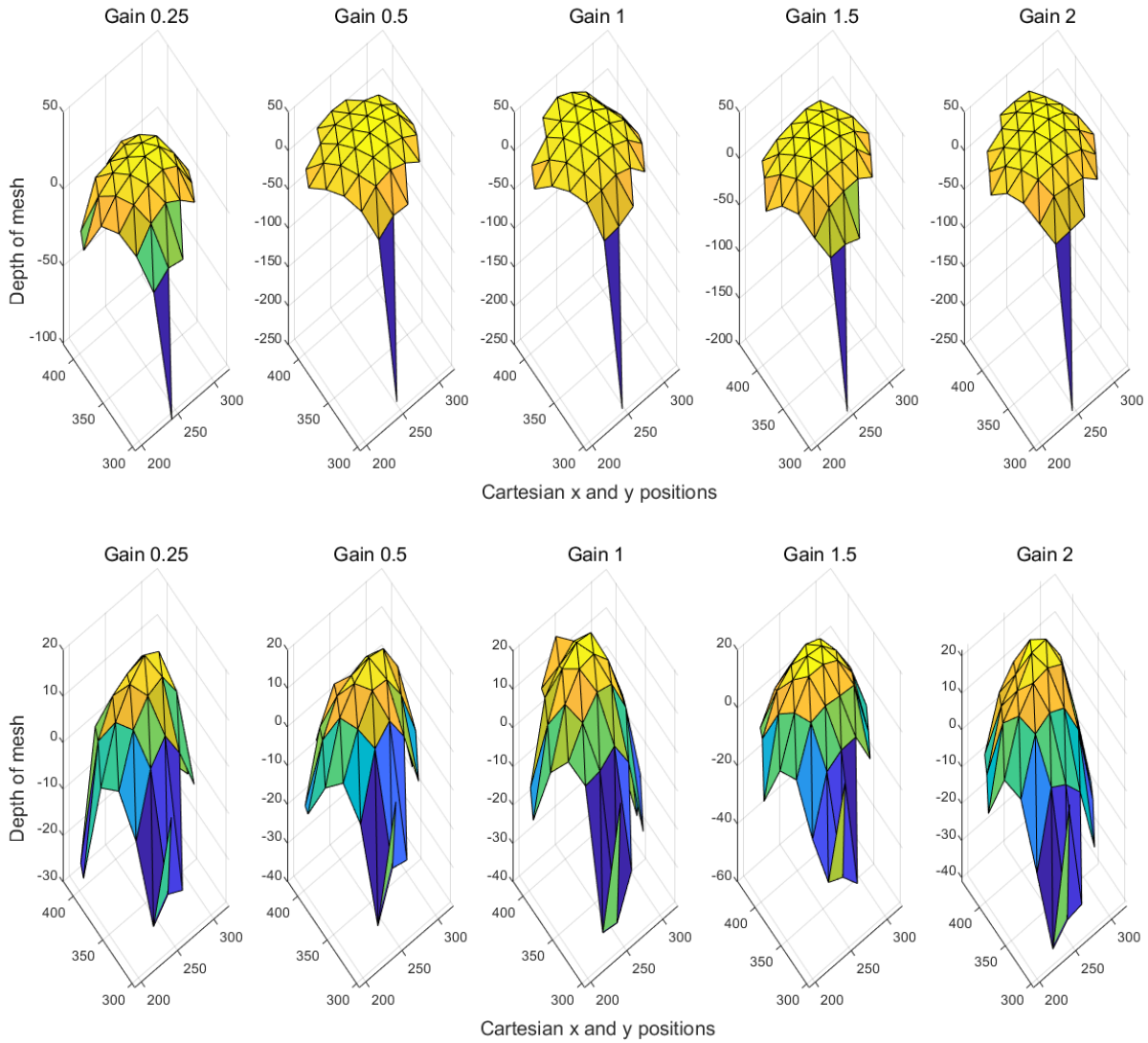


Figure 7.12: Large gradient point. Plots showing example of single data point creating extremely large gradients compared to every other setting (top), and mesh fit to the same data with that point removed (bottom).

Analysis is presented here without this point, but the full analysis was run with this point included for completeness, and the pattern of findings did not differ without it, results of which are shown in the appendix.

7.4.2 Experiment 1

As presented earlier, Koenderink et al (1992) describe their work exploring internal consistency of observations, without an objective ground truth. In this way,

their findings are operationally defined by the stimuli and task chosen. Here, this principle is used to compare the meshes against the gain of 1 to represent the original model as the operating point against which gain manipulation is compared.

Having computed the meshes for each gain, multiple linear regressions were run comparing the depth meshes for all gain conditions as a target against the gain of 1 as a reference using the equation outlined in Chapter 2: Methods. Here the predicted response ($\hat{y}$) is the depth with the predictor variable (x) being the horizontal and vertical locations, and participants' perceived depths at the reference gain of 1. This produced regression slope values for each participant for the four coefficients of the intercept (b0), x axis (bx), y axis (by) and z axis (bz).

The slopes of these regressions were used to predict the depth in the target gain condition from the depth in the reference condition with a gain of 1. The main coefficient of interest here is the slope of the z axis, showing the change in perceived depth as the gain in the rendered depth is changed. The predicted values of the depth gain should match that in the stimulus for a veridical observer, with a gain of 2 having a slope of 2 and so on. Here, the gain 1 condition shows a regression slope of 1, as this represents the target being referenced against itself. A slope of more than 1 suggests an increase in depth reported compared to the reference, and a slope of less than 1 represents less depth reported for the target compared to the reference. The means of the slopes are plotted in Figure 7.13. The slope of the affine depth stretching against the stimulus gain would take a value of 1 if observers were fully responding to the changes in depth. Further analysis was conducted to explore the trend of slopes across the gain conditions.

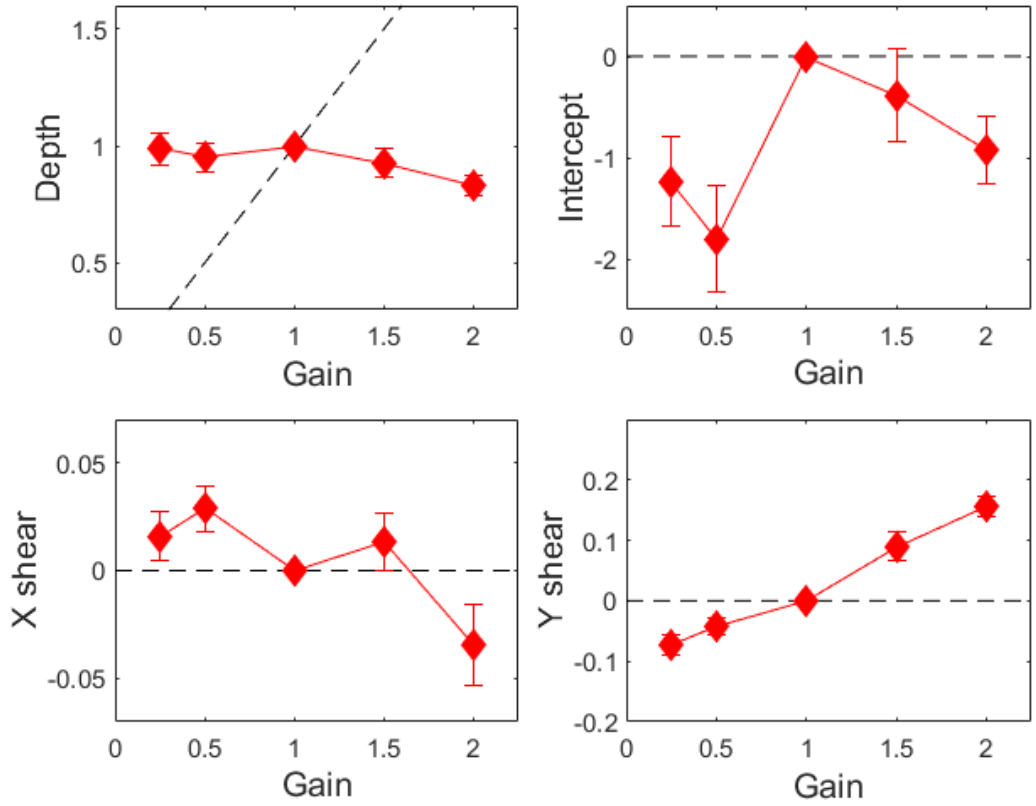


Figure 7.13: Affine transformations Experiment 1. Plots showing mean (SE) slope values for the effect of gain on the four coefficients in Experiment 1, with the dotted lines showing expected transformations for the ideal observer.

As observers completed all five gain conditions, this work has a repeated measures design, so a linear mixed effects (LME) model was selected to account for this. A general background to this approach can be found in Chapter 2: Methods. Here, the slopes of the depth of target against the reference were used to explore the change in reported depth (bz) across gain conditions (g) as a fixed effect, relative to each individual observer's settings (o) with random slopes and intercepts, and an intercept of 1:

Equation 7.1

$$bz \sim 1 + g \ + (1 + g \mid o)$$

The null hypothesis here is that the slope value will not differ from 0, meaning that there was no effect of increasing gain on the depth reported and the same depth would be reported for all gains as for the reference gain of 1.

Results show that the slope did differ significantly from 0, and thus the null hypothesis is rejected, as the fixed factor of gain was found to significantly affect depth estimates, with a beta slope value of β=-0.08, (***p*=.006**). However, this is a negative slope, contrary to what was predicted, showing that increasing the gain significantly reduced the depth reported by observers. The slope value can be taken as a direct measure of scaling, showing that increasing the gain on depth of the objects actually resulted in an 8% smaller depth estimate, and decreasing gain resulted in a bigger depth estimate. This is a small effect when compared with the expected 100% scaling with gain, although it is a reasonably reliable finding with confidence intervals (CI) placing this value between 2% and 13% (95% CI [-0.13 -0.02]). The results for all four coefficients are summarised in Table 7-1. These models include the random slopes and intercepts, grouped by observer.

Table 7-1: LME results for affine transformations Experiment 1.

Coef fi-cient	Model	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
bz	$bz \sim 1 + g + (1 + g \mid o)$	-0.0761	0.027	133	**.006****	-0.1295	-0.0227
b0	$b0 \sim 1 + g + (1 + g \mid o)$	0.456	0.263	133	.085	-0.064	0.976
bx	$bx \sim 1 + g + (1 + g \mid o)$	-0.0270	0.008	133	**.001****	-0.0430	-0.0109
by	$by \sim 1 + g + (1 + g \mid o)$	0.132	0.013	133	**<.001*****	0.108	0.157

The intercept represents the relative distance of each object, relative to the gain of 1. As all objects were presented at the same distance, a slope of 0 would be expected from the ideal observer, although here they did not differ significantly from a gain of 1 as shown in the table, with confidence intervals falling either side of 0. The x position refers to shearing on the x axis. This should also have a slope of 0 for the ideal observer. Negative numbers here show a slight deviation for all objects relative to a gain of 1, although this was not found to be significant, with confidence intervals falling either side of 0. The y position refers to a vertical shearing in the stimuli. Again, this should be a slope of 0 for the ideal observer. These results do show a significant effect of shearing in y, (*β=0.13,* ***p*<.001**), meaning that between gain conditions the perceived objects sheared as a function of vertical position compared to the gain of 1. This is estimated to be around 13%, with confidence intervals putting it between a very narrow range of 11% to 16%.

Finally, the overall depth in the reconstructed meshes was compared by calculating the relative depth range. This was the difference between the maximum and minimum depth values in the meshes, divided by the gain of 1 to normalise to the standard model.

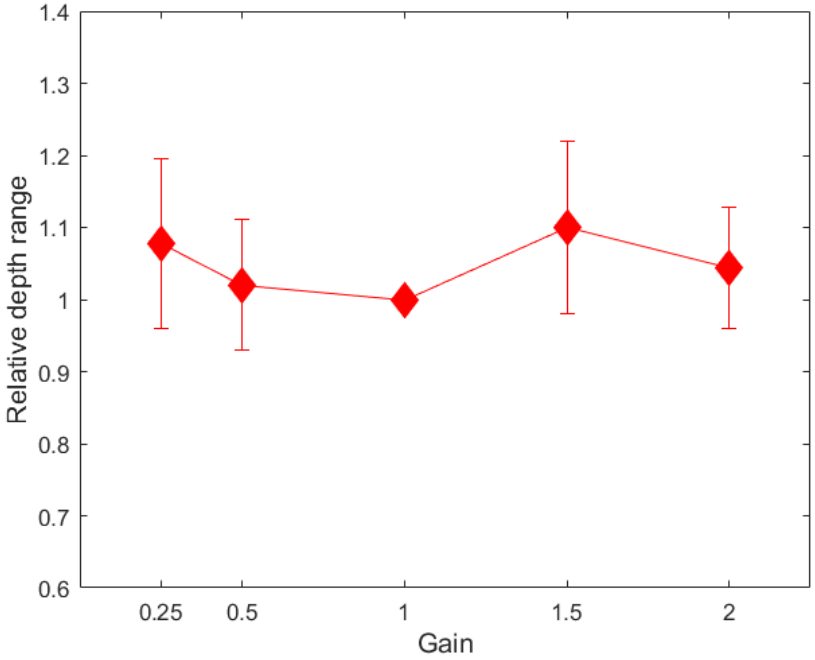


Figure 7.14: Relative depth Experiment 1. Plot shows the mean (SE) relative depth range Experiment 1, calculated as the distance between the nearest and furthest points on the reconstructed meshes. Depth ranges are normalised to a gain of 1.

An LME found no evidence of an effect of pictorial gain on the relative depth range of the reconstructed meshes, (*β=0.050, p*=.423, 95% CI [-0.072 0.173]), as shown in Figure 7.14.

7.4.3 Experiment 2

The analysis of Experiment 2 looks at the effect of gain on the single object scene, the object within the cluttered scene, and then a comparison of the two. It begins with the single object conditions to see if the negative effect of gain in Experiment 1 is replicable.

7.4.3.1 Single object scene

To begin with, the single object scene was analysed using the same procedure as Experiment 1. As before, data was first verified visually by rendering the meshes of the perceived surfaces of the objects, and example of which can be seen in Figure 7.15.

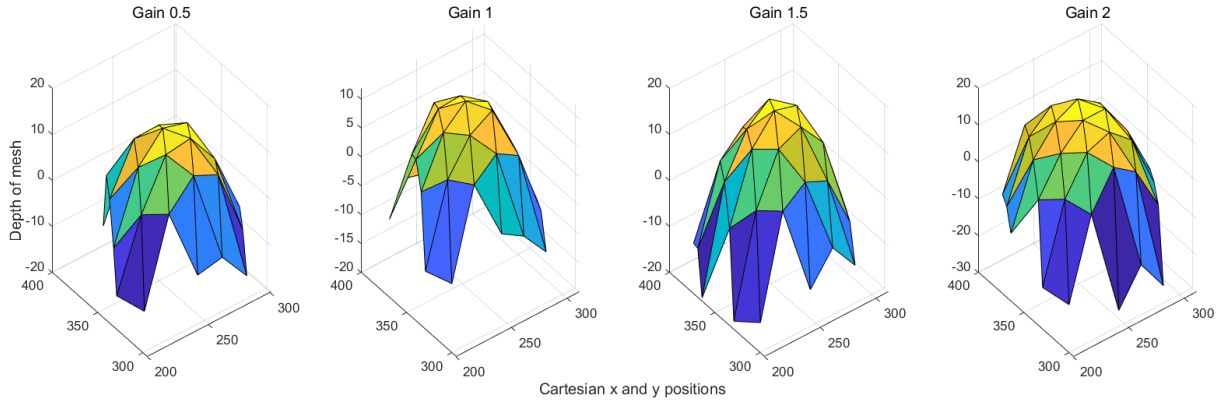


Figure 7.15: Reconstructed meshes Experiment 2. Plots showing example of mesh recreation across the four gain conditions for one participant.

Once again, multiple regressions compared the change in perception of the surface through affine transformation. The means and standard error (SE) of the slopes are plotted in Figure 7.16.

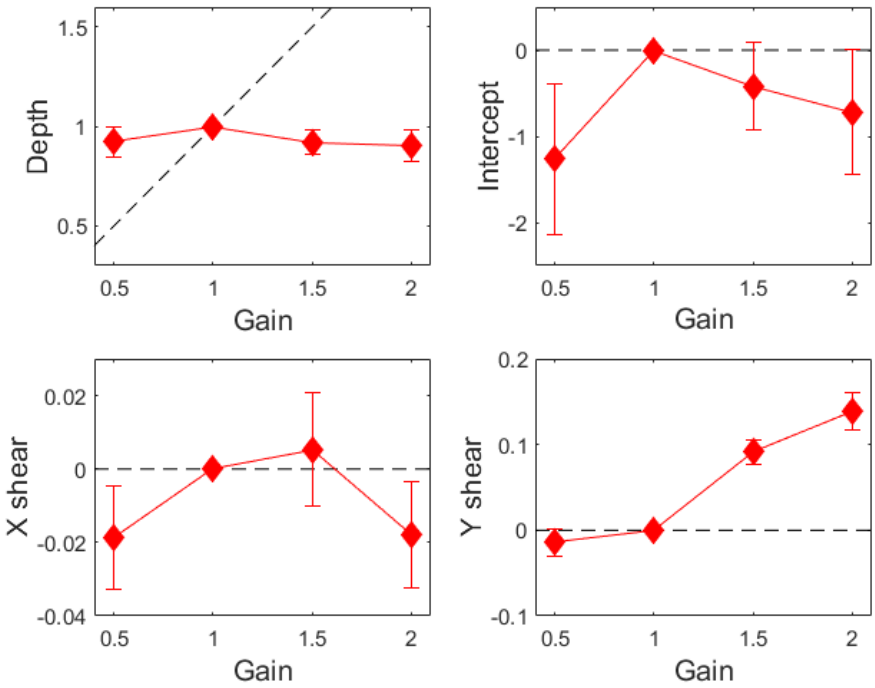


Figure 7.16: Affine transformation Experiment 2 single object. Plots show the mean (SE) effect of gain on the four coefficients, dashed line shows expected performance for the ideal observer.

A linear mixed effects model explored the transformation of each of the coefficients as a result of pictorial gain. Given the full model including random slopes and intercepts was included in Experiment 1, the same model is presented here in Equation 7.2.

Equation 7.2

$$bz \sim 1 + g \ + (\, 1 + g \mid o \,)$$

There was no significant effect of gain on reported depth for the single object scene (p=.525, 95% CI [-0.110 0.057]). Again, significant y shearing was observed (***p*<.001**), estimated to be 11% vertical increase with increasing pictorial gain, with confidence intervals placing it between 8 and 14%. LME results for all coefficients are shown in Table 7-2.

Table 7-2: LME results for affine transformations for the single object Experiment 2

Coeffi-cient	Model	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
bz	$bz \sim 1 + g + (1 + g \mid o)$	-0.027	0.042	74	.525	-0.110	0.057
b0	$b0 \sim 1 + g + (1 + g \mid o)$	0.241	0.405	74	.554	-0.566	1.047
bx	$bx \sim 1 + g + (1 + g \mid o)$	0.0015	0.0127	74	.907	-0.0239	0.0268
by	$by \sim 1 + g + (1 + g \mid o)$	0.110	0.014	74	**<.001****	0.083	0.137

Once again, the relative depth range was calculated and analysed using LME. Here, pictorial gain was found to increase the relative depth range by 16%, (*β*=0.161, ***p*=.035**, 95% CI [0.011 0.311]). Mean relative depth ranges per gain are compared between the single object and cluttered scene conditions below.

7.4.3.2 Cluttered scene object

Next the analysis will be presented for the same object within the cluttered scene. After visual verification of the data through mesh reconstruction, regression slopes were calculated as before to quantify the affine transformations. The means and standard error (SE) of the slopes are plotted in Figure 7.17.

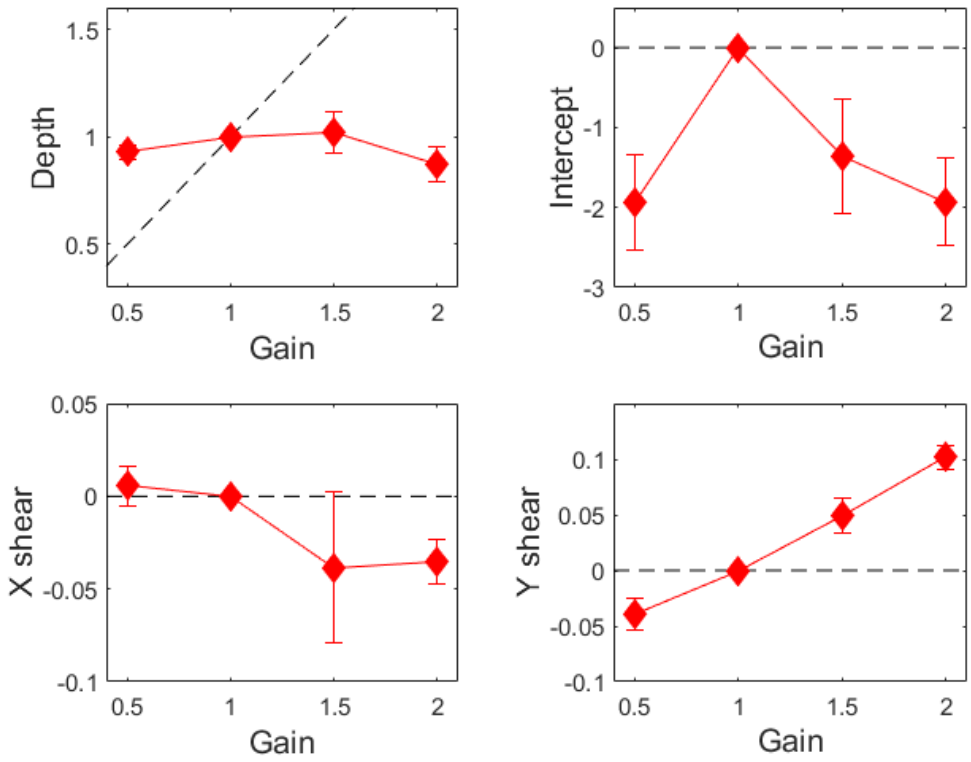


Figure 7.17: Affine transformation Experiment 2 cluttered scene. Plots show the mean (SE) effect of gain on the four coefficients for the scene object, dotted line showing expected performance.

A particularly large standard error was observed in x shearing at a gain of 1.5 compared to the other gains, but this represents a very small increase in shearing overall so does not affect results.

As before, the coefficient of particular interest is the depth transformation, as shown by the change in the z dimension. The full random slopes and intercepts model is presented here in Equation 7.3:

Equation 7.3

$$bz \sim 1 + g \ + (1 + g \mid o)$$

Linear mixed effects models were conducted to explore the transformations of the four coefficients with increasing gain. No significant effect was found of gain on

depth reported in the cluttered scene object condition, (*p*=.626, 95% CI [-0.156 0.095]). There was significant shearing on the y axis, similar to the previous analysis, (***p*<.001**), with an estimate of 9% shift on the y axis with increasing gain, with confidence intervals placing this strongly somewhere between 8 and 11%.

Table 7-3: LME results for affine transformations for the cluttered scene object in Experiment 2.

Coeffi-cient	Model	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
bz	$bz \sim 1 + g + (1 + g \mid o)$	-0.03	-0.031	74	.629	-0.156	0.095
b0	$b0 \sim 1 + g + (1 + g \mid o)$	-0.267	0.420	74	.527	-1.103	0.569
bx	$bx \sim 1 + g + (1 + g \mid o)$	-0.032	0.019	74	.099	-0.070	0.006
by	$by \sim 1 + g + (1 + g \mid o)$	0.094	0.009	74	**<.001*** **	0.077	0.113

As before, depth ranges were calculated, relative to a gain of 1. An LME found a 20% increase in relative depth range with pictorial gain, (***β***=0.208*,* ***p*=.016**, 95% CI [0.040 0.376]).

7.4.3.3 Single object and cluttered scene object comparison

Having presented the results for the object both as a single object and as part of a cluttered scene, a comparison between the two will now be presented. The x, y and z coefficients from the single object conditions were used to predict the depth transformation across gains in the cluttered scene condition with multiple linear regression. Means (SE) are plotted in Figure 7.18.

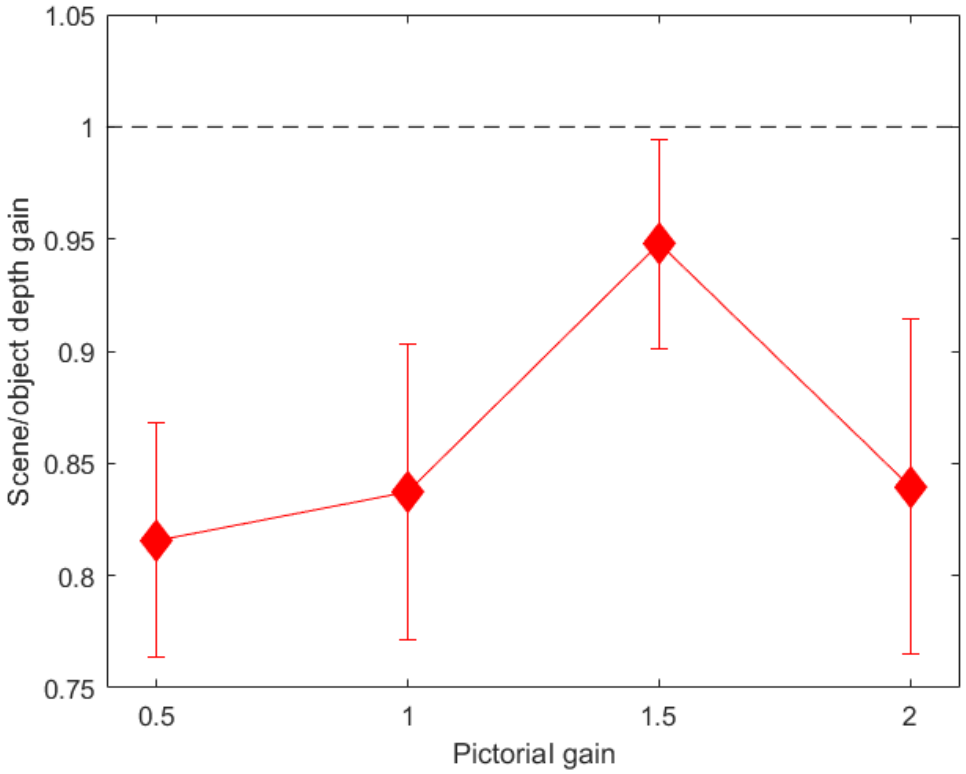


Figure 7.18: Depth gain. Graph showing comparison of the mean (SE) depth gain for the single object against the scene object depth transformation across gains, with the dashed line showing perfect performance.

This found that the rate at which the location of vertices in the single object condition predicts the depth of those points on the object in the scene condition did not change linearly with increasing gain, (R^2=0.157, p=.604). It had been predicted that depth in the scene condition would be greater than in the object condition, as shown with a depth gain greater than one, due to the increased context by which to scale depth information, and that this would change linearly with increasing pictorial gain.

However, depth was seen to be consistently lower in the cluttered scene than the single object condition, as shown by slope values consistently less than 1, which would indicate equal depth in each condition. A *t*-test found that the mean of depth transformations was significantly different from 1, meaning that consistently less

depth was reported in the scene condition than the single object condition, (t(18)= -4.2234, SD=0.144, ***p*<.001**), with an effect size of -0.140, 95% CI [-0.209 -0.070]. This shows that less depth was perceived for the same object when viewed as part of a cluttered scene than without the context of other objects.

The relative depth ranges between the object and scene conditions were then compared, means (SE) are shown in Figure 7.19.

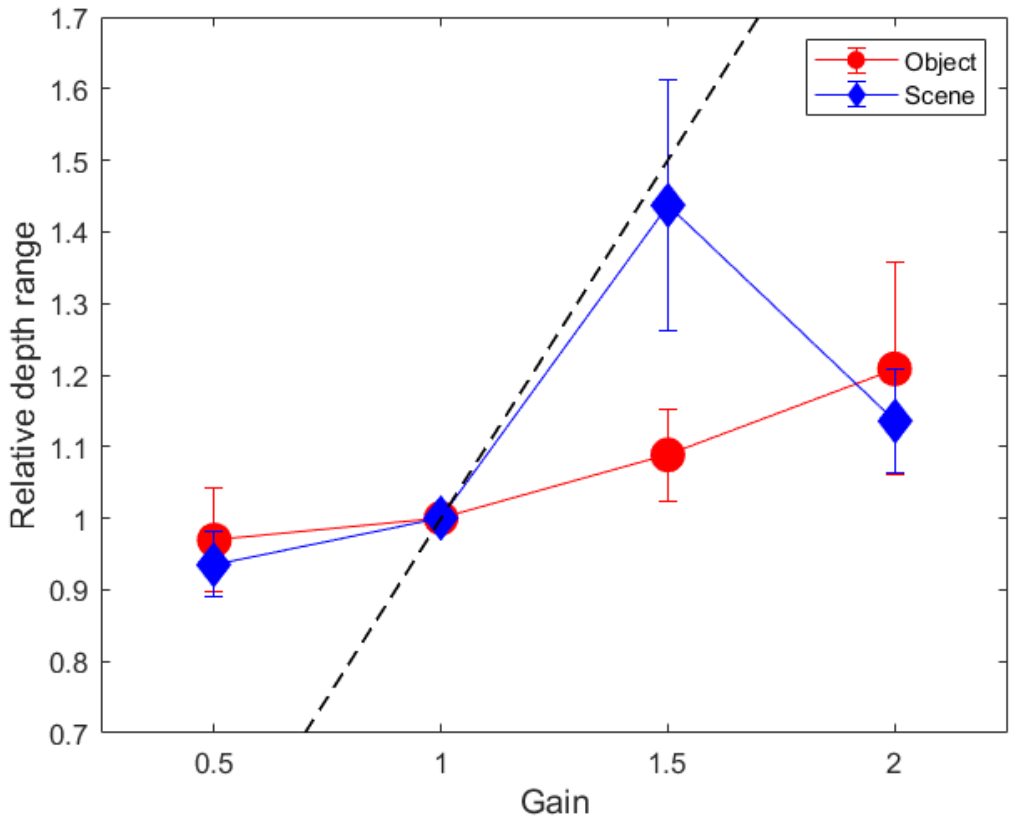


Figure 7.19: Relative depth Experiment 2. Showing mean (SE) relative depth range comparison between the single object and the object in the cluttered scene. Black dashed line shows predicted performance for pictorial cues only.

An LME was used to see if relative depth (r) could be predicted by object or scene condition (os), gain (g), or an interaction of the two, using the following equation:

Equation 7.4

$$r \sim 1 + g * os + (1 + os|o)$$

This model contains random intercepts and random slopes of object or scene condition. Results are shown in Table 7-4.

Table 7-4: LME results for relative depth range object and scene comparison.

Model	Variable	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
$r \sim 1 + g * os + (1 + os\|o)$	g	0.161	0.075	148	**.032***	0.014	1.085
	os	0.002	0.157	148	.990	-0.309	0.312
	$g * os$	0.047	0.106	148	.657	-0.162	0.255

A 16% increase in relative depth range with pictorial gain was found between both object and scene conditions. However, depth ranges were not found to differ significantly between object and scene conditions as had been predicted, and there was no interaction with gain. These results show a modest effect of pictorial gain on relative depth scaling, although no effect of the cluttered scene context on pictorial weighting.

7.5 Discussion

7.5.1 Findings from Experiment 1 and 2

Results from both Experiments 1 and 2 showed that observers did not perceive an increase in depth with increasing pictorial gain as might be predicted. This means that observers did not perceive more depth when objects were stretched, and less

depth when objects were squashed, as would be expected and was predicted in this work.

For Experiment 1, while observers did not see an increase in depth with increasing gain as predicted for the ideal observer, a small but significant negative effect of gain on depth was found, meaning observers perceived on average 8% less depth with increasing pictorial gain. These findings suggest that observers saw more depth compared to the original object when objects were squashed, and less depth compared to the original when objects were stretched. This is contrary to the predictions for the ideal observer. A similar study was conducted in Experiment 2 to see if this result was replicated. However, contrary to Experiment 1, the slope of depth transformation across gain conditions did not differ significantly from 0, meaning despite the stimuli being squashed or stretched relative to a gain of 1, observers did not perceive a change in depth of the object, and the results of Experiment 1 were not replicated, suggesting that they were an artefact of the experiment design rather than an empirical finding.

The lack of depth transformation with increasing pictorial gain overall could suggest that observers are utilising Bayesian a priori information which dictates that certain objects, such as the apple used in Experiment 2, are expected to be spherical. Therefore, when the gain for pictorial cues is increased, this no longer matches this a priori information, and the pictorial cues of occluding contour and shape from shading are therefore weighted less in the overall depth estimate, creating a slope closer to 0 which shows no perceived change in depth across conditions.

The comparison between the single object and the object in the scene did not find that depth transformation changed with gain. However, consistently more depth was reported in the single object condition than for the cluttered scene overall. This suggests that the addition of other depth cues in the cluttered scene resulted in a different interpretation of the depth cues present. For instance, shape from shading is not an absolute depth cue, but rather provides relative depth information about the orientation and depth order of objects in the scene (Brooks & Horn, 1985), with the additional objects in the scene condition providing this relative information. Additionally, the direction of illumination plays a key role, as the location of the light sources helps to interpret shading information (Todd & Mingolla, 1983). The lack of background or other context in the single object scene does not provide this key information, and therefore may have led to a different interpretation of the depth cues than for the scene condition, where additional information about the light source by which to interpret shape from shading is available.

These findings suggest that occluding contour and shape from shading are not sufficient in isolation to fix a metric scaling of the perceived depth relief and therefore, do not contribute strongly to the absolute metric depth. This may reflect the fact that observers are primarily able to recover depth relief up to an arbitrary scaling, rather than metric depth, from these pictorial cues (Koenderink, 1998; Koenderink, van Doorn, Kappers, & Todd, 2001; Glennerster, Rogers, & Bradshaw, 1996). The stretching of the stimuli in depth prior to rendering, while increasing the depth range in the stimuli, does not affect depth relief. This supports the findings of Experiment 2, where an increased depth range was observed with pictorial gain, but no evidence of increased depth relief from affine transformations.

Koenderink et al (1992) also discuss how one observer in their work had outliers in their depth gradients consistent with vertices at the contour, which they state ‘spoils’ the linear regression analysis. As points were taken across the entire surface of the objects in the present work, this included the boundary contours themselves. This could account for the extreme values observed from some observers, and in particular the point removed from the main analysis.

Nefs (2008) presents a problem for occluding contours called the false target problem, whereby any number of differently shaped objects may produce the same occluding contour if they sit tightly within the same boundaries created by the viewing position. The resulting retinal image can cause ambiguity for the visual system when decoding the shape of these objects.

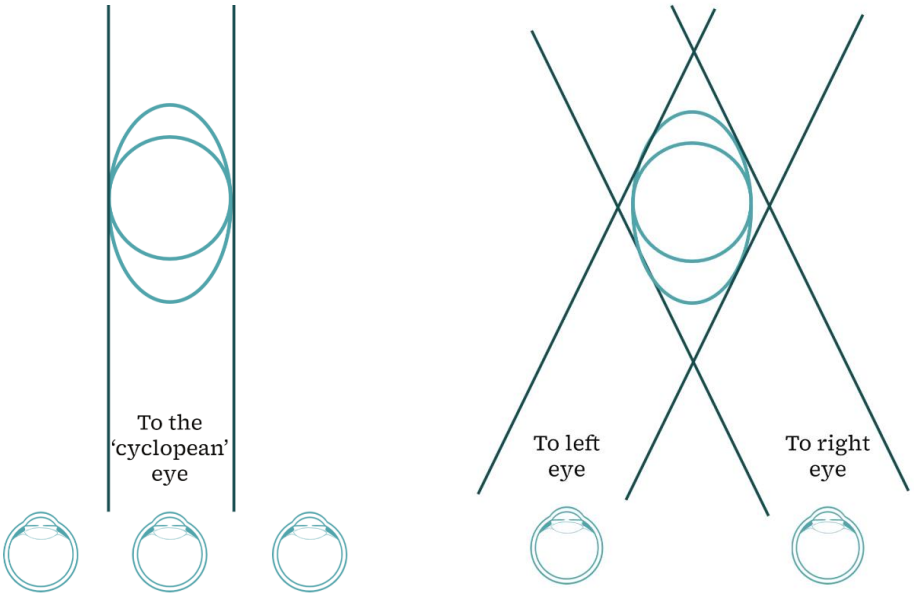


Figure 7.20: Cyclopean versus binocular viewing. Diagram showing how a stretched object and the original object sit within the boundaries for both cyclopean and binocular viewing.

In terms of the present work, Figure 7.20 shows how both a stretched object and the original object fall within the same boundaries for cyclopean viewing on the left. As such, the stretched stimuli here would have fallen within the same boundaries, and therefore fallen foul of the false target problem. However, as is show on the right, these two objects would not sit tightly within boundaries for binocular viewing. As such, this experiment should be repeated for binocular viewing using methods similar to those used in the previous chapter, Chapter 6, to investigate the effect of differing boundary on the perception of the occluding contour. Further areas of potential new research are discussed below.

7.5.2 Future studies

Following on from the above suggestion to address Nefs' false target problem (2008), Koenderink (1984) discusses how the space curve of the occluding contour itself changes as the observer moves between vantage points. This could be explored in the above suggested work by adopting several vantage points for the stimuli, as the present work only presented the models at a single vantage point given the online nature of the study. In addition, the stimuli here is only of a convex shape, and future work could incorporate natural objects with a concave portion to look for an effect of gain on these regions.

Here, the shape of the object was manipulated directly by stretching and squashing the models of the objects during rendering of the scenes to create a gain on pictorial cues. An alternative to this is to manipulate the image surface, rather than the model. The next chapter will explore this idea by creating a gain on pictorial

cues by manipulating the appearance of shadows on the surface of the images, while keeping the model shape constant.

Although it did not significantly affect the pattern of results, some observers reported having ‘misclicked’ for one or more trials, meaning they had pressed the mouse button accidentally having not made their gauge setting. The experiments in Chapter 6 took an average from three settings per location to address this, with previous studies such as Koenderink, van Doorn and Kappers (1992) finding that repeated settings made by a single observer are highly correlated. Here, the number of settings required across stimuli meant a single location setting was chosen to ensure timely data collection. Other work in this area should be mindful of this potential issue where smaller numbers of settings are made, meaning the effect of a missed trial would have a great effect on results.

All experiments showed evidence of significant vertical shearing. This is not something that was predicted, and is not of interest to the main hypothesis of the work. Koenderink, van Doorn, Kappers and Todd (2001), note a large variance in shearing in their results of a pictorial relief gauge figure task. They point out, however, that while the extent of shearing differed between observers, the shearing direction was similar, which suggests a similar interpretation of semantic image information. Other work has attributed shearing to the direction of illumination (Egan & Todd, 2015; Todd, Koenderink, van Doorn, & Kappers, 1996). However, future work could explore this experiment in a laboratory-based setting and see if the shearing still occurs. This is suggested as shearing in this way may be accounted for by a subtly different viewpoint. Todd, Koenderink, van Doorn and Kappers (1996) probed local surface attitude of torsos with a gauge figure task under varying viewing

location conditions. This was achieved by presenting 2D photographs of the same stimuli taken from different angles. They found that affine transformations accounted for 95% of the variance between viewing conditions, and conclude that shearing simulates a change of perspective, showing that changing viewpoint produces a significant effect on shearing. As these experiments were conducted online, there is no way of knowing if participants continued to view the stimuli from the same viewpoint for the duration of the experiments. This may be a physically different viewpoint, as used in the study described above, or a change in the 'mental eye' view, denoting a change in the perceived viewpoint pictured by the observer (Koenderink, van Doorn, Kappers, & Todd, 2001). Therefore, it cannot be ruled out that between blocks, and therefore between conditions, participants shifted their viewpoint just enough to influence the perceived location of the stimuli, although this alone would not account for the systematic nature of the finding.

As this experiment was conducted online, only one viewing distance was presented. The objects were presented at a simulated distance of 20cm. An interesting follow up to this experiment would be to see if these results are consistent across viewing distances, either real by way of moving the computer monitor, or by increasing the rendered distance of the objects in the scene, as pictorial cues are weighted more heavily at further distances when information from binocular disparity becomes less reliable (Viguier, Clément, & Trotter, 2001; Harris, 2004; Keefe, Hibbard, & Watt, 2011).

Given the finding of consistent under-reporting of depth relative to the gain, it is also possible that binocular information is playing a role here, and averaging the slope closer to that of binocular information which has a slope of 0, and would report

that the surface is flat due to the screen. Future work could explore this by presenting the scenes monocularly as well as binocularly, which would follow previous work that presented stimuli to one eye only (Koenderink, van Doorn, & Kappers, 1992). Given the strength of binocular disparity in the depth estimate due to its high reliability, especially at close distances, removing this cue could provide a way for pictorial information to be more highly weighted in the depth estimate and it would be interesting to observe if the contribution of this cue increases under those conditions.

8 Makeup filters narrow and elongate the appearance of facial features

8.1 Abstract

Makeup is a ubiquitous day-to-day application of manipulating perception, whether through reducing surface texture of wrinkles, to increasing contrast with eyeliner, to deepening cheekbones with contour. Digital filters that can digitally apply makeup to facial images via a smartphone app have risen hugely in popularity in recent years. The present work explored these filters in the context of a depth perturbation experiment, where we present faces with differing levels of digital makeup. Participants were asked to probe the local surface orientation using a gauge figure task. Depth gradients from the slant and tilt settings of the gauge allowed for reconstruction of facial meshes. We analysed both the full face and a localised zone of facial features on which the filter was applied. We found that depth ranges increased significantly with increasing level of digital makeup, but only in the localised features. This shows that makeup filters alter the apparent depth of facial features, but not the entire face. Additionally, we found that the shape of reconstructed meshes changed significantly with a filter applied compared to the no makeup condition. Taken together with the increased depth range, these results suggest that the highlight and contour filter enhanced the perceived depth of facial features through a narrowing and sharpening of the facial angles compared to the original model. These results show a clear effect of pictorial manipulation of digital makeup on the perception of depth and shape of facial features.

8.2 Introduction

8.2.1 Makeup and perception

Makeup has been used throughout history to manipulate perception. Egyptian women in around 3000 BCE blackened their eyelids and lashes with kohl made from soot, and Romans in the first century whitened their complexions with chalk (Danesi, 2018). Archaeologists recently unearthed a 2,700-year-old cosmetic face cream from the tomb of a nobleman in the Loess Plateau in China (Han, et al., 2021). Queen Elizabeth I of England was said to use a white lead-based make up to create the 'mask of youth' (Jain & Chaudhri, 2009). Nineteenth century British military officers wore face powder and reddened their cheeks with rouge to conceal any signs of fear from their subordinates (Carter, 1998).

The current day makeup world is a multi-billion-dollar industry (Eldridge, 2015). Cosmetic advertising promotes perfecting the physical appearance and critical gaze of the self and others (McCabe, Malefyt, & Fabri, 2017). Modern makeup products and techniques are driven by fashion and cultural values of appearance (Han, et al., 2021) and are used to cover unwanted facial features such as wrinkles or discolouration to increase perceived evenness of the skin (Batres, Russell, & Workowski, 2023), or to emphasise other features such as the eyes or lips (Scherbaum, et al., 2011). The application of eyeliner has been shown to increase the perceived size of the eyes (Morikawa, Matsushita, Tomita, & Yamanami, 2015). Additionally, accentuating the contrast between the eyes and lips and the rest of the face is seen to increase attractiveness (Jones, Russell, & Ward, 2015). Russell (2003) presents evidence that this relationship between enhanced luminance and

attractiveness is stronger for female faces, and therefore the present work chose to present female faces.

Makeup is also used to align the face more closely with cosmetic trends and standards of beauty. The golden ratio (shown in Figure 8.1), also known as the Fibonacci ratio, is found throughout nature, from flowers to shells, and can also be applied to the human face to denote a beautiful ratio of features (Singh, Vijayan, & Mosahebi, 2019).

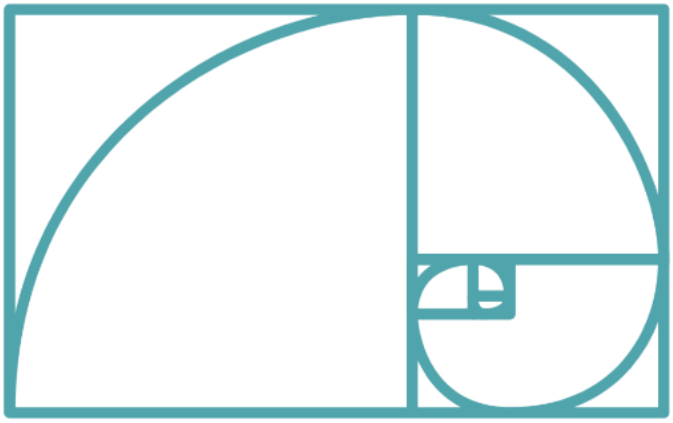


Figure 8.1: The golden ratio. Depicting the golden ratio, or Fibonacci ratio, which is found in nature and the principles of which are also applied to the human face.

The value of the golden ratio (here denoted as phi, φ) is calculated as follows:

Equation 8.1

$$\varphi = \frac{1 + \sqrt{5}}{2}$$

This leads to a value of approximately 1.618, and there is a long tradition in experimental psychology addressing whether spatial arrangements with this ratio optimise aesthetic appeal (Green, 1995; Russell, 2000). Choi and Baker (2022) outline ways in which cosmetics can be used to conform features to this golden ratio.

They describe the use of highlight and contour makeup to manipulate the facial features to more closely resemble the golden ratio. They define highlight as opaque pigmented powders or cream foundations several shades lighter than the rest of the foundation applied to the face, and contour as powders or foundations several shades darker to simulate shadows on the face. They state that the combination of these causes a receding effect of the structures to which this is applied. The present work uses this manipulation of highlight and contour makeup to manipulate the perceived 3D surface structure of faces.

Just as many cosmetic products are available for applying makeup, when making a judgement of depth, many cues are available for the brain to utilise. Cutting and Vishton (1995) outline these depth cues and discuss their individual considerations. Here, we focus on pictorial cues, which are cues derived entirely from the image on the retina. Specifically, the highlight and contour makeup are used in the present work to manipulate shape from shading (Brooks & Horn, 1985; Horn, 1990).

Chapter 4 introduced shape from shading in the context of the ‘Dark is Deep rule’, whereby surfaces locally distance from the light source appear darker, due to diffuse lighting conditions, such as from the sun on a cloudy day (Tyler, 1998; Cooper & Norcia, 2014; Todd, Egan, & Kallie, 2015). Additionally, the sun in a cloudless sky creates directional lighting whereby the surface luminance is dependent on the orientation of the surface relative to the light source (Hibbard, Goutcher, Hornsey, Hunter, & Scarfe, 2023). Application of makeup products such as highlight and contour can mirror the natural patterns of skin luminance created by an overhead light source, such as the sun, and deepen the shadows that would

naturally be created by different sizes or shapes of facial features, for instance, increasing the buccal hollow of the cheek to imply a more well-defined cheekbone.

The use of computer packages to apply makeup have become popular, both for makeup users and for researchers wishing to study the effects of makeup. Scherbaum et al (2011) present a system to automate the application of makeup to a professional standard. They used images of human faces with and without makeup professionally applied, and captured detailed facial information including surface normal, reflectance and specularity. This facial information was mapped in 3D to obtain positional information, and a best fitting makeup application was applied. They argue that makeup can be seen as changes in appearance as a function of the shape of facial features, meaning that the changes can be mapped and modelled. However, in their model they assume that makeup has no bearing on geometry and subsequently omit the surface normal information from their work. Here, the perceived direction of the surface is expected to be altered by the different shape from shading information provided by the makeup filter.

Tong, Tang, Brown and Xu (2007) present digital means to conduct ‘makeup transfer’, that is the process of applying makeup to a specification, such as before and after images a makeup artist might use to advertise their services. They outlined a process whereby the makeup from an example reference image could be transferred to a target image, preserving the facial features such as the eyes, skin colour and texture information of the target image, and only applying the makeup style. As with the work detailed above, this was achieved by mapping the surface geometry of the faces to a canonical facial model. These are the underlying principles behind the makeup filter application used in the present work.

The study in this chapter uses several levels of digital makeup to assess the change in perception with varying degrees of manipulation. The original face without a makeup filter is shown to provide a baseline against which to compare the makeup filter conditions. Two differing levels of makeup filter were created - a ‘subtle’ makeup filter, created by applying the highlight and contour filter with 50% opacity, and a ‘dramatic’ makeup filter, created by applying the filter with 100% opacity.

8.2.2 Measuring depth estimates for pictorial cues

As with the last few chapters, the local surface attitude of the faces was measured with a surface normal gauge figure task, where observers rotated a pushpin shaped figure until it was perceived as lying flat on the surface of the face in pictorial space, with the rod normal to the surface (Koenderink, van Doorn, & Kappers, 1995). As before, we implemented methods outlined clearly by Nefs (2008) and used successfully in previous works (Koenderink, van Doorn, & Kappers, 1992) to turn the slant and tilt settings captured during the gauge figure task into depth gradients, by which to measure the magnitude and orientation of perceived surfaces.

These gradients were used to create a best-fitting mesh surface from the data which allowed for comparison between the differing filter gain conditions. This was conducted using three main approaches. Firstly, we calculated the depth range as the difference between the closest and furthest pixels in the reconstructed meshes to give a measure of the change in depth between conditions. Secondly, we used affine transformation comparisons which have been used successfully in the preceding chapters to define the change in global shape meshes between conditions. Finally, we examined the surface roughness to look for an effect of makeup filter gain on

surface texture captured in the reconstructed meshes. This method, which was introduced in Chapter 5, uses equations designed for testing the implication of mesh simplification for highly detailed 3D models (Wu, Hu, Tai, & Sun, 2001), and is used again here to quantify the change in steepness of surface texture as a way to identify changes in the magnitude of depth across facial features. These analyses were completed both for the full face, and also for a localised feature zone of the face where the makeup filter manipulated the surface colour. The areas affected by the colour manipulation of the digital makeup filter are shown in Figure 8.2.

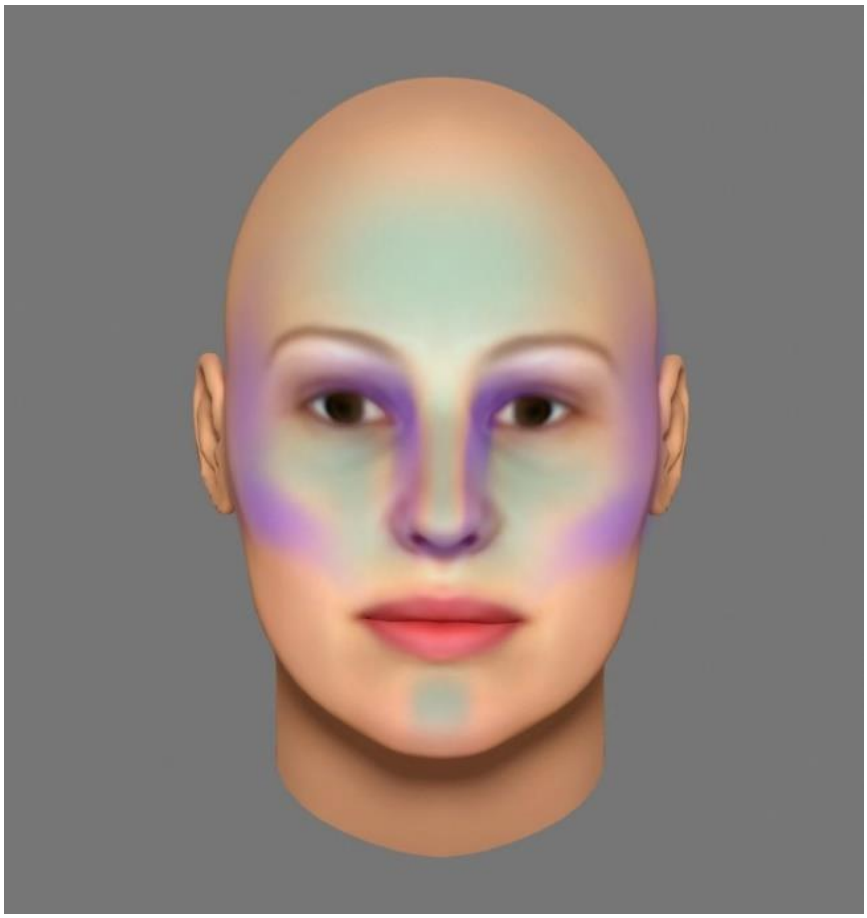


Figure 8.2: Highlight and contour. Showing zones where makeup effects are applied with the filter. Green shows highlight zones, purple shows contour zones.

Here, green shows the zones where the highlight effect was applied, whereby the pixels were lightened via the software adjusting the image colour in that region, and contour zones shown in purple whereby pixels were darkened. Colours for the makeup filter used in the experiment were selected using the original image as a base in order to appear more natural. More general background on these methods can be found in Chapter 2: Methods.

8.3 Methods

8.3.1 Participants

Nine participants aged between 28 and 63 took part in this study. Four (44%) identified as female, four (44%) male and one (11%) as a trans man. Participants included one researcher and seven people naïve to the purpose of the experiment. Naïve participants were recruited through word of mouth and were compensated financially for their time. All participants were required to have normal or corrected-to-normal vision (glasses or lenses) and were asked to wear glasses or lenses if they typically would do to use a computer.

8.3.2 Apparatus and stimuli

The 3D model of the face used in this experiment was created using FaceGen software (Singular Inversions Inc, Toronto), which created a 3D model mesh, as well as an image of the face. A Caucasian face was created with ‘feminine’ features, and a neutral expression. The face was rendered in the software Blender against a medium grey background, with a central light source placed diagonally above the face. The makeup filter was applied to the image of the face using FaceFilter Pro

(Reallusion, California). The filter consisted of a lighter ‘highlight’ element, and a darker ‘contour’ element. The areas of the face selected for each are predetermined within the software and were calibrated to the face by matching facial landmarks during calibration. The highlight colour was selected as the lightest value from the original image with an RGB value of (246,194,172), and likewise the contour colour selected was chosen from a value of an existing shadow within the original image, RGB (127,80,50). The filter was applied at three levels: 0%, 50% and 100% filter, corresponding to the filter’s opaqueness, shown in Figure 8.3.

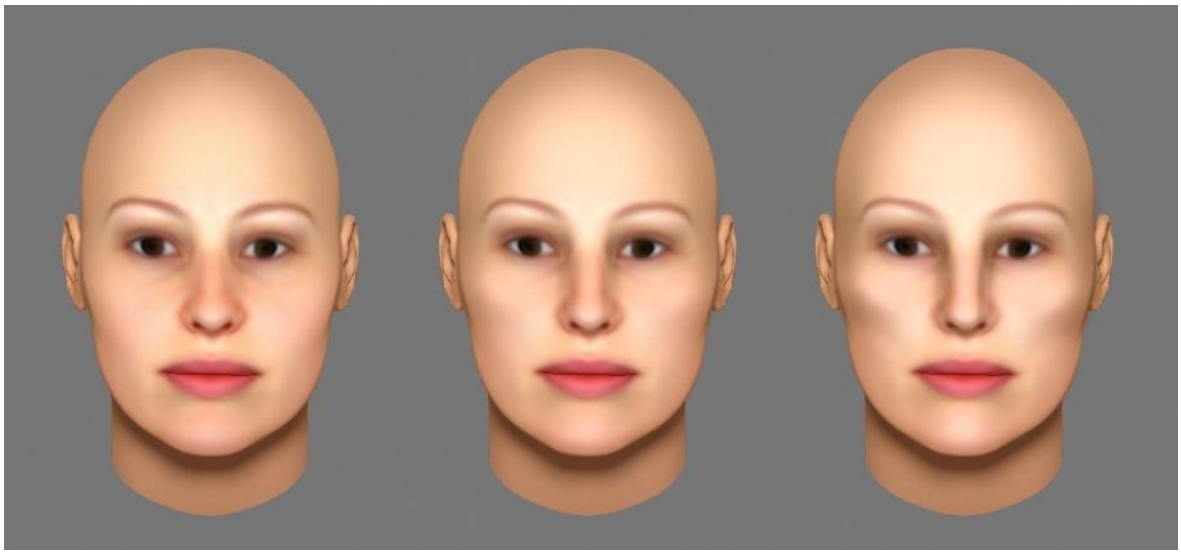


Figure 8.3: Face stimuli. Showing the face stimuli used in the experiment showing 0%, 50% and 100% opacity of highlight and contour filter applications respectively.

The gauge figures used in this experiment were white with a base diameter of 20 pixels. The possible locations for gauge figures are shown in Figure 8.4, which shows a triangulation mesh across an area defined by an ellipse. The centre of each triangle represents the barycentre, or a possible gauge figure location.

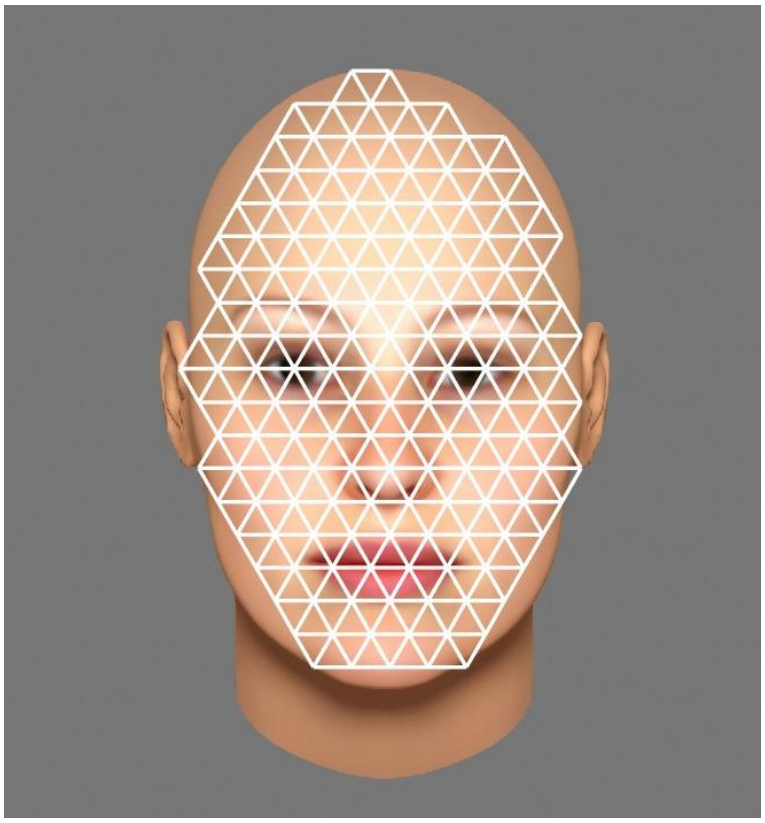


Figure 8.4: Gauge locations. Showing possible gauge figure locations defined by triangulation across an ellipse.

JavaScript code to render the gauge figures on the scene was adapted from work by Wijntjes and Van Zuijlen (2004). The experiment was created using the P5.js library (https://p5js.org/) in Qualtrics (https://qualtrics.com) and completed online using the participants' own desktop or laptop computer. Responses were recorded using the participants' computer mouse or laptop trackpad.

8.3.3 Procedure

The experiment began with a few practice trials to ensure participants understood the task. This was completed with a different 3D model of a tomato to ensure participants were not primed for the main task. Participants were asked to

capture the shape of the practice object by rotating the gauge figure so that the circle element appeared to be painted flat onto the surface of the object (Koenderink, van Doorn, & Kappers, 1995), with the pin element normal to the surface. This was achieved by rotating the mouse, and once happy with the placement, participants clicked the left mouse button to record their answer and move on to the next trial. Once participants felt happy they understood the task, they moved on to the experimental trials, where an estimate of surface slant was taken at each of the possible gauge figure locations in a randomised order. When all locations had been sampled the block ended and participants began a new block. In total there were nine blocks, corresponding to three blocks of trials each for 0%, 50% and 100% filter opaqueness. That is, the trials for each level of filter opaqueness were split across three blocks each to allow participants breaks if required. Blocks were presented in a random order and once all blocks had been completed, participants were debriefed.

8.4 Results

8.4.1 Data treatment

Raw scores were taken from Qualtrics of phi (slant) and theta (tilt) values linked to the x and y positions of the gauge figures. Given these were presented in a random order, these were first reordered for analysis. The nine blocks were combined into three data sets for analysis corresponding to the slant and tilt settings for each filter level. Following methods from Nefs (2008), slant and tilt settings were converted into depth gradients to give a measure of the direction and magnitude of the perceived surface of the face. From these depth gradients, meshes were then fit to the results using the face and vertices information with the observed orientations

using a best fit least squares deviation approach (Koenderink, van Doorn, & Kappers, 1992). An example can be seen in Figure 8.5. The faces are rendered as if looking upwards, with the left side of the forehead closest to the screen. They are clearly recognisable as faces and these coherent facial meshes show that observers made sensible settings.

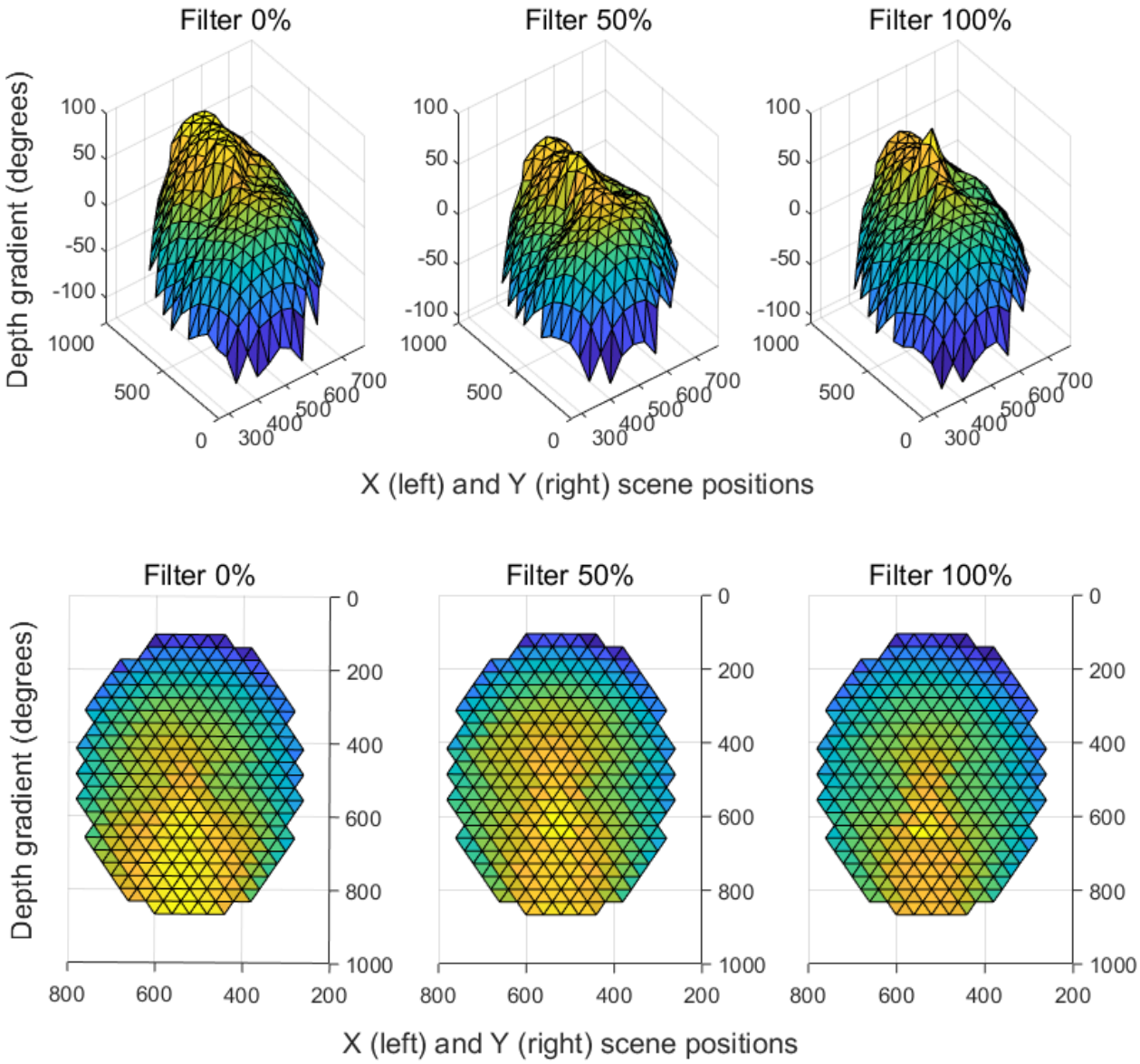


Figure 8.5: Reconstructed meshes. Showing example of the reconstructed meshes of faces perceived for each level of filter gain for one participant. Top row shows side view, and bottom row shows top-down view of the same data.

8.4.2 Global versus localised feature analysis

As presented in the introduction, the highlight and contour filters were applied to different zones of the face by the software. During analysis, the global percept was measured using gradient responses across all points of the face. Additionally, we explored localised perception across the main features manipulated by the filter, namely the cheeks and nose, by analysing just a subsection of the reconstructed mesh of the full face. Figure 8.6 shows the points isolated for the localised analysis.

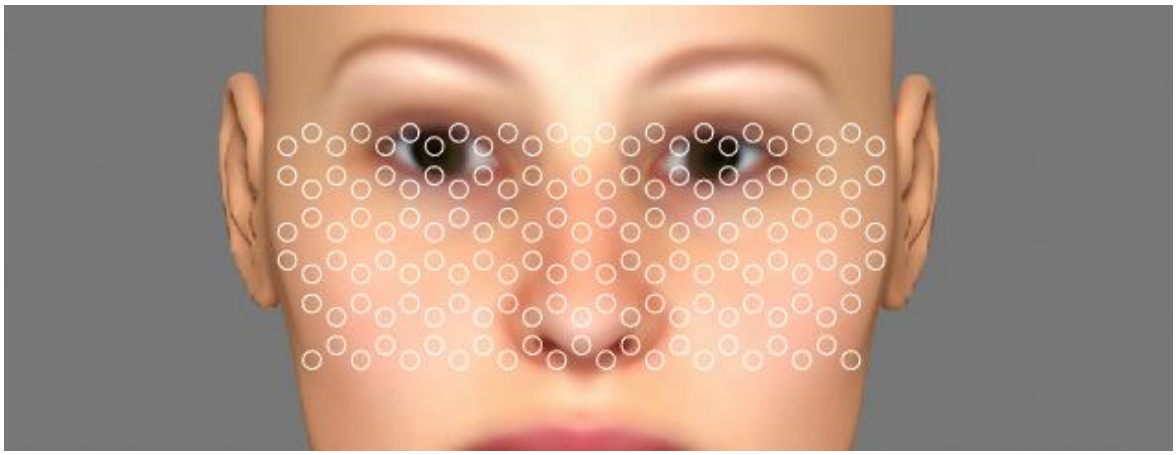


Figure 8.6: Localised feature analysis. Showing the points selected for analysis of the change in perception across a localised area including the nose and cheeks.

8.4.3 Depth range

To begin with, the relative depth ranges for each condition and participant were calculated as the difference between the closest and furthest pixels in the reconstructed meshes. As this study had a repeated measures design, a linear mixed effects (LME) model was used to compare these depth ranges across conditions, first of all for the full face, as follows:

Equation 8.2

$$dr \sim 1 + g + (1 + g \mid o)$$

This model evaluates how well the relative depth range (dr) could be predicted by the makeup filter gain (g) as a categorical variable, grouped by observer (o). Filter gain was not found to predict the depth range of reconstructed meshes for the full face for either condition, as shown in Table 8-1.

Table 8-1: LME results for depth range full face analysis.

Fixed effect coefficient	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
Filter gain (g) comparing 50% (g _2) to 0% (g _1)	0.001	0.043	24	.986	-0.088	0.090
Filter gain (g) comparing 100% (g _3) to 0% (g _1)	0.070	0.042	24	.105	-0.016	0.156

The depth range of the localised features was also calculated, with a comparison between the mean (SE) depth ranges for each condition for the full face and localised features analysis shown in Figure 8.7.

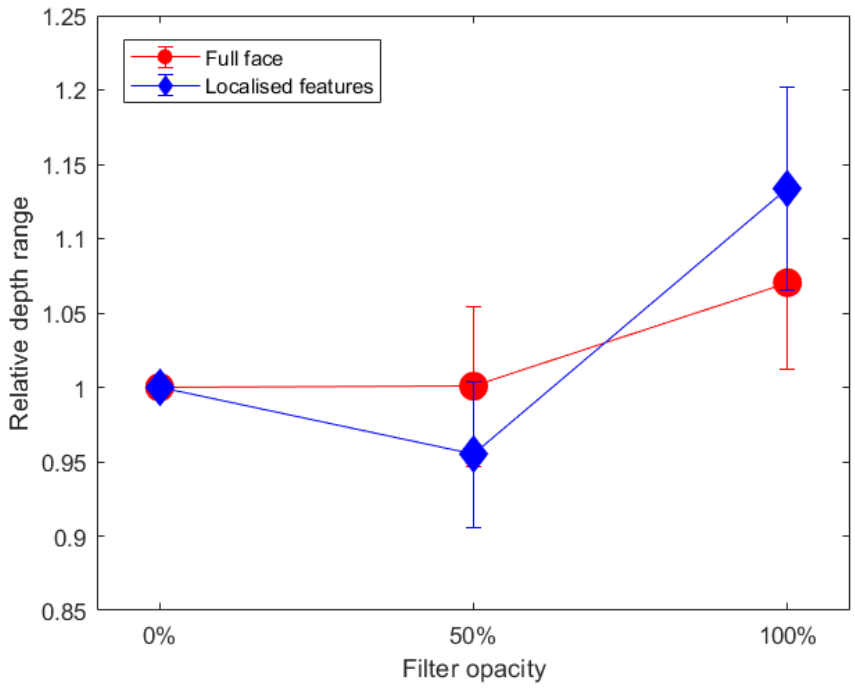


Figure 8.7: Relative depth range comparison. Showing the mean (SE) relative depth range for the full face and localised feature analysis at each filter level.

A much steeper overall effect of filter gain on relative depth range can be seen for the localised analysis compared to the full face. Here, the depth range for the 100% filter opacity condition can be seen to be considerably larger than the other two conditions. An LME explored these effects using Equation 8.2, with results shown in Table 8-2.

Table 8-2: LME results for depth range of localised features.

Fixed effect coefficient	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
Filter gain (g) comparing 50% (g_2) to 0% (g_1)	-0.069	0.038	24	.087	-0.148	0.011
Filter gain (g) comparing 100% (g_3) to 0% (g_1)	0.164	0.049	24	**.003****	0.062	0.267

Filter gain was found to have a significant effect on the depth range for the 100% filter gain condition, with the full highlight and contour shading manipulation increasing the perceived depth of features by around 16%.

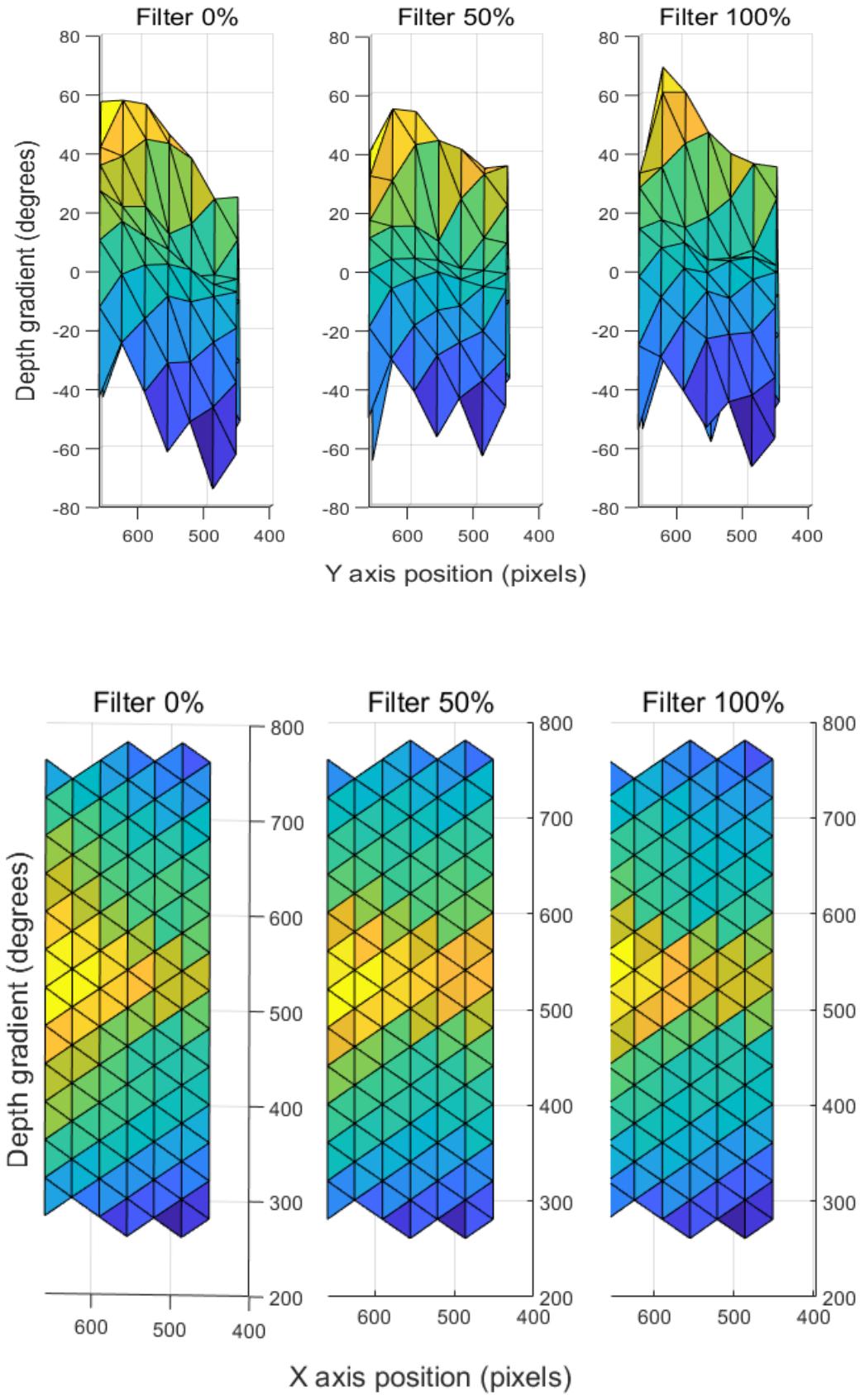


Figure 8.8: Noses in profile. Showing a profile view of the faces at each filter gain condition to show how the filter affects the depth of features. Top row shows a side

view of the nose profiles, and the bottom row shows a top-down view of the same data.

This increase in depth range can be seen clearly in Figure 8.8, which shows the localised feature zone in portrait. This shows how the nose extends further in depth with the 100% filter condition, as well as being less rounded than the no makeup condition.

8.4.4 Affine transformations

As with previous chapters, the reconstructed meshes were compared through affine transformation to quantify the change in perceived shape with filter gain. This explored if the difference in relative depth could be explained by a change in the shape of the perceived surface. Here, the filter gain conditions were compared to the 0% filter settings as a reference, as this represents the original image before the makeup filter application. The affine transformation regression produced transformation estimates for four main parameters: the intercept, x and y which represent shearing along these axes, and z which represents the relative depth of the models. The means (SE) are plotted in Figure 8.9.

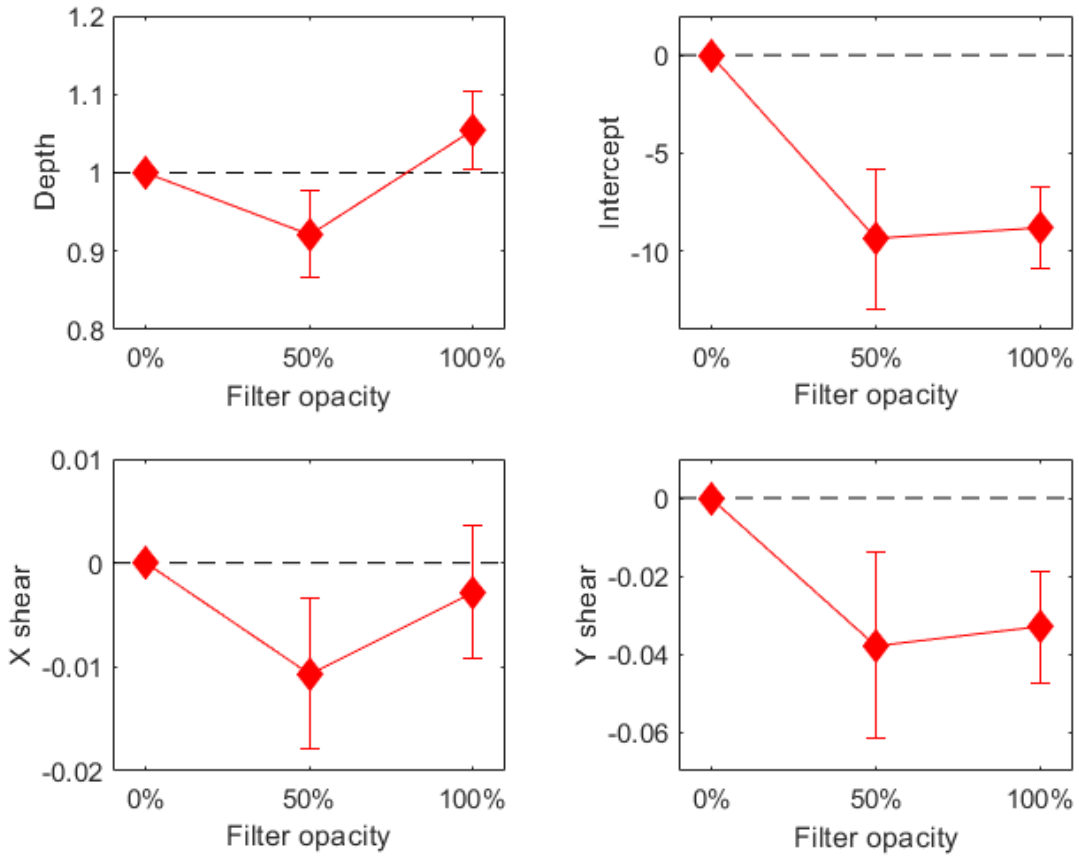


Figure 8.9: Affine transformations full face. Showing plots of mean (SE) regression slope values for the four parameters from affine transformation for the full face analysis.

As this study had a repeated measures design, a linear mixed effects (LME) model was fit to the data to explore the effect of gain on the four parameters, using the following equation:

Equation 8.3

$$b \sim 1 + g \ + (1 + g \mid o)$$

Here, how well the gain on pictorial cues (g) predicts the beta slope values of each parameter (b) is quantified, grouped by observer (o). This model contains random intercepts and slopes for gain.

Table 8-3: Depth range ANOVA results affine transformations full face analysis.

Coefficient	Model	F	DF	*p* Value
bz	$bz \sim 1 + g + (1 + g \mid o)$	4.47	2,24	**.022***
b0	$b0 \sim 1 + g + (1 + g \mid o)$	17.20	2,24	**<.001*****
bx	$bx \sim 1 + g + (1 + g \mid o)$	2.05	2,24	.151
by	$by \sim 1 + g + (1 + g \mid o)$	7.13	2,24	**.004****

A negative main effect of filter gain was found for the intercept parameter, with estimates of -9.38, (***p*=.001**, 95% CI [-14.61 -4.16]), for the 50% filter and -8.79, (***p*<.001**, 95% CI [-12.47 -5.10]) for the 100% filter conditions showing the filter faces were seen as around nine pixels further back in depth than the 0% filter condition. No shearing in x was observed, but there was a significant negative effect of filter gain on shearing in y, with estimates of -0.037, (***p*=.039**, 95% CI [-0.073 -0.002]), at 50% and -0.033, (***p*=.010**, 95% CI [-0.057 -0.008]), at 100% showing between 3 and 4% shift in the vertical position with filters. A small overall main effect of filter gain was found for relative affine depth transformations, meaning that the relative depths differed between the filter conditions, but the direction of relative depth transformation differed between the 50 and 100% filter gains, with neither the 50%, (*p*=.075), or the 100%, (*p*=.155), level alone significantly predicting the depth transformations.

Next, the difference in shape was compared for the localised features using affine transformations. The mean (SE) transformations for each of the four parameters are shown below in Figure 8.10.

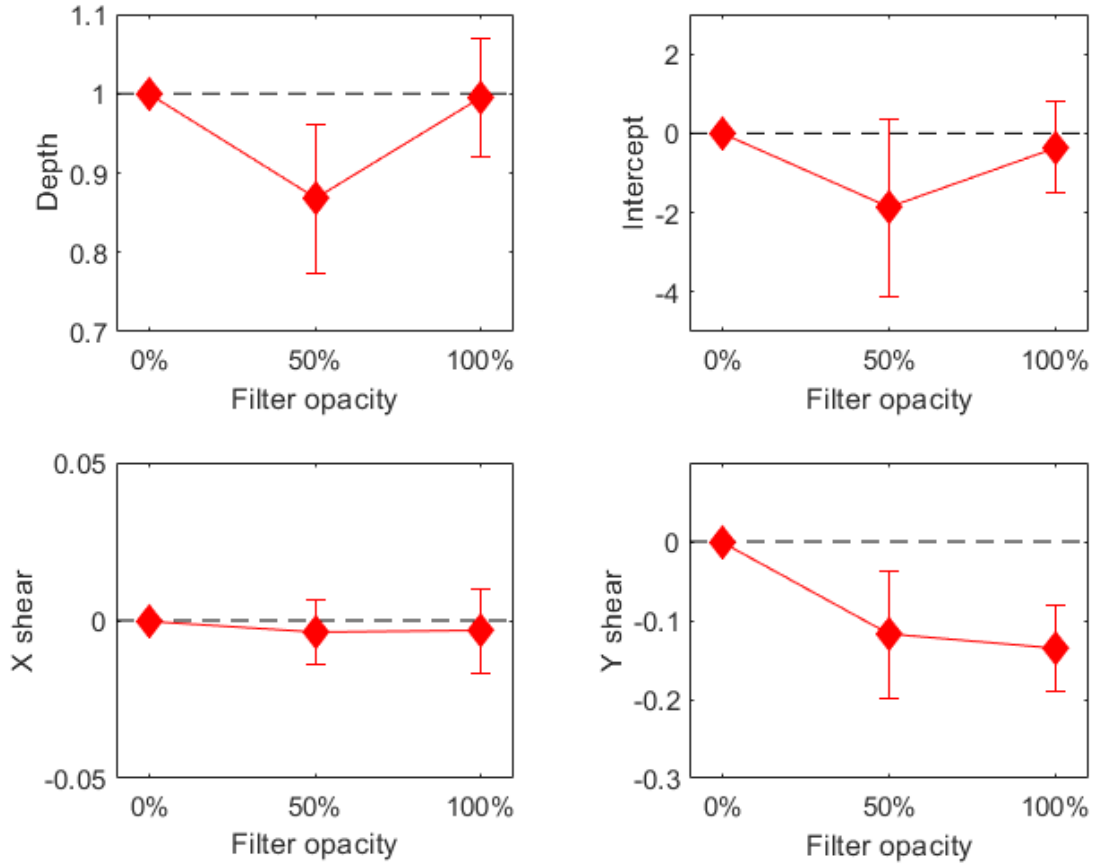


Figure 8.10: Affine transformations. Mean (SE) affine transformations for localised features for the four parameters. Black dashed lines show null hypothesis.

The affine transformation slopes were compared between conditions using the LME equation in Equation 8.3. The results are shown in Table 8-4.

Table 8-4: Depth range ANOVA results affine transformations localised feature analysis.

Coefficient	Model	F	DF	*p* Value
bz	$bz \sim 1 + g + (1 + g \mid o)$	3.45	2,24	**.048***
b0	$b0 \sim 1 + g + (1 + g \mid o)$	1.43	2,24	.259
bx	$bx \sim 1 + g + (1 + g \mid o)$	0.17	2,24	.846
by	$by \sim 1 + g + (1 + g \mid o)$	7.64	2,24	**.003****

Filter gain was found to significantly predict affine transformation of depth between the filter gain conditions for the localised feature analysis. As before, this proved to be a small negative main effect, rather than being explained by either the 50% filter gain alone, (*p*=.067), or the 100%, (*p*=.921). A significant effect on vertical shearing was also observed again, with estimates of -0.12, (***p*=.036**, 95% CI [-0.23 -0.01]), at 50% and -0.13, (***p*=.001**, 95% CI [-0.21 -0.06]), at 100% showing a much bigger shift of between a 12 and 13% in the vertical position with filters compared to the full face analysis.

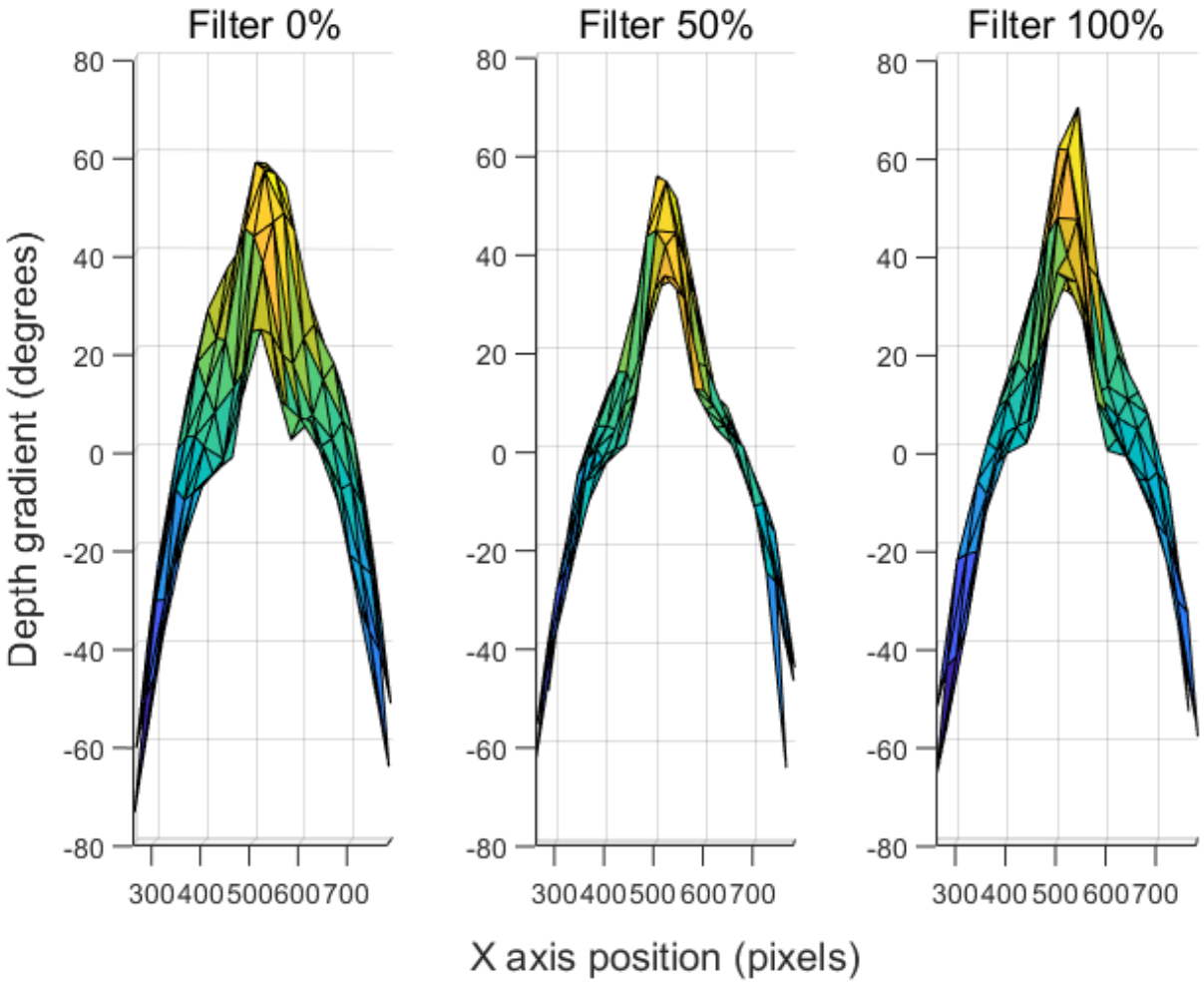


Figure 8.11: Noses from above. Showing the localised features analysis as if viewed from above.

Figure 8.11 shows the clear change in shape with filter gain, where the features can be seen to narrow and elongate between the no makeup and makeup filter conditions.

8.4.5 Surface roughness

The change in the magnitude of depth across the face was quantified with a surface roughness measure used previously in Chapter 5. This calculated the average angle of the gradient between a surface face and the ones surrounding it, giving a measure of how shallow or steep the transition between surfaces was. This was used here to examine how filter gain affected the steepness of facial features. The mean (SE) surface roughness in degrees for the full face analysis is shown in Figure 8.12.

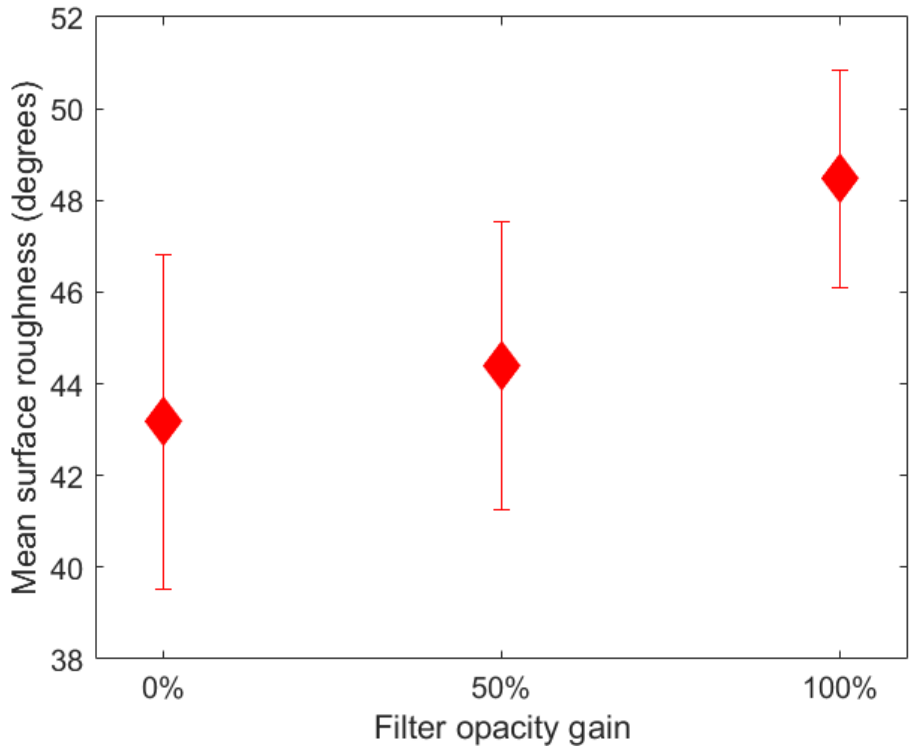


Figure 8.12: Surface roughness full face analysis. Showing the mean (SE) surface roughness in degrees for each filter gain condition for the full face analysis.

An LME was fit with the following equation, to see if surface roughness (r) could be predicted by filter gain (g), grouped by observer (o):

Equation 8.4

$$r \sim 1 + g + (1 + g \mid o)$$

This found a significant effect of filter gain on surface roughness for the 100% condition, accounting for an increase in facial steepness of around 9%, but no effect for the 50% condition. Results are shown in Table 8-5.

Table 8-5: LME results for surface roughness full face analysis.

Fixed effect coefficient	Estimate	SE	DF	*p* Value	Lower CI	Upper CI
Filter gain (g) comparing 50% (g _2) to 0% (g _1)	0.021	0.037	24	.571	-0.055	0.098
Filter gain (g) comparing 100% (g _3) to 0% (g _1)	0.092	0.037	24	**.020***	0.016	0.169

These results show the more dramatic filter resulted in increased global steepness for facial features, but the subtle filter did not result in any global change. Next, the surface roughness of the localised facial features analysis was compared. The mean (SE) surface roughness in degrees for the filter gain conditions under the localised feature analysis is shown in Figure 8.13.

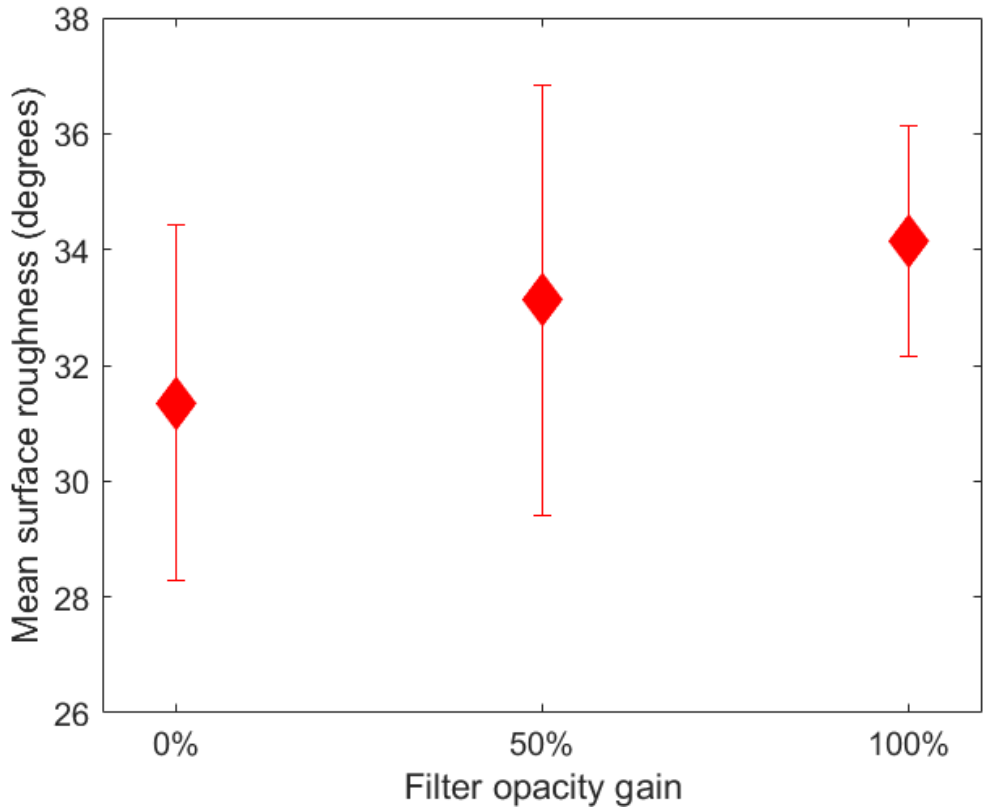


Figure 8.13: Surface roughness localised features. Showing the mean (SE) surface roughness in degrees for the filter gain conditions for the localised analysis.

The difference in surface roughness between conditions was explored with Equation 8.4. Filter gain was not found to significantly predict surface roughness for the localised features, ($F(2,24)=2.13$, $p=.141$).

8.5 Discussion

8.5.1 Findings

This study provided a practical application of the research methods outlined in this body of work. The gauge figure and cue gain methodologies developed over the previous chapters were used to explore a real world depth manipulation application in the form of makeup filters. Here, we discuss the findings. To begin with, the

reconstructed meshes provide a clear indication that the methods successfully capture the shape of surfaces as the face is clearly visible from the results.

Next, the depth range results were compared. No global effect of filter gain on depth range was found for the full face analysis. This finding is not unexpected, given that makeup is designed to enhance features, not the depth of the entire face. When we considered the localised feature analysis, we observed a significant effect of filter gain on depth range, measured at 16% for the dramatic 100% filter, indicating that applying the filter resulted in facial features such as the cheekbones appearing 16% deeper compared to the no-makeup condition.

Considering the results of the affine transformations, a small effect of filter gain was found both at a global and localised level, indicating that the filters subtly changed the global shape of the face. Specifically, a reduction in global shape between the no makeup condition and the filter conditions was found, which could indicate that some regions were perceived as receded compared to the no makeup condition. For instance, contour makeup around the nose is intended to ‘thin’ the appearance of it, which is supported by the finding here that the depth of the shape reduced with the makeup filter applied. A significant effect on shearing in y was observed for both the global and localised analysis, further supporting the conclusion of subtly different perception of the shape of faces between makeup filter conditions. As reconstructed meshes were normalised to a standard distance, the significant intercept in the global condition suggests that faces with the filter applied were viewed as slightly further back.

Finally, the surface roughness results are considered. Surface roughness was used in this work to evaluate the steepness of the change between surface faces, as

a way to quantify the change in facial features between conditions. Here, we found a significant effect of filter gain on mean surface roughness of the reconstructed meshes. This suggests that increasing filter gain resulted in greater depth gradients between points, resulting in sharper facial features than the original no-makeup condition. No effect of filter gain on surface roughness was found for the localised features. Given the large standard error this could be due to the variability in the data, or it could point to the observed surface roughness in the global face analysis resulting from other facial features not included in the localised feature analysis, such as the chin. Future research could split analysis zones even further into each individual facial feature to explore this further.

Taken together, these findings show an enhanced range of depth for faces with makeup filters applied, and a reduction in the depth of the shape of faces around certain features. This pattern of results suggests that the application of the filters results in facial features that appear narrower and deeper. For example, Figure 8.14 illustrates how the nose, as if viewed from above, may have been affected by the filters such that it exhibits a deeper range of depth after the filter is applied, as shown in turquoise, as well as a reduction in the depth of the feature shape, as shown with the purple arrows. Here, the nose is perceived as narrower and longer due to the shape from shading manipulation applied with the filter, resulting in increased depth range, but a reduction in the relative depth from affine transformations. Figure 8.8 and Figure 8.11 clearly illustrate this effect from the data on the appearance of the nose, which is seen to narrow and elongate with the makeup filter applied.

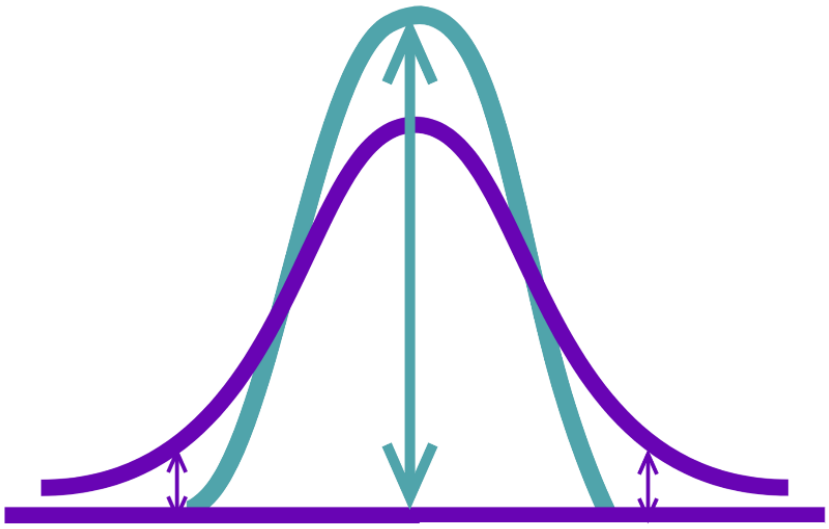


Figure 8.14: The change in depth and shape of the nose. Illustrating how facial features such as the nose may exhibit an enhanced depth range and decreased depth of shape. The noise is viewed as if from above. Purple shows the nose before the filter is applied, and turquoise shows the perception with the filter.

Our findings present clear evidence that makeup does have a geometric effect, contrary to the suggestions of Scherbaum et al. (2011). In general, results differed substantially between the subtle and dramatic makeup filter conditions. Some difference between these conditions was expected, given that the subtle makeup look represented 50% of the manipulation of the dramatic condition. However, these sometimes exhibited polarised patterns of results. As the makeup industry covers a range of looks, from ‘natural’ makeup looks through to dramatic runway makeup for a fashion show for instance, future work could expand on this by testing more levels than were used in the present work. Other considerations for future studies are discussed below.

8.5.2 Future studies

This experiment utilised technology in the form of a filter to manipulate the surface colour to apply the highlight and contour. An obvious next step would be to extend this research with the addition of real applications of makeup, as well as using digital methods to compare the two. In conjunction with this, comparing filters to real makeup applications by professionals as well as by non-professionals would be an interesting avenue for future work, given that professionally applied makeup has been found to be more effective than self-applied (Batres, Porcheron, Courrèges, & Russell, 2021), so it would be interesting to see what the performance of digital makeup filters is when compared with these.

Given that the focus of this work was to measure the contribution of depth cues to perception, the data and analysis is based around measurements of depth. However, more qualitative data could be gathered in future studies to assess the 'believability' of makeup – that is, exploring how much shadows can be manipulated before people are aware that it is a makeup filter, rather than natural shape from shading.

In addition, this work was completed using a face with 'female' characteristics within the face generating software. Future work could explore the impact of sex or gender in this paradigm. Likewise, a neutral expression was chosen for the study presented here, but results with differing facial expressions would provide further areas to explore.

Whilst setting up the experiment, it became apparent that for the face with the 100% filter application, the makeup applied was far more believable when the image was viewed full size than when it was presented as a thumbnail size. This is an

interesting area of study that future work could pick up, to explore the limits of efficacy of these filters, or with a real application of makeup.

The face in this experiment was rendered without any hair to keep the shape from shading information limited to the shading aspects of interest, rather than adding additional shading information from hair falling around the face. An interesting follow up to this work would be to present the face with and without hair, to see if this additional information, such as additional shading to the sides of the face or cheek, may affect perceptions of depth from the contour application.

When applying the makeup filters in the software, only the elements for highlight and contour were retained in the filter as the elements of interest in this experiment, but many other aspects of makeup were available. The addition of other elements of makeup, such as blush, may alter the effect of the highlight and contour manipulation, given that these elements are often used in the context of a full face of makeup. Exploring how these interact could be an interesting avenue to explore.

A final area future research could explore is to manipulate the filters in reverse, that is to lighten the contoured areas and darken the highlight. In doing this, the colour and shading information would be directly in conflict with other available cues, and the weighting for these cues could be measured by varying the degree of opacity using a similar method as this work and observing the difference in slope values.

9 General discussion

9.1 Review of thesis aims

The general aim of this thesis was to address some issues highlighted within previous depth cue combination research. Namely, few studies provide information about the weighting of cues beyond relative terms. Additionally, cue combination studies typically present either overly-simplified stimuli, stimuli under overly-simplified cue-restrictive viewing conditions, or both. The studies collated here were designed to address these issues. A summary of the main findings in this context are presented below.

9.2 Summary of main findings

Chapter 1 outlined cue combination and provided commentary on criticisms of current models of depth perception. This was examined in the context of wider criticisms within the field of visual perception, in that many traditional methods are not geared to natural viewing conditions, oftentimes due to technical challenges such as accurately recreating or measuring naturalistic stimuli. Additionally, the discussion here was mindful of the fact that we perceive surfaces, not points in space, as was previously implicitly assumed in many experimental designs using simple dots or lines.

Chapter 2 addressed the above by outlining the methods used to explore cue contribution in complex natural scenes. It detailed the use of advanced technology to capture and recreate scan data for mesh model creation, in order to present rendered realistic stimuli in complex natural scenes. The use of such methods

afforded many benefits over more traditional methods, such as the ability to directly manipulate the objects for cue perturbation, in terms of, for instance, the shape and surface colour.

The experimental work began with Chapter 3 where the natural starting point for this thesis was to adapt a traditional experimental design, to begin exploring how biases in depth cues can be introduced. This study used traditional methods, such as a '2 Alternative Forced Choice' task (2AFC) and stripped back all extraneous depth information in order to assess the single cue of vergence in isolation. Additionally, it used traditional stimuli, here simple red dots and lines on a black background. These were utilised to investigate the level of noise in the cue of vergence to denote how certain participants were of relying on this cue, and exploring how this related to shape constancy. The results replicated well-known psychophysical effects, such as imperfect distance scaling, but did not find evidence to support the theory that shape constancy is related to noise within vergence estimates and therefore certainty of vergence. While this work provided insight into the contribution of the vergence cue to binocular depth perception, the simplistic stimuli did not provide information about how we view objects in complex natural scenes, something which was remedied in the next chapter.

Having manipulated a cue in isolation using the traditional simplistic stimuli, the work moved on to exploring more naturalistic stimuli with multiple cues present in Chapter 4. This was achieved by using 3D renderings of real scanned objects and perturbating cues within the context of many others present, in order to measure their contribution to the depth estimate as part of the cue combination model. The work here followed up on research conducted by Hibbard et al. (2023), which found

evidence of the shape from shading ‘Dark is Deep’ rule, which states that surfaces locally furthest away with appear darkest. The study in Chapter 4 expanded on this work by presenting stimuli at two distances, to examine the weighting given to pictorial cues at the further distance, where binocular cues are known to be weighted less due to reduced reliability. While this chapter failed to replicate the study upon which it was based, due to a lack of luminance effect at the closer distance, results did show evidence of the ‘Dark is Deep’ rule at the further distance, as predicted. The use of rendered 3D models of real scanned objects over the use of simplistic stimuli, allowed for experimental control needed to probe the research questions, but also provided realistic viewing conditions to explore how we view things and the contribution of cues in complex natural scenes.

The previous chapter provided insight into the contributions of cues to the overall depth estimate of a complex natural scene, and measuring these by purposely introducing bias. However, as discussed previously, Gibson (1950) states that we perceive surfaces, not unconnected points, so Chapter 5 built on this previous chapter by introducing methods designed to probe the local attitude of surfaces of natural objects. When designing the work, questions arose about how these methods capture the shape of objects and what parameters were important. As such, the work in this chapter explored the impact of scale and context on local and global judgements of the surface texture of 3D objects in complex natural scenes. Further consideration in this chapter is taken to link global and local viewing to the Bayesian cue combination model, and the importance of order of stimulus presentation. Results found that both gauge size and sample location distance are important when designing a gauge figure task, with significantly more surface texture being captured with a smaller gauge sampled at points spaced closer together.

Results also found flattening of the percept with reduced context, providing supporting evidence for the inclusion of complex natural scenes in studies of this kind.

Having established the naturalistic stimuli and complex natural scene experimental design methods, the work in Chapter 6 began to measure the contribution of an individual cue to the weighted depth estimate under cue combination. One criticism of experimental methods often used to study this model is that they are unable to measure bias, and therefore all cues are assumed to be unbiased, although this is not what is observed empirically. In this work, bias was purposefully introduced in order to measure the contribution of binocular disparity to the overall weighted depth estimate. This was achieved by rendering the objects to simulate a larger or smaller IOD than the observers' own, which effectively manipulated a gain on the cue of disparity, while keeping information from other cues constant, and judgements measured with a gauge figure task. This effect was measured for both a single object, and a complex cluttered scene. Binocular disparity was found to contribute around 9% weighting to the depth estimate in the single object condition, with the additional context information raising this to 11% for the cluttered scene. This shows a modest direct weighting of binocular cues, increasing with the addition of other information as predicted, and further supports the message of this work for the inclusion of complex, natural scenes to replicate true-to-life viewing.

Chapter 7 followed on from the previous chapter, this time measuring the contribution of pictorial cues, namely the occluding contour and shape from shading. Here, rendered objects were either squashed or stretched in depth, with a gauge

figure task used to probe surface perception. Results did not find evidence that observers perceived additional depth with stretching or squashing the pictorial gain. Therefore, the occluding contour and shape from shading alone were not found to convey metric depth.

Finally, Chapter 8 took results from the previous chapters and explored their application using a real-world example of depth manipulation – makeup. This provided an excellent opportunity to explore how people perceive natural scenes, here in the context of faces, when pictorial cues have been deliberately manipulated. Digital makeup filters were used in order to convince observers of differing facial geometry, for example by applying a darker shade to a cheekbone to give the appearance of a deeper hollow to the cheek. The work in Chapter Four looked at how people perceive extra shading as extra depth as per the 'Dark is Deep' rule. Here, the work was interested in discovering if people view the extra shading as extra depth in the context of makeup, or if the awareness that a face has had makeup applied provides Bayesian knowledge that would affect the perception of the surface. Results show a clear narrowing and elongating of facial features as a result of the makeup filter application, with the filter increasing the perception of depth by around 16%. Additionally, the reconstructed meshes provide a convincing and distinctly-recognisable reconstruction of the shape of the face, providing evidence that the methods used in this thesis are well fit to the problem they were designed to investigate.

However, while the gauge figure method was a successful technique used here, it is not without its limitations. Monocular rendering of a gauge figure over a stereoscopic stimulus has been found to result in an overestimation of slant

(Bernhard, Waldner, Plank, Soltészová, & Viola, 2016). This may be relevant for the work in Chapter 6, although largely settings were congruent with the veridical slant of the 3D object. Additionally, as gauge figure tasks probe pictorial space (Koenderink, 2012), there are a number of factors that may affect the estimate. Pictorial relief is said to have affine freedom (Nefs, Koenderink, & Kappers, 2005), meaning that interpretation may differ in any direction within pictorial space. For instance, differences in pictorial reliefs captured between sessions or differing participants may be as a result of a difference in mental viewpoint (Koenderink, van Doorn, Kappers, & Todd, 2001). All of these factors may have affected the capture of perceived depth, and future work could combine a gauge figure task with other measurements to explore these further.

Similarly, when analysing the slant and tilt values garnered using the gauge figure method, some caution needs to be taken. For instance, in converting slant and tilt settings into gradients, large settings can result in extreme gradients due to the exponential nature of the relationship, which skew the rest of the data. This was mitigated in this work by applying an 85 degree gradient limit, which reduced the effects of these extreme gradients. Analysing the gradients between the reference and target meshes with affine transformations likewise has limitations. As mentioned above, pictorial space has affine freedom, which means that the resulting model created using perceptions of the slant and tilt of the surface may differ not just in depth on the z axis, but also in the x and y dimensions too which account for horizontal and vertical shearing respectively. Research shows that including these x and y shears accounts for more variability in the data compared with linear transformations, which look at depth alone (Egan & Todd, 2015; Koenderink, van Doorn, Kappers, & Todd, 2001). While the depth transformation was the main

interest of this work given that the gain manipulation was expected to affect depth, future studies in this area could explore the horizontal and vertical shearing effects further. For instance, x shearing may be interpreted as a differing mental viewpoint (Koenderink, van Doorn, Kappers, & Todd, 2001), and y shearing may be related to changes in vertical disparities, or could show that observers interpreted the gain manipulation as a rotating of objects rather than a stretching in depth, as seen by the significant y shearing in Chapter 7.

This work explored a number of depth cues to calculate their individual weightings within the cue combination estimate. The cue perturbation methods used here manipulated perception using large gains, such as half or double the original stimuli. These extreme gains were essential to the design of the work, to pick up on the relatively small effects by using large values. Research has shown that cues are sometimes no longer integrated with extreme conflicts (Cheng, Shettleworth, Huttenlocher, & Rieser, 2007; van Ee, van Dam, & Erkelens, 2002). This relates back to the work of Girshick and Banks (2009), who explored the idea of robust averaging. Where these extreme conflicts were observed, robust averaging of cues discounted the outlying data point, resulting in a depth estimate that does not include the conflicted cue. In this context, it was expected that large gains may result in a robust averaging effect. To address this, gains were presented in increments to allow for an effect to become evident, such as in Chapter 6 where the effect of a gain on disparity is observed for levels less than 1, but not linearly across all gains. Therefore, where no effect has been observed, we can be confident that our methods would have picked it up, were there one there to be found.

In fact, due to the very nature of the design of this work, these exaggerated gains were essential. In many previous studies in this field, for example the work of Hillis, Watt, Landy and Banks (2004), who explored cue combination for slant perception from disparity and texture. In their work they manipulated the cues to have equal weighted, which in this case meant 50% equal weighting in the estimate. This meant that when one cue was manipulated, they could measure the relative weightings by observing the estimate and seeing which cue was accounting for the higher weighting.

However, in the studies in this thesis, many more cues are available on which to base estimates, and their natural weightings may be very low. For instance, were a cue to provide just 10% of the depth estimate under natural full-cue viewing, and we manipulated an object of 10 cm by a gain of 1.1, this should be cause the object to be 10% bigger, but as the cue is only providing 10% of the estimate, the difference of 11mm would almost certainly be too small to be picked up as a statistically significant change. As such, much greater gains were used to ensure the natural weightings of the cues did not prevent an effect from being measured.

Overall, while binocular cues were found to bring some sense of metric depth from the scaling of vergence, the findings from the studies presented here predominantly show evidence of observers recovering a purely pictorial relief, with an arbitrary depth scaling, rather than scaled metric depth.

9.3 Laboratory versus online data collection

Having discussed the broad findings, some thought is taken to the type of data collection used. The studies presented within this body of work represent both online

and laboratory-based data collection methods. Originally, the intention had been to carry out all experimentation in person under laboratory conditions. However, like many, we were forced to adjust our work plan due to the Covid19 pandemic. While online data collection provided the only means through which to continue with the research at the time, it also opened up opportunities, which will be discussed here.

9.3.1 Pros of online data collection

Online experiments are used successfully in many areas of psychological research, allowing for quick, convenient data collection, with a bigger and broader range of participants than is practical with physical data collection (Reips, 2000). Utilising online methods also removes physical restrictions researchers often experience, such as availability of technology in a shared laboratory, as well as allowing for concurrent participants, which speeds up the data collection stage (Gosling, Vazire, Srivastava, & John, 2004).

While many areas of psychology are naturally suited to an online environment, psychophysics has traditionally been contained within, often darkened, laboratories. Semmelmann and Weigelt (2017) present work exploring a number of well-known psychophysical effects through online data collection. They ran experiments either in the laboratory using ‘Gold Standard’ control, in the laboratory using the online data collection software in their web-in-lab condition, or via the internet. The results of the online studies were found to be well matched to those collected in the laboratory and the web-in-lab conditions, supporting the use of online data collection in psychophysical experiments. They did note some possible effect of web browser on reaction times, but this is not something on interest in the work specified in this

thesis, so would not impact online collection for similar studies. Likewise, these effects have been found not to affect results for within-participant experimental designs, such as the work presented here (Reimers & Stewart, 2015).

9.3.2 Cons of online data collection

Although the benefits of continuing the research during the pandemic far outweighed the cons of online data collection, there were a few downsides, the implications of which, and possible solutions for, are discussed here.

Despite online methods allowing for a greater reach for the participant pool, the lack of experimenter control in online studies compared to laboratory-based work has been found to lead to substantially noisier data. Hirao, Koizuma, Ikeda and Ohira (2021) conducted an investigation into online data collection for perception studies, specifically the impression of faces. They completed the same experiment both online and under laboratory conditions, where participants studied facial photographs and completed questionnaires of impression measures such as trustworthiness and attractiveness. Results from the online data collection were far noisier and of reduced quality compared to the laboratory conditions. Nonetheless, the researchers do report that this can be offset using very large participant numbers. From their modelling, 2000 observers reduced the effect of increased noise in the results (Hirao, Koizumi, Ikeda, & Ohira, 2021). Large numbers are often required to combat the so-called replication crisis, where effects are not replicated between studies due to small observation numbers. However, numbers of these sizes are unlikely to be required for psychophysical experiments, which traditionally have far fewer

participants complete far more observations, with the observer themselves acting as the replication agent (Smith & Little, 2018).

One difference between the lab-based and online studies in this thesis, was the inability to screen observers using the same equipment from the laboratory, such as was used in the experiments in this thesis to screen for sufficient visual acuity and stereovision. Instead, the experiments outlined the restrictions of having normal or correct-to-normal vision in both eyes. However, this relied on observer self-report. As this body of work was interested in the pattern of results across conditions, this is unlikely to have affected the conclusions, although future online studies should explore options for screening observers remotely. For instance, visual acuity testing methods have been developed for mobile device applications (Black, et al., 2013; Rodríguez-Vallejo, Ferrando, Montagud, Monsoriu, & Furlan, 2017), allowing for remote testing and automatic scoring that would address this issue going forward.

One restriction on data collection for observers was that the experiments had to be completed on a computer or laptop. Whilst this did not hamper data collection for these experiments, it may impact uptake of future online experiments given the popularity of smartphones for completing other types of studies, such as surveys (Cunningham, Neighbors, Bertholet, & Hendershot, 2013). One solution to this is presented by Marin-Campos, Dalmau, Compte and Linares (2021), who have developed a mobile device application on which psychophysical experiments can be conducted. This offers an interesting avenue to explore for future research in this area, to further increase the participant pool.

Finally, due to the nature of the online experiments within Qualtrics, there was not a feasible way to build in catch trials as would traditionally be used to ensure

observers are paying attention (Brebner & Flavel, 1978). Full written instructions were provided, as well as images depicting the task, but the online nature of collection did not allow the experimenter to verify participants' understanding of the task. Additionally, there is no 'correct' answer to the placement of the gauge figure, given that it is testing the perceived fit to the surface, which differs between observers. Therefore, it was not possible to provide perceptual feedback to observers to reassure them they were completing the task correctly. The only option was for the observer to contact the researcher, such as outlined in Chapter 5, although from the reconstructed meshes it is clear that not all observers were aware of their misunderstanding of the task, given the lack of coherent surface shape in some conditions. Additionally, where students were awarded course credit for participation, nothing within the system could verify if they had completed the task correctly. As such, several datasets were abandoned, see Chapter 5, as observers had evidently 'clicked through' the experiment with nothing to stop them or penalise them for doing so. To mitigate this, further observers were recruited through word of mouth to ensure sufficient data collection.

9.3.3 Conclusions

As shown, there are many pros and cons of using online data collection. The cost and benefits of conducting research in this way suggests that a pragmatic way forward could be to combine these data collection styles, ensuring the rigidity of control of laboratory-based studies, as well as the flexibility and convenience of online collection. One way to encompass both within a research plan could be to conduct pilot studies in a laboratory setting to ensure the experiments are set up

effectively, and complete data collection online, or run the two studies concurrently and compare results, as previous works have done.

For practical reasons, such as the use of specialised laboratory equipment in the laboratory-based studies and subsequent lack of access to these during the Covid19 pandemic, it was not feasible to complete the same experiment both in the laboratory and online. This would be a good avenue for future research to allow a direct comparison between methods and findings to explore the potential impact of using online data collection for studies of this nature.

9.4 Future directions

Many specific potential avenues for future work are presented within the chapters themselves, but a broad overview of the future direction of this work is considered here.

To begin with, the focus of this body of work was on complex scenes of naturalistic stimuli. 3D rendered objects created from real scanned fruits and vegetables, as well as other complex naturalistic stimuli in the form of computer-generated faces, were used to achieve this. However, stimuli of this kind can create artifacts, such as issues with digital elements conforming to the laws of physics within the rendering pipeline (Koenderink, 1999). As such, future studies could include physical elements as well as digital. One suggestion for this area would be the use of another rising advanced technology – 3D printing. This would offer the benefit of perfectly replicating the 3D models, and also provide a direction for multisensory experiments, such as including haptic feedback.

Overall, the use of linear mixed effects models within this work allowed for analysis mindful of the variance exhibited by observers in the various tasks. While this provided a robust method to account for the expected individual differences, these differences themselves were not considered within the scope of this work. However, future work could explore modelling these differences within the context of experiments such as those presented here.

On a similar note, some recent work has found that subjective evaluation of performance by observers contributes to cue weightings (Chen, McNamara, Kelly, & Wolbers, 2017). While many observers did find tasks such as the gauge figure task intuitive as highlighted earlier in the work, a small number of observers produced meshes that suggested they did not feel confident with the task, even with illustrative diagrams, and written, and where applicable verbal, instruction. Likewise, observers in Chapter 3, reported being unsure of making the judgements for the nonius lines task. That chapter explored a quantitative estimate of certainty of vergence from vergence noise, but an interesting follow up would be to include an element of metacognition and see if those observers who verbally felt more unsure about making alignment judgements also show greater uncertainty quantitively.

Overall, this body of work, while not the path envisaged at the start, presents a series of exciting roads for possible future research, ones which the author cannot wait to travel.

References

Abraham, W. T., & Russell, D. W. (2008). Statistical Power Analysis in Psychological. *Social and Personality Psychology Compass, 2*(1), 283-301. doi:10.1111/j.1751-9004.2007.00052.x

Alais, D., & Burr, D. (2019). Cue combination within a Bayesian framework. In A. Lee, M. Wallace, A. Coffin, A. Popper, & R. Fay (Eds.), *Multisensory processes. Springer handbook of auditory research* (Vol. 68, pp. 9-31). Springer. doi:https://doi.org/10.1007/978-3-030-10461-0_2

Allison, R. S. (2004). The camera convergence problem revisited. *Proceedings of the SPIE, 5291*, 167-178. doi:10.1117/12.526278

Anderson, B. L., & Marlow, P. J. (2023). Perceiving the shape and material properties of 3D surfaces. *Trends in Cognitive Sciences, 27*(1), 98-110. doi:https://doi.org/10.1016/j.tics.2022.10.005

Anderson, N. H. (1970). Functional measurement and psychophysical judgment. *Psychological Review, 77*(3), 153-170. doi:https://doi.org/10.1037/h0029064

Artec 3D. (n.d.). Luxembourg. Retrieved from https://www.artec3d.com/

Baird, J. C. (1970). A cognitive theory of psychophysics. I. *Scandinavian Journal of Psychology, 11*(1), 35-46. doi:10.1111/j.1467-9450.1970.tb00715.x

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2021). Power Contours: Optimising Sample Size and Precision in Experimental. *Psychological Methods, 26*(3), 295-314. doi:http://dx.doi.org/10.1037/met0000337

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:https://doi.org/10.18637/jss.v067.i01

Batres, C., Porcheron, A., Courrèges, S., & Russell, R. (2021). Professional versus self-applied makeup: Do makeup artists add value? *Perception, 50*(8), 709-719. doi:http://dx.doi.org/10.1177/03010066211029218

Batres, C., Russell, R., & Workowski, M. (2023). Makeup applied to facial features increases perceived skin evenness. *Vision Research, 202*, 1-4. doi:https://doi.org/10.1016/j.visres.2022.108144

Battaglia, P. W., Di Luca, M., Ernst, M. O., Schrater, P. R., Machulla, T., & Kersten, D. (2010). Within- and cross-modal distance information disambiguate visual size-change perception. *Plos Computational Biology, 6*(3), e1000697. doi:https://doi.org/10.1371/journal.pcbi.1000697

Bazargani, M., Anjos, A. D., Lobo, F., Mollahosseini, A., & Shahbazkia, H. (2012). Affine image registration transformation estimation using a real coded genetic algorithm with SBX. *arXiv*, 1-14. doi:https://doi.org/10.48550/arXiv.1204.2139

Benzeroual, K., Allison, R. S., & Wilcox, L. M. (2011). Distortions of space in stereoscopic 3D content. *SMPTE 2nd Annual International Conference on Stereoscopic 3D for Media and Entertainment*, (pp. 1-10). New York. doi:10.5594/M001420

Bernhard, M., Waldner, M., Plank, P., Soltészová, V., & Viola, I. (2016). The accuracy of gauge-figure tasks in monoscopic and stereo displays. *IEEE Computer Graphics and Applications, 36*(4), 56-66. doi:10.1109/MCG.2016.45

Biederman, I. (1981). Do background depth gradients facilitate object identification? *Perception, 10*(5), 573-578. doi:https://doi.org/10.1068/p100573

Black, J. M., Jacobs, R. J., Phillips, G., Chen, L., Tan, E., Tran, A., & Thompson, B. (2013). An assessment of the iPad as a testing platform for distance visual acuity in adults. *BMJ Open, 3*(6), e002730. doi:10.1136/bmjopen-2013-002730

Bradshaw, M. F., Elliott, K. M., Watt, S. J., Hibbard, P. B., Davies, I. R., & Simpson, P. J. (2004). Binocular cues and the control of prehension. *Spatial Vision, 17*(1-2), 95-110. doi:https://doi.org/10.1163/156856804322778288

Bradshaw, M. F., Parton, A. D., & Glennerster, A. (2000). The task-dependent use of binocular disparity and motion parallax information. *Vision Research, 40*, 3725-3734.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433-436.

Brebner, J., & Flavel, R. (1978). The effect of catch-trials on speed and accuracy among introverts and extraverts in a simple RT task. *British Journal of Psychology, 69*(1), 9-15. doi:https://doi.org/10.1111/j.2044-8295.1978.tb01627.x

Bredow, R., & Imageworks, S. P. (2002). Renderman on film. *SIGGRAPH 2002 Course 16 notes, RenderMan in Production*, pp. 103-128.

Brenner, E., & van Damme, W. J. (1998). Judging distance from ocular convergence. *Vision Research, 38*(4), 493-498.

Brooks, M. J., & Horn, B. K. (1985). Shape and source from shading. *Proceedings of the International Joint Conference on Artificial Intelligence*, (pp. 932-936).

Bülthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America, A, Optics, Image & Science, 5*(10), 1749-1758. doi:https://doi.org/10.1364/JOSAA.5.001749

Burge, J., Girshick, A. R., & Banks, M. S. (2010). Visual-haptic adaptation is determined by relative reliability. *Journal of Neuroscience, 30*(22), 7714-7721. doi:https://doi.org/10.1523/JNEUROSCI.6427-09.2010

Carter, M. (1998). Facials: the aesthetics of cosmetics and makeup. *Literature & Aesthetics*, 8.

Chen, C. C., & Tyler, C. W. (2015). Shading beats binocular disparity in depth from luminance gradients: Evidence against a Maximum Likelihood principle for cue combination. *PLoS ONE, 10*(8), e0132658. doi:https://doi.org/10.1371/journal.pone.0132658

Chen, X., McNamara, T. P., Kelly, J. W., & Wolbers, T. (2017). Cue combination in human spatial navigation. *Cognitive Psychology, 95*, 105-144. doi:https://doi.org/10.1016/j.cogpsych.2017.04.003

Cheng, K., Shettleworth, S. J., Huttenlocher, J., & Rieser, J. J. (2007). Bayesian integration of spatial information. *Psychological Bulletin, 133*(4), 625-637. doi:https://doi.org/10.1037/0033-2909.133.4.625

Choi, Y. K., & Baker, S. B. (2022). The effects of contour and highlighting makeup on the perception of facial form. *Aesthetic Surgery of the Facial Skeleton*, 485-499. doi:https://doi.org/10.1016/B978-0-323-48410-7.00053-8

Chopin, A., Levi, D. M., & Bavelier, D. (2017). Dressmakers show enhanced stereoscopic vision. *Scientific Reports, 7*. doi:https://doi.org/10.1038/s41598-017-03425-1

Chopin, A., Levi, D. M., Knill, D., & Bavelier, D. (2016). The absolute disparity anomaly and the mechanism of relative disparities. *Journal of Vision, 16*(8), 2. doi:https://doi.org/10.1167/16.8.2

Cooper, E. A., & Norcia, A. M. (2014). Perceived depth in natural images reflects encoding of low-level luminance statistics. *Journal of Neuroscience, 34*(35), 11761-11768. doi:10.1523/JNEUROSCI.1336-14.2014

Cormack, R., & Fox, R. (1985). The computation of retinal disparity. *Perception & Psychophysics*, 176-178.

Cornsweet, T. N. (1962). The staircase-method in psychophysics. *The American Journal of Psychology, 75*(3), 485-491. Retrieved from https://www.jstor.org/stable/1419876

Cunningham, J. A., Neighbors, C., Bertholet, N., & Hendershot, C. S. (2013). Use of mobile devices to answer online surveys: Implications for research. *BCM Research Notes, 6*, 1-4. doi:https://doi.org/10.1186/1756-0500-6-258

Cutting, J. E., & Vishton, P. (1995). Perceiving layout and knowing distances: The interaction, relative potency, and contextual use of different information about depth. *Perception of Space and Motion*, 69-177.

Danesi, M. (2018). Makeup: Why do we put it on? In M. Danesi, *Of Cigarettes, high heels, and other interesting things: An introduction to semiotics* (pp. 49-69). New York: Palgrave Macmillan. doi:https://doi.org/10.1057/978-1-349-95348-6

Dodgson, N. A. (2004). Variation and extrema of human interpupillary distance. *Stereoscopic Displays and Virtual Reality Systems XI.* SPIE 5291. doi:https://doi.org/10.1117/12.529999

Domini, F. (2023). The case against probabilistic inference: A new deterministic theory of 3D visual processing. *Philosophical Transaction Royal Society B, 378*(1869), 20210458. doi:https://doi.org/10.1098/rstb.2021.0458

Domini, F., & Braunstein, M. L. (1998). Recovery of 3-D structure from motion is neither euclidean nor affine. *Journal of Experimental Psychology: Human Perception and Performance, 24*(4), 1273-1295. doi:https://doi.org/10.1037/0096-1523.24.4.1273

Domini, F., & Caudek, C. (2003). 3-D structure perceived from dyamic information: A new theory. *Trends in Cognitive Sciences, 7*(10), 444-449. doi:https://doi.org/10.1016/j.tics.2003.08.007

Domini, F., Caudek, C., & Richman, S. (1998). Distortions of depth-order relations and parallelism in structure from motion. *Perception & Psychophysics, 60*(7), 1164-1174. doi:https://doi.org/10.3758/BF03206166

Egan, E. J., & Todd, J. T. (2015). The effects of smooth occlusions and directions of illumination on the visual perception of 3-D shape from shading. *Journal of Vision, 15*(2), 1-11. doi:https://doi.org/10.1167/15.2.24

Eldridge, L. (2015). *Face paint: The story of makeup.* Abrams.

Ernst, M. O. (2006). A Bayesian view on multimodal integration cue. In *Human body perception from the inside out* (p. 105).

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429-433. doi:https://doi.org/10.1038/415429a

Ernst, M. O., Banks, M. S., & Bülthoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience, 3*, 69-73. doi:https://doi.org/10.1038/71140

Feldman, J. (2009). Bayes and the simplicity principle in perception. *Psychological Review, 116*(4), 875-887. doi:10.1037/a0017144

Fernandez, J. M., & Farell, B. (2009). Is perceptual space inherently non-Euclidean? *Journal of Mathematical Psychology, 53*(2), 86-91. doi:10.1016/j.jmp.2008.12.006

Foley, J. M. (1980). Binocular distance perception. *Psychological Review, 87*(5), 411-434. doi:https://doi.org/10.1037/0033-295X.87.5.411

Foley, J. M., & Held, R. (1972). Visually directed pointing as a function of target distance, direction and available cues. *Perception & Psychophysics, 12*(3), 263-268. doi:https://doi.org/10.3758/BF03207201

Foster, R., Fantoni, C., Caudek, C., & Domini, F. (2011). Integration of disparity and velocity information for haptic and perceptual judgments of object depth. *Acta Psychologia, 136*(3), 300-310. doi:https://doi.org/10.1016/j.actpsy.2010.12.003

Fowler, S., Cutting, C., Kennedy, J., Leonard, S., Gabriel, F., & Jaeschke, W. (2021). Technology enhanced learning environments and the potential for enhancing spatial reasoning: A mixed methods study. *Mathematics Education Research Journal, 34*(4), 887-910. doi:https://doi.org/10.1007/s13394-021-00368-9

Gepshtein, S., Burge, J., Ernst, M. O., & Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision, 5*(11), 1013-1023. doi:https://doi.org/10.1167/5.11.7.

Gibson, J. J. (1950). The perception of visual surfaces. *The American Journal of Psychology, 63*(3), 367-384.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception: Classic edition.* Boston: Houghton Mifflin. doi:10.2307/429816

Girshick, A. R., & Banks, M. S. (2009). Probabilistic combination of slant information: Weighted averaging and robustness as optimal percepts. *Journal of Vision, 9*(9), 1-20. doi:https://doi.org/10.1167/9.9.8

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research, 36*(21), 3441-3456. doi:https://doi.org/10.1016/0042-6989(96)00090-9

Glennerster, A., Tcheang, L., Gilson, S. J., Fitzgibbon, A., & Parker, A. J. (2006). Humans ignore motion and stereo cues in favor of a fictional stable world. *Current Biology: CB, 16*(4), 428-432. doi:10.1016/j.cub.2006.01.019

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15*(1), 20-25. doi:https://doi.org/10.1016/0166-2236(92)90344-8

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet

questionnaires. *American Psychologist, 59*(2), 93-104. doi:https://doi.org/10.1037/0003-066X.59.2.93

Green, C. D. (1995). All that glitters: A review of psychological research on the aesthetics of the golden section. *Perception, 24*(8), 937-968. doi:10.1068/p240937

Gregory, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature, 199*, 678-680.

Gregory, R. L. (1973). *Eye and brain: The psychology of seeing* (2nd ed.). London: Weidenfeld and Nicolson.

Hallahan, M., & Rosenthal, R. (1996). Statistical power: concepts, procedures, and applications. *Behaviour Research and Therapy, 34*(5-6), 489-99. doi:10.1016/0005-7967(95)00082-8

Han, B., Chong, J., Sun, Z., Jiang, X., Xiao, Q., Zech, J., . . . Yang, Y. (2021). The rise of the cosmetic industry in ancient China: Insights from a 2700-year-old face cream. *Archaeometry, 63*(5), 1042-1058. doi:https://doi.org/10.1111/arcm.12659

Harris, J. M. (2004). Binocular vision: Moving closer to reality. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 362*(1825), 2721-2739. doi:https://doi.org/10.1098/rsta.2004.1464

Hartle, B., & Wilcox, L. M. (2022). Stereoscopic depth constancy for physical objects and their virtual counterparts. *Journal of Vision, 22*(4), 9. doi:https://doi.org/10.1167/jov.22.4.9

Hatfield, G., & Epstein, W. (1985). The status of the Minimum Principle in the theoretical analysis of visual perception. *Psychological Bulletin, 97*(2), 155-186. doi:https://doi.org/10.1037/0033-2909.97.2.155

Helbig, H. B., & Ernst, M. O. (2007). Knowledge about a common source can promote visual--haptic integration. *Perception, 36*(10), 1523-1533. doi:https://doi.org/10.1068/p5851

Hess, R. (2010). Blender foundations: The essential guide to learning Blender 2.6. Focal Press.

Hibbard, P. B. (2021). Estimating the contributions of pictorial, motion and binocular cues to the perception of distance. *Perception, 50*, 152.

Hibbard, P. B., & Bradshaw, M. F. (2003). Reaching for virtual objects: binocular disparity and the control of prehension. *Experimental Brain Research, 148*, 196-201. doi:https://doi.org/10.1007/s00221-002-1295-2

Hibbard, P. B., Bradshaw, M. F., Langley, K., & Rogers, B. J. (2002). The stereoscopic anisotrophy: Individual differences and underlying mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 469-476.

Hibbard, P. B., Goutcher, R., Hornsey, R. L., Hunter, D. W., & Scarfe, P. (2023). Luminance contrast provides metric depth information. *Royal Society Open Science, 10*(2), 220567. doi:https://doi.org/10.1098/rsos.220567

Hibbard, P. B., Hornsey, R. L., & Asher, J. M. (2022). Binocular information improves the reliability and consistency of pictorial relief. *Vision, 7*(1), 1. doi:https://doi.org/10.3390/vision7010001

Hibbard, P. B., van Dam, L. C., & Scarfe, P. (2020). The implications of interpupillary distance variability for virtual reality. *2020 International Conference on 3D Immersion (IC3D)* (pp. 1-7). Brussels, Belgium: IEEE. doi:10.1109/IC3D51119.2020.9376369.

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science (New York, N.Y.), 298*(5598), 1627-1630. doi:https://doi.org/10.1126/science.1075396

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*(12), 967-992. doi:https://doi.org/10.1167/4.12.1

Hirao, N., Koizumi, K., Ikeda, H., & Ohira, H. (2021). Reliability of online surveys in investigating perceptions and impressions of faces. *Frontiers in Psychology, 12*, 1-7. doi:10.3389/fpsyg.2021.733405

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision, 8*(3), 1-30. doi:https://doi.org/10.1167/8.3.33

Horn, B. K. (1970, November). Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. *Project MAC*. M.I.T.

Horn, B. K. (1990). Height and gradient from shading. *International Journal of Computer Vision, 5*(1), 37-75. doi:10.1007/BF00056771

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis.* New York: Oxford University Press.

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D., & Newman, T. B. (2013). Appendix 6C. In *Designing clinical research: an epidemiologic approach* (4th ed., p. 79). Lippincott Williams & Wilkins.

Hwang, A. D., & Peli, E. (2014). Instability of the perceived world while watching 3D stereoscopic imagery: A likely source of motion sickness symptoms. *i-Perception, 5*, 515-535. doi:dx.doi.org/10.1068/i0647

Inoue, T., Bounyong, S., Kato, Y. O., & Ozawa, J. (2013). Fixation distance estimation using vergence eye movement for automatic focusing glasses. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2013*, (pp. 4674-4677). doi:10.1109/EMBC.2013.6610590

Jain, N. K., & Chaudhri, S. K. (2009). History of cosmetics. *Asian Journal of Pharmaceutics, 3*(3), 164-167. doi:10.4103/0973-8398.56292

Jaschinski, W. (1997). Fixation disparity and accommodation as a function of viewing distance and prism load. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists), 17*(4), 324-339.

Jaschinski, W., Bröde, P., & Griefahn, B. (1999). Fixation disparity and nonius bias. *Vision Research, 39*(3), 669-677.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research, 31*(7-8), 1351-1360. doi:https://doi.org/10.1016/0042-6989(91)90056-B

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research, 34*(17), 2259-2275. doi:https://doi.org/10.1016/0042-6989(94)90106-6

Johnston, E. B., Cumming, B. G., & Parker, A. J. (1993). Integration of depth modules: Stereopsis and texture. *Vision Research, 33*(5-6), 813-826. doi:https://doi.org/10.1016/0042-6989(93)90200-G

Jones, A. L., Russell, R., & Ward, R. (2015). Cosmetics alter biologically-based factors of beauty; Evidence from facial contrast. *Evolutionary Psychology, 13*(1), 210-229. doi:147470491501300113

Julesz, B. (1971). *Foundations of cyclopean perception.* U. Chicago Press.

Kaiser, P. K. (2017). Calculation of visual angle. In *The joy of visual perception: A web book.* York: York University.

Karpicka, E., & Howarth, P. A. (2013). Heterophoria adaptation during the viewing of 3D stereoscopic stimuli. *Opthalmic and Physiological Optics, 33*(5), 604-610. doi:https://doi.org/10.1111/opo.12081

Kaufman, P. L., & Alm, A. (Eds.). (2003). *Adler's physiology of the eye* (10th ed.). St Louis, MO: Mosby Inc.

Keefe, B. D., Hibbard, P. B., & Watt, S. J. (2011). Depth-cue integration in grasp programming: No evidence for a binocular specialism. *Neuropsychologia, 49*(5), 1246-1257. doi:https://doi.org/10.1016/j.neuropsychologia.2011.02.047

Kęsik, J., Żyła, K., Montusiewicz, J., Neamtu, C., & Juszczyk, M. (2023). A methodical approach to 3D scanning of heritage objects being under

continuous display. *Applied Sciences, 13*, 441: 1-20. doi:https://doi.org/10.3390/app13010441

Kim, J., Kane, D., & Banks, M. S. (2014). The rate of change of vergence–accommodation conflict affects visual discomfort. *Vision Research, 105*, 159-165. doi:https://doi.org/10.1016/j.visres.2014.10.021

Kingdom, F. A. (2003). Color brings relief to human vision. *Nature Neuroscience, 6*(6), 641-644. doi:https://doi.org/10.1038/nn1060

King-Smith, P. E., & Rose, D. (1997). Principles of an adaptive method for measuring the slope of the psychometric function. *Vision Research, 37*(12), 1595-1604. doi:https://doi.org/10.1016/S0042-6989(96)00310-0

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 36*, 1-16.

Knill, D. C. (1998). Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research, 38*(17), 2635-2656. doi:https://doi.org/10.1016/S0042-6989(97)00415-X

Knill, D. C., & Richards, W. (Eds.). (2008). *Perception as Bayesian inference.* Cambridge: Cambridge University Press.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research, 43*(24), 2539-2558. doi:https://doi.org/10.1016/S0042-6989(03)00458-9

Knill, D. C., Kersten, D., & Yuille, A. (2008). Introduction: A Bayesian formulation of visual perception. In D. C. Knill, & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 1-21). Cambridge: Cambridge University Press.

Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception, 13*(3), 321-330. doi:https://doi.org/10.1068/p130321

Koenderink, J. J. (1998). Pictorial relief. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 356*(1740), 1071-1086.

Koenderink, J. J. (1999). Virtual psychophysics. *Perception, 28*(6), 669-674. doi:10.1068/p2806ed

Koenderink, J. J. (2012). *Pictorial space.* Utrecht: De Clootcrans Press.

Koenderink, J. J., & van Doorn, A. J. (1982). The shape of smooth objects and the way contours end. *Perception, 11*(2), 129-137. doi:https://doi.org/10.1068/p110129

Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. (1992). Surface perception in pictures. *Perception & Psychophysics, 52*(5), 487-496.

Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. (1995). Depth relief. *Perception, 24*, 115-126.

Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. (1996). Pictorial surface attitude and local depth comparisons. *Perception & Psychophysics*, 163-173.

Koenderink, J. J., van Doorn, A. J., & Lappin, J. S. (2000). Direct measurement of the curvature of visual space. *Perception, 29*(1), 69-79. doi:https://doi.org/10.1068/p2921

Koenderink, J. J., van Doorn, A. J., & Wagemans, J. (2011). Depth. *i-Perception*, 541-564.

Koenderink, J. J., van Doorn, A. J., & Wagemans, J. (2015). Part and whole in pictorial relief. *i-Perception, 6*(6), 1-21. doi:10.1177/2041669515615713

Koenderink, J. J., van Doorn, A. J., Kappers, A. M., & Lappin, J. S. (2002). Large-scale visual frontoparallels under full-cue conditions. *Perception, 31*(12), 1467-1475. doi:10.1068/p3295

Koenderink, J. J., van Doorn, A. J., Kappers, A. M., & Todd, J. T. (2001). Ambiguity and the 'mental eye' in pictorial relief. *Perception, 30*(4), 431-448. doi:https://doi.org/10.1068/p3030

Koenderink, J. J., van Doorn, A. J., Kappers, A. M., & Todd, J. T. (2002). Pappus in optical space. *Perception & Psychophysics, 64*(3), 380-391. doi:https://doi.org/10.3758/bf03194711

Komoda, M. K., & Ono, H. (1974). Oculomotor adjustments and size-distance perception. *Perception & Psychophysics, 15*, 353-360. doi:10.3758/BF03213958

Kunnapas, T. (1968). Distance perception as a function of available visual cues. *Journal of Experimental Psychology*, 523-529. doi:https://doi.org/10.1037/h0026050

Lambooij, M. T., Ijsselsteijn, W. A., & Heynderickx, I. (2007). Visual discomfort in stereoscopic displays: A review. *Proceedings SPIE 6490, Stereoscopic Displays and Virtual Reality Systems XIV*, 649001. doi:https://doi.org/10.1117/12.705527

Landis, H. (2002). Production-ready global illumination. *Siggraph Course Notes, 16*, 87-101.

Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America. A, Optics, Image Science and Vision, 18*(9), 2307-2320. doi:https://doi.org/10.1364/josaa.18.002307

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defence of weak fusion. *Vision Research, 35*(3), 389-412. doi:https://doi.org/10.1016/0042-6989(94)00176-M

Likova, L. T., & Tyler, C. W. (2003). Peak localization of sparsely sampled luminance patterns is based on interpolated 3D surface representation. *Vision Research, 43*(25), 2649-2657. doi:10.1016/s0042-6989(02)00575-8

Linton, P. (2020). Does vision extract absolute distance from vergence? *Attention, Perception, & Psychophysics, 82*, 3176-3195. doi:10.3758/s13414-020-02006-1

Linton, P. (2022). Minimal theory of 3D vision: new approach to visual scale and visual shape. *Philosophical Transactions Royal Society B, 378*(1869), 20210455. doi:https://doi.org/10.1098/rstb.2021.0455

Linton, P., Morgan, M. J., Read, J. C., Vishwanath, D., Creem-Regehr, S. H., & Domini, F. (2022). New approaches to 3D vision. *Philosophical Transactions Royal Society B, 378*, 20210443. doi:https://doi.org/10.1098/rstb.2021.0443

Lovell, P. G., Bloj, M., & Harris, J. M. (2012). Optimal integration of shading and binocular disparity for depth perception. *Journal of Vision, 12*(1), 1-18. doi:https://doi.org/10.1167/12.1.1

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition, 81*(1), B1-9. doi:10.1016/s0010-0277(01)00116-0

Marin-Campos, R., Dalmau, J., Compte, A., & Linares, D. (2021). StimuliApp: Psychophysical tests on mobile devices. *Behavior Research Methods, 53*(3), 1301-1307. doi:10.3758/s13428-020-01491-4

Marr, D. (1977). Analysis of occluding contour. *Proceedings of the Royal Society B, 197*(1129), 441-475. doi:https://doi.org/10.1098/rspb.1977.0080

Marr, D. (1982). *Vision: A computational approach.* San Francisco, CA: Freeman and Co..

Mather, G., & Smith, D. R. (2004). Combining depth cues: Effects upon accuracy and speed of performance in a depth-ordering task. *Vision Research*, 557-562.

McCabe, M., Malefyt, T. W., & Fabri, A. (2017). Women, makeup, and authenticity: Negotiating embodiment and discourses of beauty. *Journal of Consumer Culture, 20*(4), 1-22. doi:https://doi.org/10.1177/1469540517736558

Melmoth, D. R., & Grant, S. (2006). Advantages of binocular vision for the control of reaching and grasping. *Experimental Brain Research, 171*(3), 371-388. doi:https://doi.org/10.1007/s00221-005-0273-x

Méndez-Feliu, À., & Sbert, M. (2009). From obscurances to ambient occlusion: A survey. *The Visual Computer, 25*(2), 181-196. doi:10.1007/s00371-008-0213-4

Mon-Williams, M., & Tresilian, J. R. (1999). Some recent studies on the extraretinal contribution to distance perception. *Perception, 28*(2), 167-181. doi:https://doi.org/10.1068/p2737

Mon-Williams, M., Tresilian, J. R., & Roberts, A. B. (2000). Vergence provides veridical depth perception from horizontal retinal image disparities. *Experimental Brain Research, 133*, 407-413.

Morikawa, K., Matsushita, S., Tomita, A., & Yamanami, H. (2015). A real-life illusion of assimilation in the human face: Eye size illusion caused by eyebrows and eye shadow. *Frontiers in Human Neuroscience, 9:139*, 1-9. doi:https://doi.org/10.3389/fnhum.2015.00139

Morrell, C. H., Pearson, J. D., & Brant, L. J. (1997). Linear transformations of linear mixed-effects models. *The American Statistician, 51*(4), 338-343.

Muller, C. M., Brenner, E., & Smeets, J. B. (2009). Testing a counter-intuitive prediction of optimal cue combination. *Vision Research, 49*(1), 134-139. doi:https://doi.org/10.1016/j.visres.2008.10.006

Nadenau, M. J., Reichel, J., & Kunt, M. (2002). Performance comparison of masking models based on a new psychvisual test method with natural scenery stimuli. *Signal Processing: Image Communication, 17*(10), 807-823.

Nefs, H. T. (2008). Three-dimensional object shape from shading and contour disparities. *Journal of Vision, 8*(11), 1-16. doi:https://doi.org/10.1167/8.11.11

Nefs, H. T., Koenderink, J., & Kappers, A. M. (2005). The influence of illumination direction on the pictorial reliefs of lambertain surfaces. *Perception, 34*(3), 275-287. doi:10.1068/p5179

Nefs, H. T., O'Hare, L., & Harris, J. M. (2010). Two independent mechanisms for motion-in-depth perception: evidence from individual differences. *Frontiers in Psychology, 1*. doi:https://doi.org/10.3389/fpsyg.2010.00155

Norman, J. F., & Todd, J. T. (1992). The visual perception of 3-dimensional form. In G. A. Carpenter, & S. Grossberg (Eds.), *Neural networks for vision and image processing* (pp. 93-110). Cambridge, MA: MIT Press.

Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics, 53*(3), 279-291. doi:https://doi.org/10.3758/BF03205183

O'Kane, L. M., & Hibbard, P. B. (2007). Vertical disparity affects shape and size judgements across surfaces separated in depth. *Perception, 36*, 696-702. doi:DOI:10.1068/p5406

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *The Behavioural and Brain Sciences, 24*(5), 939-1031. doi:https://doi.org/10.1017/s0140525x01000115

Oruç, İ., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43*(23), 2451-2468. doi:https://doi.org/10.1016/S0042-6989(03)00435-8

O'Shea, R. P. (1991). Thumb's rule tested: Visual angle of thumb's width. *Perception, 20*(3), 415-418. doi:https://doi.org/10.1068/p200415

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.

Porac, C., & Coren, S. (1976). The dominant eye. *Psychological Bulletin, 83*(5), 880-897. doi:https://doi.org/10.1037/0033-2909.83.5.880

Potetz, B., & Lee, T. S. (2003). Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes. *Journal of the Optical Society of America, 20*(7), 1292-1303. doi:10.1364/JOSAA.20.001292

Prins, N. (2013). The psi-marginal adaptive method: How to give nuisance parameters the attention they deserve (no more, no less). *Journal of Vision, 13*(7), 1-17. doi:https://doi.org/10.1167/13.7.3

Prins, N., & Kingdom, F. A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes Toolbox. *Frontiers in Psychology, 9:1250*, 1-14. doi:https://doi.org/10.3389/fpsyg.2018.01250

Purves, D., & Lotto, R. B. (2011). *Why we see what we do redux: A wholly empirical theory of vision.* Sunderland: Sinaeur Associates.

Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behaviour and Brain Sciences, 41*(e223), 1-66. doi:https://doi.org/10.1017/S0140525X18000936

Rajamanickam, M. (2002). *Modern Psychophysical and Scaling Methods and Experimentation.* New York: John Wiley & Sons, Inc.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 331*(6152), 163-166. doi:https://psycnet.apa.org/doi/10.1038/331163a0

Raman, R., & Sarkar, S. (2016). Predictive coding: A possible explanation of filling-in at the blind spot. *PLoS One, 11*(3), e0151194. doi:10.1371/journal.pone.0151194

Reallusion. (n.d.). California, United States. Retrieved from https://www.reallusion.com/facefilter/

Reimers, S., & Stewart, N. (2015). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript web experiments. *Behaviour Research Methods, 47*, 309-327. doi:https://doi.org/10.3758/s13428-014-0471-1

Reips, U. D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the internet* (pp. 89-117). Academic Press. doi:https://doi.org/10.1016/B978-012099980-4/50005-8

Revina, Y., & Maus, G. W. (2020). Stronger perceptual filling-in of spatiotemporal information in the blind spot compared with artificial gaps. *Journal of Vision, 20*(4), 1-17. doi:https://doi.org/10.1167/jov.20.4.20

Rideaux, R., & Welchman, A. E. (2018). Proscription supports robust perceptual integration by suppression in human visual cortex. *Nature Communications, 9*(1), 1502. doi:https://doi.org/10.1038/s41467-018-03400-y

Rodríguez-Vallejo, M., Ferrando, V., Montagud, D., Monsoriu, J. A., & Furlan, W. D. (2017). Stereopsis assessment at multiple distances with an iPad application. *Displays, 50*, 35-40. doi:http://dx.doi.org/10.1016/j.displa.2017.09.001

Rosas, P., & Wichmann, F. A. (2011). Cue combination: Beyond optimality. In J. Trommershäuser, K. Kording, & M. S. Landy (Eds.), *Sensory cue integration*

(pp. 144-152). doi:https://doi.org/10.1093/acprof:oso/9780195387247.003.0008

Rosas, P., Wichmann, F. A., & Wagemans, J. (2007). Texture and object motion in slant discrimination: Failure of reliability-based weighting of cues may be evidence for strong fusion. *Journal of Vision, 7*(6), 1-21. doi:https://doi.org/10.1167/7.6.3

Rushton, S. K., Harris, J. M., Lloyd, M. R., & Wann, J. P. (1998). Guidance of locomotion on foot uses perceived target location rather than optic flow. *Current Biology, 8*(21), 1191-1194. doi:https://doi.org/10.1016/S0960-9822(07)00492-7

Russell, P. A. (2000). Testing the aesthetic significance of the golden-section rectangle. *Perception, 29*(12), 1413-1422. doi:https://doi.org/10.1068/p3037

Russell, R. (2003). Sex, beauty, and the relative luminance of facial features. *Perception, 32*(9), 1093-1107. doi:DOI:10.1068/p5101

Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nature Neuroscience, 8*(12), 1647-1650. doi:10.1038/nn1606

Saunders, J. A., & Backus, B. T. (2006). Perception of surface slant from oriented textures. *Journal of Vision, 6*(9), 3. doi:https://doi.org/10.1167/6.9.3

Saunders, J. A., & Chen, Z. (2015). Perceptual biases and cue weighting in perception of 3D slant from texture and stereo information. *Journal of Vision, 15*(2), 1-24. doi:https://doi.org/10.1167/15.2.14

Scaccia, M., & Langer, M. S. (2018). Signs of depth-luminance covariance in 3-D cluttered scenes. *Journal of Vision, 18*(5), 1-13. doi:https://doi.org/10.1167/18.3.5

Scarfe, P. (2022). Experimentally disambiguating models of sensory cue. *Journal of Vision, 22*(1), 1-23. doi:https://doi.org/10.1167/jov.22.1.5

Scarfe, P., & Hibbard, P. B. (2006). Disparity-defined objects moving in depth do not elicit three-dimensional shape constancy. *Vision Research, 46*(10), 1599-1610.

Scarfe, P., & Hibbard, P. B. (2011). Statistically optimal integration of biased sensory estimates. *Journal of Vision, 11*(7), 1-17. doi:https://doi.org/10.1167/11.7.12

Scarfe, P., & Hibbard, P. B. (2017). A Bayesian model of distance perception from ocular convergence. *Journal of Vision, 17*(10), 159-159. doi:10.1167/17.10.159

Scherbaum, K., Ritschel, T., Hullin, M., Thormählen, T., Blanz, V., & Seidel, H.-P. (2011). Computer-suggested facial makeup. *Computer Graphics Forum, 30*(2), 485-492. doi:https://doi.org/10.1111/j.1467-8659.2011.01874.x

Schofield, A. J., Rock, P. B., & Georgeson, M. A. (2011). Sun and sky: Does human vision assume a mixture of point and diffuse illumination when interpreting shape-from-shading? *Vision Research, 51*(21-22), 2317-2330. doi:10.1016/j.visres.2011.09.004

Sedgwick, H. A. (1986). Space perception. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (pp. 21.1-21.57). Wiley.

Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: reaction time effects in cognitive experiments. *Behavior Research Methods, 49*, 1241-1260. doi:10.3758/s13428-016-0783-4

Servos, P., Goodale, M. A., & Jakobson, L. S. (1992). The role of binocular vision in prehension: a kinematic analysis. *Vision Research, 32*(8), 1513-1521. doi:https://doi.org/10.1016/0042-6989(92)90207-Y

Singh, P., Vijayan, R., & Mosahebi, A. (2019). The golden ratio and aesthetic surgery. *Aesthetic Surgery Journal, 39*(1), 4-5. doi:10.1093/asj/sjy240

Singular Inversions Inc. (n.d.). *FaceGen*. Toronto, Canada. Retrieved from https://facegen.com/

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review, 25*, 2083-2101. doi:https://doi.org/10.3758/s13423-018-1451-8

Squire, M., & Platt, V. J. (2017). *The frame in classical art.* Cambridge: Cambridge University Press.

Stereo Optical. (2020, April 20). *Stereotests & color tests*. Retrieved from Stereo Optical: https://www.stereooptical.com/products/stereotests-color-tests/butterfly/

Stevens, K. A. (1983). Surface tilt (the direction of slant): A neglected psychophysical variable. *Perception & Psychophysics, 33*(3), 241-250.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects.* New Jersey: John Wiley & Sons.

Stidwell, D., & Fletcher, R. (2011). *Normal binocular vision: Theory, investigation and practical aspects.* Chicester: Wiley-Blackwell.

Svarverud, E., Gilson, S. J., & Glennerster, A. (2010). Cue combination for 3D location judgements. *Journal of Vision, 10*(1), 1-13. doi:https://doi.org/10.1167/10.1.5

Takeda, T., Hashimoto, K., Hiruma, N., & Fukui, Y. (1999). Characteristics of accommodation toward apparent depth. *Vision Research, 39*(12), 2087-2097. doi:https://doi.org/10.1016/S0042-6989(98)00258-2

Taya, S. (2023). Do observers use their own interpupillary distance in disparity scaling? *Optical Review*. doi:https://doi.org/10.1007/s10043-022-00780-x

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 663-678. doi:https://doi.org/10.1037/0096-1523.21.3.663

Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics, 48*(5), 419-430. doi:https://doi.org/10.3758/BF03211585

Todd, J. T., & Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance, 9*(4), 583-595. doi:https://doi.org/10.1037/0096-1523.9.4.583

Todd, J. T., & Norman, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception & Psychophysics, 50*(6), 509-523. doi:https://doi.org/10.3758/BF03207535

Todd, J. T., & Reichel, F. D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychological Review, 96*(4), 643-657. doi:https://doi.org/10.1037/0033-295X.96.4.643

Todd, J. T., Egan, E. J., & Kallie, C. S. (2015). The darker-is-deeper heuristic for the perception of 3D shape from shading: Is it perceptually or ecologically valid? *Journal of Vision, 15*(15), 2. doi:https://doi.org/10.1167/15.15.2

Todd, J. T., Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. (1996). Effects of changing viewing conditions on the perceived structure of smoothly curved surfaces. *Journal of Experimental Psychology: Human Perception and Performance, 22*(3), 695-706. doi:https://doi.org/10.1037/0096-1523.22.3.695

Tong, W. S., Tang, C. K., Brown, M. S., & Xu, Y. Q. (2007). Example-based cosmetic transfer. *15th Pacific Conference on Computer Graphics and Applications* (pp. 211-218). IEEE. doi:DOI 10.1109/PG.2007.31

Tresilian, J. R., & Mon-Williams, M. (2000). Getting the measure of weights for cue combination in nearness perception. *Experimental Brain Research*, 28-35.

Tresilian, J. R., Mon-Williams, M., & Kelly, B. M. (1999). Increasing confidence in vergence as a cue to distance. *Proceedings: Biological Sciences, 266*(1414), 39-44. doi:http://www.jstor.org/stable/51340

Trommershäuser, J., Körding, K., & Landy, M. S. (2011). *Sensory cue integration.* Oxford: Oxford University Press.

Tyler, C. W. (1998). Diffuse illumination as a default assumption for shape-from-shading in the absence of shadows. *Journal of Imaging Science and Technology, 42*(4), 319-325.

Tyler, C. W. (2019). Depth cue combination: A quantitative critique. *Perception, 48*(9), 765-768. doi:https://doi.org/10.1177/0301006619865899

Tyler, C. W. (2020). An accelerated cue combination principle accounts for multi-cue depth perception. *Journal of Perceptual Imaging, 3*(1), 10501-1 - 10501-9. doi:https://doi.org/10.2352/J.Percept.Imaging.2020.3.1.010501

van Ee, R., van Dam, L. C., & Erkelens, C. J. (2002). Bi-stability in perceived slant when binocular disparity and monocular perspective specify different slants. *Journal of Vision, 2*(9), 597-607. doi:https://doi.org/10.1167/2.9.2

Viguier, A., Clément, G., & Trotter, Y. (2001). Distance peception within near visual space. *Perception, 30*(1), 115-124. doi:https://doi.org/10.1068/p3119

Vishwanath, D. (2022). From pictures to reality: modelling the phenomenology and psychophysics of 3D perception. *Philosophical Transactions of the Royal Society B, 378*(1869), 1-17. doi:https://doi.org/10.1098/rstb.2021.0454

VPixx Technologies. (2020, April 20). *VIEWPixx*. Retrieved from VPixx Technologies: Vision Science Solutions: https://vpixx.com/products/viewpixx/

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin, 138*(6), 1218-1252. doi:https://doi.org/10.1037/a0029334

Wagner, M. (1985). The metric of visual space. *Perception & Psychophysics, 38*(6), 483-495. doi:https://doi.org/10.3758/BF03207058

Wahl, B. (1991). Chapter 4: Calculating fractal dimensions. In B. Wahl, *Fractal Explorer.* Retrieved from www.wahl.org/fe/HTML_version/link/FE4W/c4.htm

Wallach, H., & Zuckerman, C. (1963). The constancy of stereoscopic depth. *The American Journal of Psychology, 76*(3), 404-412. doi:https://doi.org/10.2307/1419781

Wang, Y. J., & Lin, Y. H. (2021). Liquid crystal technology for vergence-accommodation conflicts in augmented reality and virtual reality systems: A review. *Liquid Crystal Reviews, 9*(1), 35-64. doi:https://doi.org/10.1080/21680396.2021.1948927

Ware, C. (2021). *Visual thinking for information design* (2nd ed.). Burlington, Massachusetts: Morgan Kaufmann, Elsevier Inc. doi:https://doi.org/10.1016/C2016-0-01395-5

Warren, P. A., & Mamassian, P. (2010). Recovery of surface pose from texture orientation statistics under perspective projection. *Biological Cybernetics, 103*(3), 199-212. doi:https://doi.org/10.1007/s00422-010-0389-3

Warren, W. H., & Hannon, D. J. (1988). Direction of self-motion is perceived from optical flow. *Nature, 336*(6195), 162-163. doi:https://doi.org/10.1038/336162a0

Watt, S. J., & Bradshaw, M. F. (2000). Binocular cues are important in controlling the grasp but not the reach in natural prehension movements. *Neuropsychologia, 38*(11), 1471-1481. doi:10.1016/s0028-3932(00)00065-8

Watt, S., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision, 5*(10), 834-862. doi:https://doi.org/10.1167/5.10.7

Wijntjes, M. W., & van Zuijlen, M. J. (2004). Sketch-and-test: picture-centered research with p5.js assisted crowdsourcing. *arXiv*, 08198. doi:https://doi.org/10.48550/arXiv.2004.08198

Wilcox, L. M. (2016). Depth magnitude from stereopsis: Assessment techniques and the role of experience. *Vision Research, 125*, 64-75. doi:https://doi.org/10.1016/j.visres.2016.05.006

Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence, 17*(1-3), 17-45.

Wixted, J. T. (Ed.). (2018). *Stevens' handbook of experimental psychology and cognitive neuroscience* (4th ed., Vol. 5: Methodology). New York: Wiley.

Wu, J. H., Hu, S. M., Tai, C. L., & Sun, J. G. (2001). An effective feature-preserving mesh simplification scheme based on face constriction. *Proceedings Ninth Pacific Conference on Computer Graphics and Applications* (pp. 12-21). Tokyo, Japan: IEEE. doi:10.1109/PCCGA.2001.962853

Yamanoue, H., Okui, M., & Yuyama, I. (2000). A study on the relationship between shooting conditions and cardboard effect of stereoscopic images. *IEEE Transactions on Circuits and Systems for Video Technology, 10*(3), 411-416. doi:10.1109/76.836285

Yuille, A. L., & Bülthoff, H. H. (2008). Bayesian decision theory and psychophysics. In D. C. Knill, & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123-161). Cambridge: Cambridge University Press.

Zhukov, S., Iones, A., & Kronin, G. (1998). An ambient light illumination model. In *Eurographics workshop on rendering techniques* (pp. 45-55). Vienna, Austria: Springer. doi:10.1007/978-3-7091-6453-2_5

10 Appendix

10.1 Chapter 3

The full linear mixed effects model with random slopes and intercepts was included in the main analysis, as these had the lowest AIC values, indicating the best fit of the model. Here, the model without random slopes or intercepts is presented for completeness. Estimates were not found to change between models, showing robustness, with p values and confidence intervals only changing to a small extent.

Table 10-1: Goodness of fit comparison for PSE and JND of both tasks.

Variable	Model	AIC	Estimate	*p* Value	Lower CI	Upper CI
Depth PSE (pd)	$pd \sim 1 + d + (1 + d \mid o)$	603.33	0.0647	**<.001*** **	0.0484	0.0809
	$pd \sim 1 + d + (1 \mid o)$	626.92	0.0647	**<.001*** **	0.0513	0.0780
Depth JND (sd)	$sd \sim 1 + d + (1 + d \mid o)$	-207.43	0.00115	**.013***	0.00024	0.00205
	$sd \sim 1 + d + (1 \mid o)$	-207.37	0.00115	**.007****	0.00031	0.00198
Vergence PSE (pv)	$pv \sim 1 + d + (1 + d \mid o)$	1891.5	-2.252	**.001****	-3.514	0.990
	$pv \sim 1 + d + (1 \mid o)$	1895.7	-2.252	**<.001*** **	-3.422	-1.082
Vergence JND (sv)	$sv \sim 1 + d + (1 + d \mid o)$	2021.9	-5.35	**<.001*** **	-8.13	-2.57
	$sv \sim 1 + d + (1 \mid o)$	2044.8	-5.35	**<.001*** **	-7.60	-3.09

Here the model comparison is shown for the fixation disparity LME against distance, which shows the best model as the full random intercepts and slopes model with the lowest AIC value, with similar results.

Table 10-2: Goodness of fit comparison for fixation disparity against distance.

Variable	Model	AIC	Estimate	*p* Value	Lower CI	Upper CI
Fixation disparity (fd)	$fd \sim 1 + d + (1 + d \mid o)$	321.38	0.993	**<.001*** **	0.985	1.001
	$fd \sim 1 + d + (1 \mid o)$	386.79	0.993	**<.001*** **	0.987	0.998

10.2 Chapter 4

4.1.1 Model comparisons

Here, the AIC values for the ground truth LME are shown. The random slopes and intercepts model was selected as having the lowest AIC value.

Table 10-3: LME results ground truth comparison.

Model	AIC
$d \sim 1 + g + (1 + g \mid o)$	**68552**
$d \sim 1 + g + (1 \mid o)$	68767

Here, we present the model comparisons for the LME to predict the error in distance judgements. The model has two predictor variables as the main fixed effects, as well as fixed interactions between these. The best fit to the data from the lowest AIC value was the random intercepts and random slopes for condition (c) and monocular or binocular viewing (e) model for all groups except the object far group.

However, the full random intercepts and slopes model was chosen for consistency as the data is from a repeated measures design, and selecting this model over the random slopes for monocular and binocular viewing alone did not affect the estimates or significance pattern.

Table 10-4: LME results relative distance error for the four groupings.

Group	Model	AIC
Object close	$err \sim 1 + c * e + (1 + c + e \mid o)$	**-46.39**
	$err \sim 1 + c * e + (1 + c \mid o)$	-35.87
	$err \sim 1 + c * e + (1 + e \mid o)$	-37.16
	$err \sim 1 + c * e + (1 \mid o)$	-34.52
Object far	$err \sim 1 + c * e + (1 + c + e \mid o)$	0.58
	$err \sim 1 + c * e + (1 + c \mid o)$	1.52
	$err \sim 1 + c * e + (1 + e \mid o)$	**-4.00**
	$err \sim 1 + c * e + (1 \mid o)$	-1.24
Scene close	$err \sim 1 + c * e + (1 + c + e \mid o)$	**55.31**
	$err \sim 1 + c * e + (1 + c \mid o)$	70.30
	$err \sim 1 + c * e + (1 + e \mid o)$	64.17
	$err \sim 1 + c * e + (1 \mid o)$	66.40
Scene far	$err \sim 1 + c * e + (1 + c + e \mid o)$	**-52.27**
	$err \sim 1 + c * e + (1 + c \mid o)$	-39.08
	$err \sim 1 + c * e + (1 + e \mid o)$	-48.31
	$err \sim 1 + c * e + (1 \mid o)$	-43.03

10.3 Chapter 5

10.3.1 Experiment 1

Here the AIC values for each model considered are presented for completeness. The model chosen as having the lowest AIC value was the maximal random slopes for gauge size, texture and the interaction and random intercepts model. Estimates and patterns of significance values did not differ broadly.

Table 10-5: Goodness of fit comparisons for depth range Experiment 1.

Model	AIC
$dr \sim 1 + g * t + (1 + g * t \mid o)$	**972.78**
$dr \sim 1 + g * t + (1 + g + t \mid o)$	992.44
$dr \sim 1 + g * t + (1 + g \mid o)$	993.52
$dr \sim 1 + g * t + (1 + t \mid o)$	988.87
$dr \sim 1 + g * t + (1 \mid o)$	990.11

The model chosen for the surface roughness comparison was the random intercepts only model, given its low AIC value. However, as before, patterns of results did not differ largely between models.

Table 10-6 Goodness of fit comparisons for surface roughness Experiment 1.

Model	AIC
$r \sim 1 + g * t + (1 + g * t \mid o)$	753.14
$r \sim 1 + g * t + (1 + g + t \mid o)$	755.85
$r \sim 1 + g * t + (1 + g \mid o)$	753.33
$r \sim 1 + g * t + (1 + t \mid o)$	750.02
$r \sim 1 + g * t + (1 \mid o)$	**748.92**

10.3.2 Experiment 2

Here the model comparison for the roughness measure. Estimates and significance values did not differ considerably between models, and therefore the random slope and intercepts model was selected as having the lowest AIC value.

Table 10-7: Goodness of fit comparisons for surface roughness Experiment 2.

Model	AIC
$r \sim 1 + g + (1 + g\|o)$	**-3.85**
$r \sim 1 + g + (1\|o)$	54.61

This table lists the model comparisons for the depth range for Experiment 2. The maximal model was selected as having the lowest AIC value.

Table 10-8: Goodness of fit comparisons for depth range Experiment 2.

Model	AIC
$r \sim 1 + g + (1 + g\|o)$	**50.24**
$r \sim 1 + g + (1\|o)$	128.96

10.4 Chapter 6

Here the model comparisons are presented for completeness.

10.4.1 Experiment 1

The model comparisons for the LME to predict set gradients show the best fitting model is the random intercepts and slopes of physical gradient model by the lowest AIC value.

Table 10-9: Goodness of fit comparison between LME models for set depth.

Model	AIC
$s \sim 1 + p * g * d + (1 + p * g * d \mid o)$	33788
$s \sim 1 + p * g * d + (1 + p * g \mid o)$	33785
$s \sim 1 + p * g * d + (1 + p * d \mid o)$	33788
$s \sim 1 + p * g * d + (1 + g * d \mid o)$	33873
$s \sim 1 + p * g * d + (1 + p + g \mid o)$	33784
$s \sim 1 + p * g * d + (1 + p + d \mid o)$	33784
$s \sim 1 + p * g * d + (1 + g + d \mid o)$	33866
$s \sim 1 + p * g * d + (1 + p \mid o)$	**33782**
$s \sim 1 + p * g * d + (1 + g \mid o)$	33784
$s \sim 1 + p * g * d + (1 + d \mid o)$	33864
$s \sim 1 + p * g * d + (1 \mid o)$	33863

Here the model comparisons for affine transformations for Experiment 1 are shown, giving the AIC value for each model. As depth is the parameter of interest in this work, the best fitting model for this parameter was chosen for the full LME analysis, including the squashed and stretched gain analysis. The lowest AIC values for each parameter are shown in bold. The best model selected by AIC value was random intercepts and random slopes for distance.

Table 10-10: Goodness of fit comparison for affine transformations Experiment 1.

Model	bz	b0	bx	by
$b{\sim}1 + d * g + (1 + d * g \mid o)$	36.54	748.16	-466.7	-530.59
$b{\sim}1 + d * g + (1 + d + g \mid o)$	32.98	746.86	-470.44	-531.73
$b{\sim}1 + d * g + (1 + d \mid o)$	**26.99**	**742.81**	**-471.57**	**-536.15**
$b{\sim}1 + d * g + (1 + g \mid o)$	33.61	764.22	-454.24	-522.56
$b{\sim}1 + d * g + (1 \mid o)$	33.10	761.63	-450	-526.25

10.4.2 Experiment 2

Here, the model comparisons are shown for the reported depth. The random intercepts and slopes for distance and gain was selected as having the lowest AIC value.

Table 10-11: Goodness of fit comparison for reported depth Experiment 2.

Model	AIC
$r \sim 1 + d * g + (1 + d * g \mid o)$	-39.62
$r \sim 1 + d * g + (1 + d + g \mid o)$	**-44.32**
$r \sim 1 + d * g + (1 + d \mid o)$	-43.11
$r \sim 1 + d * g + (1 + g \mid o)$	-41.59
$r \sim 1 + d * g + (1 \mid o)$	-40.92

Here, the affine transformation LME models are compared.

Table 10-12: Goodness of fit comparison for affine transformations Experiment 2.

Model	bz	b0	bx	by
$b \sim 1 + d * g + (1 + d * g \mid o)$	62.09	479.02	-273.14	-302.45
$b \sim 1 + d * g + (1 + d + g \mid o)$	59.47	471.72	**-280.8**	-305.8
$b \sim 1 + d * g + (1 + d \mid o)$	53.85	465.72	-277.18	**-308.78**
$b \sim 1 + d * g + (1 + g \mid o)$	55.61	465.96	-257.09	-299.46
$b \sim 1 + d * g + (1 \mid o)$	**51.62**	**461.97**	-257.69	-298.48

10.5 Chapter 7

This section provides more detail on the additional analysis.

10.5.1 Experiment 1

10.5.1.1 Participant 27

Upon visual inspection of the data, one participant was seen to have not completed the task as instructed, having set the slant and tilt such that the mesh was inverted, as if they had set the normals facing into the object instead of facing away:

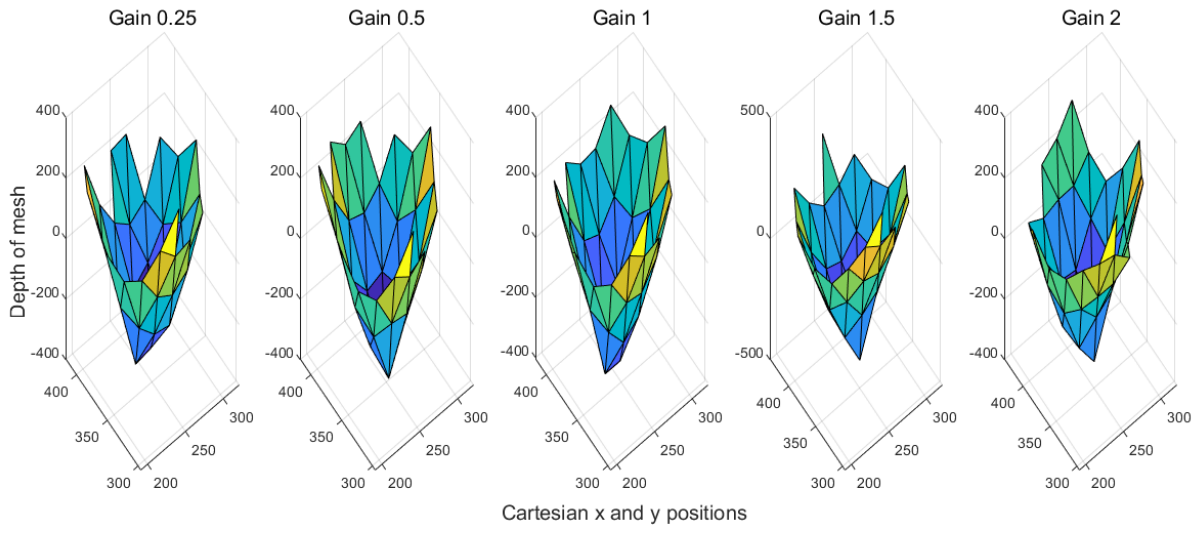


Figure 10.1: Participant 27. Plots showing meshes from Participant 27 which show the task completed incorrectly, with inverted meshes.

This participant was subsequently removed from the analysed data set, but here we provide analysis including this participant for completeness. The table presents the mean (standard deviations) of the multiple linear regressions used in the affine transformations.

Table 10-13: Showing means (SDs) of slope values from regression analysis in Experiment 1 including participant 27.

Coefficient	**Gain**				
Mean (SD)	**0.25**	**0.5**	**1**	**1.5**	**2**
Slope of intercepts	-0.34 (3.21)	-3.02 (14.25)	0.00 (0.00)	-2.98 (15.10)	-2.66 (12.01)
Slope of X position	-0.01 (0.14)	0.03 (0.08)	0.00 (0.00)	0.01 (0.12)	-0.01 (0.11)
Slope of Y position	-0.060 (0.13)	0.03 (0.29)	0.00 (0.00)	0.04 (0.43)	0.13 (0.33)
Slope of Z position	1.04 (0.39)	1.09 (0.40)	1.00 (0.00)	0.97 (0.38)	0.89 (0.28)

Results for depth transformation as shown by the slope of the z positions were significant both when the data was analysed with participant 27 present and without in the main analysis. This shows that the perceived depth, the area of interest in this work, did not differ significantly with the removal of participant 27. However, there was a significant result for shearing of the y axis which was not present with participant 27 included, which follows given the opposite pattern of results for this dimension from this participant. However, as the shearing of y is not a main concern of the work, this does not affect the overall results.

The goodness of fit for the model including participant 27 was similar to the main analysis, with both producing similar estimates and significance values:

Table 10-14: Goodness of fit comparison for affine transformations Experiment 1 including participant 27.

Model	AIC	Estimate	*p* Value	Lower CI	Upper CI
$bz{\sim}1 + g + (1 + g\vert o)$	43.24	-0.0986	**.003****	-0.1630	-0.0343

10.5.1.2 Analysis including extreme gradient vertex

One vertex was found to be causing extreme gradients. Analysis was completed without this point, although the pattern of results was not found to vary. Analysis including this point is presented here for completeness.

Table 10-15: LME results including vertex producing extreme depth gradients.

Coeffi-cient	Model	AIC	Estimate	*p* Value	Lower CI	Upper CI
bz	$bz{\sim}1 + g + (1 + g\vert o)$	43.24	-0.0999	**.003****	-0.1666	-0.0332
b0	$b0{\sim}1 + g + (1 + g\vert o)$	653.31	0.0124	.969	-0.6128	0.6376
bx	$bx{\sim}1 + g + (1 + g\vert o)$	-351.23	-0.0175	.050	-0.0350	0.0002
by	$by{\sim}1 + g + (1 + g\vert o)$	-195.80	0.146	**<.001*****	0.114	0.177

10.5.1.3 Goodness of fit of main analysis model

The model included in the main analysis includes the random slopes and intercepts, grouped by observer. Additional analysis was conducted to ensure goodness of fit of the model. The table below shows the AIC scores for each version of the model for all four coefficients:

Table 10-16: Goodness of fit comparison for affine transformations Experiment 1.

Coefficient	Model	AIC
bz	$bz \sim 1 + g + (1 + g \mid o)$	**-18.90**
	$bz \sim 1 + g + (1 \mid o)$	-15.88
b0	$b0 \sim 1 + g + (1 + g \mid o)$	580.92
	$b0 \sim 1 + g + (1 \mid o)$	**580.66**
bx	$bx \sim 1 + g + (1 + g \mid o)$	**-366.89**
	$bx \sim 1 + g + (1 \mid o)$	-364.48
by	$by \sim 1 + g + (1 + g \mid o)$	**-298.52**
	$by \sim 1 + g + (1 \mid o)$	-294.40

10.5.2 Experiment 2

10.5.2.1 Single object scene

The model included in the main analysis includes the random slopes and intercepts, grouped by observer. Here, the AIC is lower in general in the model without random slopes, however this does not affect the estimate and *p* value and confidence internals to a large extent, and therefore the main model was presented for completeness. Goodness of fit for the model for each of the four coefficients from AIC values is compared in the table below.

Table 10-17: Goodness of fit comparison for affine transformations Experiment 2 single object.

Coefficient	Model	AIC
bz	$bz \sim 1 + g + (1 + g \mid o)$	5.27
	$bz \sim 1 + g + (1 \mid o)$	**3.83**
b0	$b0 \sim 1 + g + (1 + g \mid o)$	358.85
	$b0 \sim 1 + g + (1 \mid o)$	**354.87**
bx	$bx \sim 1 + g + (1 + g \mid o)$	**-220.78**
	$bx \sim 1 + g + (1 \mid o)$	-219.07
by	$by \sim 1 + g + (1 + g \mid o)$	**-192.64**
	$by \sim 1 + g + (1 \mid o)$	-191.10

10.5.2.2 Cluttered scene object

Here, the LME models for each coefficient are compared. Estimates did not change when removing random slopes for any of the coefficients.

Table 10-18: Goodness of fit comparison affine transformations Experiment 2 scene object.

Coefficient	Model	AIC
bz	$bz \sim 1 + g + (1 + g \mid o)$	**13.618**
	$bz \sim 1 + g + (1 \mid o)$	26.514
b0	$b0 \sim 1 + g + (1 + g \mid o)$	353.47
	$b0 \sim 1 + g + (1 \mid o)$	**350.09**
bx	$bx \sim 1 + g + (1 + g \mid o)$	-137.38
	$bx \sim 1 + g + (1 \mid o)$	**-137.58**
by	$by \sim 1 + g + (1 + g \mid o)$	-236.28
	$by \sim 1 + g + (1 \mid o)$	**-240.02**

10.5.2.3 Object and scene comparison

Here the model comparisons are shown for the object and scene relative depth range analysis. The model chosen as having the lowest AIC and therefore best fit to the data was the random intercepts and random slopes of object or scene condition.

Table 10-19: Goodness of fit comparison relative depth Experiment 2.

Model	AIC
$r \sim 1 + g * os + (1 + g * os \mid o)$	169.86
$r \sim 1 + g * os + (1 + g + os \mid o)$	170.25
$r \sim 1 + g * os + (1 + g \mid o)$	170.51
$r \sim 1 + g * os + (1 + os \mid o)$	**168.03**
$r \sim 1 + g * os + (1 \mid o)$	172.82

10.6 Chapter 8

10.6.1 Depth range

Table 10-20: Goodness of fit comparisons for depth range.

Analysis	Model	AIC
Full face	$dr \sim 1 + g + (1 + g \mid o)$	**-189.20**
	$dr \sim 1 + g + (1 \mid o)$	-25.10
Localised features	$dr \sim 1 + g + (1 + g \mid o)$	**-191.37**
	$dr \sim 1 + g + (1 \mid o)$	-20.86

10.6.2 Affine transformations

Here, the models are compared for the affine transformations. In both the full face and the localised feature analysis the random intercepts and slopes model was the best fit to the data from the lowest AIC values.

Table 10-21: Goodness of fit comparison for affine transformations.

Analysis	Model	bz	b0	bx	by
Full face	$b \sim 1 + g + (1 + g \mid o)$	-195.54	23.05	-301.90	**-248.57**
	$b \sim 1 + g + (1 \mid o)$	-28.73	189.89	-138.82	-80.34
Localised features	$b \sim 1 + g + (1 + g \mid o)$	-166.97	6.84	-283.23	**-195.75**
	$b \sim 1 + g + (1 \mid o)$	-4.81	162.94	-110.18	-16.72

10.6.3 Surface roughness

Table 10-22: Goodness of fit comparisons for surface roughness full face.

Analysis	**Model**	**AIC**
Full face	$r \sim 1 + g + (1 + g \mid o)$	-24.32
	$r \sim 1 + g + (1 \mid o)$	**-30.46**
Localised features	$r \sim 1 + g + (1 + g \mid o)$	**-47.16**
	$r \sim 1 + g + (1 \mid o)$	-38.31