

Spatial and human factors affecting image quality and viewer experience of stereoscopic 3D in television and cinema

Paul Hands

Doctor of Philosophy

Institute of Neuroscience, Faculty of Medical
Sciences

June 2016

Abstract

The horizontal offset in the two eyes' locations in the skull means that they receive slightly different images of the world. The visual cortex uses these disparities to calculate where in depth different objects are, absolutely (physical distance from the viewer, perceived very imprecisely) and relatively (whether one object is in front of another, perceived with great precision). For well over a century, stereoscopic 3D (S3D) technology has existed which can generate an artificial sense of depth by displaying images with slight disparities to the different retinas. S3D technology is now considerably cheaper to access in the home, but remains a niche market, partly reflecting problems with viewer experience and enjoyment of S3D. This thesis considers some of the factors that could affect viewer experience of S3D content. While S3D technology can give a vivid depth percept, it can also lead to distortions in perceived size and shape, particularly if content is viewed at the wrong distance or angle. Almost all S3D content is designed for a viewing angle perpendicular to the screen, and with a recommended viewing distance, but little is known about the viewing distance typically used for S3D, or the effect of viewing angle. Accordingly, Chapter 2 of this thesis reports a survey of members of the British public. Chapters 3 and 4 report two experiments, one designed to assess the effect of oblique viewing, and another to consider the interaction between S3D and perceived size. S3D content is expensive to generate, hence producers sometimes "fake" 3D by shifting 2D content behind the screen plane. Chapter 5 investigates viewer experience with this fake 3D, and finds it is not a viable substitute for genuine S3D while also examining whether viewers fixate on different image features when video content is viewed in S3D, as compared to 2D.

Acknowledgements

This thesis was part-funded by BSkyB and EPSRC as a CASE PhD studentship supporting PH. Thanks to Suzanne Pinkney and Ann Fitchett for running the volunteer recruitment scheme. Data in chapters 2, 3, and 5 was collected solely by author PH for use in this thesis. Data in chapter 4 was collected by author PH or undergraduate students, Patrick Taylor and Aniketa Khushu, for use in this study. Thanks to Chris Johns and staff at BSkyB for assistance preparing the film clips used in chapter 5. Thanks to Masuji Suto for the use of the stereo Movie Maker used in chapter 5. The conversion of the EyeLink® 1000 Edf files in chapter 5 was done with the Edf2Mat Matlab Toolbox designed and developed by Adrian Etter at the University of Zurich.

Contents

1. Introduction	1
1.1. S3D display technologies	3
1.2. Filming S3D content	5
1.3. The different cues to depth	8
1.4. The vergence/accommodation conflict	13
1.5. Aims	15
2. Television viewing distance and S3D viewing habits in British households	17
2.1. Introduction	17
2.2. Material and methods	21
2.2.1. Participants	21
2.2.2. Procedure	21
2.3. Results	23
2.3.1. S3D viewing	23
2.3.2. Viewing distance and screen size	26
2.4. Discussion	30
2.5. Conclusion	34
3. Stereoscopic 3D content appears relatively veridical when viewed from an oblique angle	35
3.1. Introduction	35
3.2. Materials and methods	41
3.2.1. Participants	41
3.2.2. Apparatus	42
3.2.3. Stimulus generation	43
3.2.4. Experimental design	44
3.2.5. Experimental parameters	47
3.2.6. Modelling	48
3.2.7. Fitting	51
3.3. Results	51
3.3.1. Statistical analysis	55
3.3.2. Sensitivity to rendering angle	60
3.3.3. Effect of oblique viewing angle	61
3.3.4. Model fitting	64
3.3.5. Quantifying the preference for normal rendering vs. geometrical correctness	66
3.3.6. Effect of S3D	69
3.3.7. Effect of frame occlusion	70
3.3.8. Effect of object motion	71
3.4. Discussion	72
3.4.1. Sensitivity to viewing angle	73
3.4.2. Range over which compensation operates	74
3.4.3. Regression to expected shape	74
3.4.4. Differences between stereoscopic 3D and 2D	75
3.4.5. Effect of frame visibility and object motion	76
3.4.6. Limitations	77
3.5. Conclusion	79
4. The interaction between familiar size and vergence depth cues in stereoscopic 3D displays	80
4.1. Introduction	80
4.2. Initial experiment	85

4.2.1. Material and methods - Participants	85
4.2.2. Material and methods – Equipment	85
4.2.3. Material and methods – Procedure	86
4.2.4. Material and methods – Parameters	88
4.2.5. Results	90
4.2.6. Discussion	94
4.3. General methods	95
4.3.1. Participants	95
4.3.2. Equipment	97
4.3.3. Procedure	97
4.3.4. Parameters	98
4.3.5. Mathematical modelling	98
4.4. Experiment 1: Vergence-based depth judgements	101
4.4.1. Methods	101
4.4.2. Model	102
4.5. Experiment 2: Size judgements in 2D	102
4.5.1. Methods	102
4.5.2. Model	103
4.6. Experiment 3: Size/vergence interaction	103
4.6.1. Methods	104
4.6.2. Model	104
4.7. Results	104
4.8. Discussion	115
4.9. Conclusions	119
5. An eyetracker study on shifting the screen plane	121
5.1. Introduction	121
5.2. Material and methods	128
5.2.1. Equipment	128
5.2.2. Stimuli	128
5.2.3. Procedure	131
5.2.4. Participants	132
5.3. Results	133
5.3.1. Subjective ratings	133
5.3.2. Vergence eye measurements	134
5.3.3. Gaze position during the clips	139
5.3.4. Fixation analysis	150
5.4. Discussion	153
5.4.1. Subjective ratings	153
5.4.2. Eyetracking discussion	156
5.5. Conclusion	158
6. Conclusion	160
7. References	165

List of tables and figures

Fig. 1.1. Gemetrical diagram explaining S3D displays	3
Fig. 1.2. Side by side and mirror rig configuration	7
Fig. 1.3. Different possible camera configurations	8
Fig. 1.4. Example of optimal cue combination	12
Fig. 1.5. An example of 3D structure from 2D	14
Fig. 2.1. Mathematical relationship between viewing distance and screen size	19
Table 2.1. List of questions asked in survey	22
Fig. 2.2. Histogram of responses to question 3	24
Fig. 2.3. Histogram of number of people participants watched television with	25
Fig. 2.4. Comparison of viewing distance and screen size	26
Fig. 2.5. Comparison with forced zero intercept	27
Fig 2.6. Comparison of 3D only data	28
Fig. 2.7. Comparison of 3D only, forced zero intercept	29
Fig 2.8. Comparison with Lund and Ardito estimations	33
Fig 3.1. Explanation of how different seats when watching content can affect viewing angle	37
Fig 3.2. Experimental setup and example stimulus	46
Fig 3.3. Rendering for a different viewing angle warps object shape	48
Fig 3.4. Model predictions with perfect rendering	50
Fig 3.5. Data only, frame visible results	52
Fig 3.6. Data only, frame occluded results	53
Table 3.1. Fitted model parameters	58
Table 3.2. Main effects and interactions	60
Fig 3.7. Frame visible results	63
Fig 3.8. Frame occluded results	64
Table 3.3. Main effects and interactions, only considering when the geometrically correct cube was not perpendicular	65
Table 3.4. Main effects and interactions of compensation index	68
Fig 3.9. Compensation index, considered in terms of binocularity condition	69
Fig 4.1. How perceived card size would change with depth	87
Fig 4.2. Credit card stimulus	88
Fig 4.3. Credit card stimulus on noisy background	88
Table 4.1. Experimental widths and corresponding parallaxes	89
Fig 4.4. Psychometric results for black and noisy backgrounds	92
Table 4.2. Participant information for separate distances	96
Fig 4.5. Bias and noise effect on size signal detection	100
Fig 4.6. Bias and noise effect on parallax signal detection	100
Table 4.3. Factors and interaction ANOVA for different backgrounds	105
Fig 4.7. Data and model fits, viewing distance of 1m, black background	107
Fig 4.8. Data and model fits, viewing distance of 0.5m, black background	108
Fig 4.9. Data and model fits, viewing distance of 1m, noisy background	109
Fig 4.10. Data and model fits, viewing distance of 0.5m, noisy background	110

Fig 4.11. Data and model fits, viewing distance of 2m, noisy background	111
Fig 4.12. Comparison of weight and noise ratios, in Log coordinates	112
Fig 4.13. Data and 4 parameter model fits, viewing distance of 2m, black background	113
Fig 4.14. Data and 4 parameter model fits, viewing distance of 0.5m, noisy background	114
Fig 5.1. Geometry of experiment, to explain shift concept	130
Fig 5.2. Explanation of shifted clip concept	130
Fig 5.3. Depth quality scores for all viewing conditions across participants	133
Fig 5.4. Parallax measurements for all viewing conditions for one participant	135
Fig 5.5 Mean parallax measurements across clips for all subjects	137
Fig 5.6. Eye gaze position for one frame of three clips.	140
Fig 5.7. Example correlation of gaze position for configurations	142
Fig. 5.8. Heat map for participant 1	144
Fig. 5.9. Heat map averaged across participants and gaze positions	145
Table 5.1. Correlation values for all participants	146
Table 5.2. Mean correlation values for each configuration	149
Fig 5.10. Participant example of fixations for one clip	150
Fig 5.11. Fixation durations per clip	151
Fig 5.12. Number of fixations per clip	152
Table 5.3. Repeated measures ANOVA analysis on fixations	153

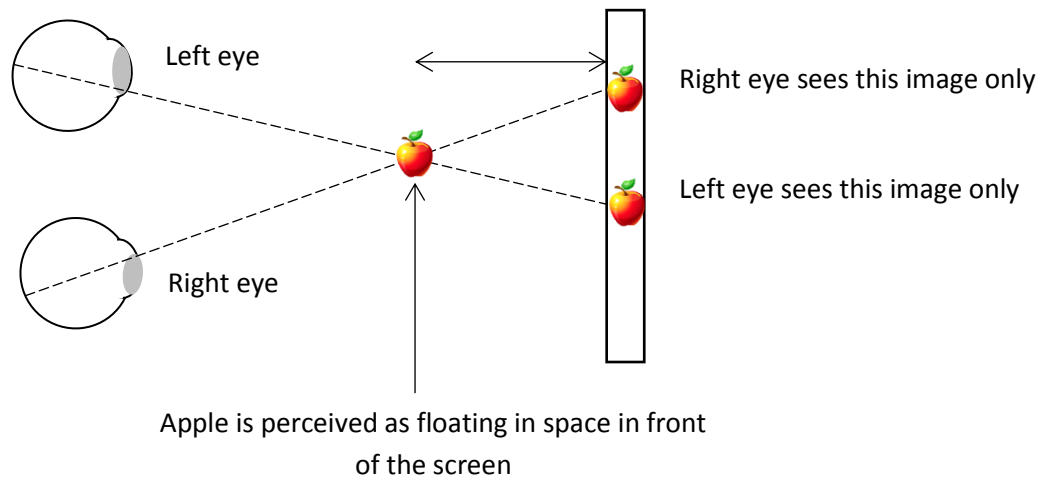
1 1. Introduction

2 Humans use the visual system, comprising of the eyes, through the central nervous
3 system up to the brain, to see the environment around them, by detecting and correctly, or
4 sometimes incorrectly, interpreting light in the visible spectrum. The information from the
5 light travels along the optic nerve once it enters the eye, and the majority of the axons from
6 the nerve go to the lateral geniculate nucleus, located in the thalamus of the brain (Nave,
7 2014). The information is then passed into the primary visual cortex (also known as V1, or
8 the striate cortex) and ascends through the different levels of the visual hierarchy (V2
9 neurons, V3, etc.). Information processing gets more refined as the signal goes up the
10 hierarchy, with lines and contours of specific orientation causing neurons to fire in V1 and
11 V2, but more complex objects are responded to in the higher regions of the visual cortex.
12 There is still discussion over the exact way this information is processed. A strong hypothesis
13 is that the dorsal and ventral streams are two distinct pathways. In this hypothesis
14 information in the dorsal stream is related to spatial attention and awareness (and hence
15 the dorsal stream is considered the 'where' stream). The ventral stream (the 'what' stream)
16 is believed to be associated more with object classification and recognition (Mishkin, M. et
17 al., 1982). Bishop and Pettigrew first discovered disparity selective cells in the striate cortex
18 (V1) neurons of a cat. This discovery was not expected and hence was not published until
19 over ten years later (Bishop, P. & Pettigrew, J. D., 1986). However, despite being present as
20 early in the visual system as V1, binocularly sensitive cells exist throughout the neural
21 pathway, with most of the cells beyond V4 being binocular. However neurons in V1 are only
22 sensitive to absolute disparity, whereas the disparity selective neurons beyond V1 are
23 sensitive to relative disparity (Parker, A., 2007).

24 The visual system extracts different information about the world the viewer perceives, such
25 as, taking spectral information to consider colour, texture and luminance in the world. Each

26 eye, due to its structure, is only capable of taking a 2D retinal image of the world that is being
27 perceived in any given time and location. A vital challenge faced by the visual system is the
28 necessity to reconstruct the 3 dimensional world based on two different 2D retinal images.
29 It is possible to generate some information about the depth in the scene from monocular
30 cues, such as occlusion, perspective lines and motion parallax (Banks, Read, Allison and Watt;
31 2012). However humans have evolved two separate eyes in different locations in the skull,
32 which allows for the two retinal images to be combined in the visual cortex to establish more
33 depth information than would be possible with only one eye, by the use of binocular vision
34 and stereopsis.

35 Due to the horizontal offset between the two eyes that humans have, they receive slightly
36 different retinal images when viewing the environment around them. These small binocular
37 disparities are detected in the visual cortex of the brain, which uses the information from the
38 disparity to draw inferences about the depth of objects around them. Even in the absence of
39 other depth cues, these disparities suffice to create a vivid perception of depth (Julesz, 1971).
40 This effect is exploited in stereoscopic 3D (S3D) displays, which present separate images to
41 the two eyes. This horizontal offset, were the eyes locked in position, would be enough to
42 calculate absolute depth, accurately and quickly. However the eyes can move in their
43 sockets, via the rectus and oblique muscles around the eye. This means that humans need
44 to be able to calculate the vergence angle (the angle between the line of vision from each
45 eye). Humans are not very good at estimating this vergence angle though, which is a
46 reflection as to why they are more confident at calculating relative depth (i.e. depth in terms
47 of 'nearer' or 'further away' from a different object) compared to absolute depth (i.e. depth
48 in a metric sense, measured in some unit such as centimetres or metres). A diagram showing
49 the basic geometry of stereopsis, particularly for S3D technology, is shown below in fig. 1.1.



50

51 **Fig. 1.1.** A diagram of the basic geometry of S3D displays and stereopsis. Using disparity, the
 52 apple can be perceived as floating in front of the screen. In this instance estimating the
 53 physical distance the apple is away from the viewer would be absolute disparity. Being able
 54 to ascertain (via disparity) that the apple is displayed in front of the screen is relative
 55 disparity.

56 1.1 S3D display technologies

57 The first S3D display was the Wheatstone stereoscope (Wheatstone, 1838).
 58 Although the concept of S3D technology has therefore been around for nearly two centuries,
 59 recent improvements in the field have allowed expansion into exciting new territories, such
 60 as medical surgery (McCloy & Stone, 2001), and home S3D cinema systems. Cinema theatres,
 61 home television systems with S3D capabilities and some game consoles use different types
 62 of S3D displays, including passive and active stereo and parallax barrier technology (Karajeh,
 63 Maqableh, & Masa'deh, 2014).

64 Passive S3D displays show the left and right eye images at the same time on the screen. The
 65 two most common forms of passive display use colour anaglyph (red/blue or red/green) or
 66 polarising filter technology. With polarising filters, the images are typically presented row
 67 interleaved (i.e. the odd pixel rows show the left eye image and the even show the right eye,

68 or vice versa), and a polarising filter is used to separate the two images and display only the
69 desired pixel rows to each eye. Polarising filter is better for colour preservation but there is
70 still a loss in vertical resolution with the row-interleaving technique.

71 To avoid this, active displays show the left and right eye images on subsequent frames (i.e.
72 frame 1 left eye, frame 2 right eye, etc.) and have shutter glasses which obstruct the
73 appropriate eye in synchrony with the display. If the image is shown at a high enough frame
74 rate then flicker is not perceived (Fröhlich et al., 2005), although other perceptual artefacts
75 can occur due to the temporal delay between left and right eyes (Hoffman et al., 2008).

76 Parallax barrier technology has both images displayed at the same time in a similar procedure
77 to that of passive displays. However the filter used does not distort the colour. In industry
78 this technology is sometimes referred to as autostereoscopic, that is, S3D without the need
79 for glasses. A mesh barrier is applied on top of the image, so if the viewer is in the correct
80 position the barrier enables each eye to see only the content intended for it. This 'sweet
81 spot' is usually quite small in conventional displays. Progress is being made to create larger
82 autostereoscopic displays with a larger number of more lenient sweet spots (that is, sweet
83 spots that are larger for viewers, so some head movement is acceptable) (Woodgate & Ezra,
84 1995).

85 A developing display technology is that of virtual reality (VR) and augmented reality (AR).
86 Typically in these technologies the display is head mounted, and allows for a much more
87 immersive experience. Many different elements of the display can be manipulated, such as
88 using optics to choose at what depth to display the content, despite the screens being
89 physically very close to the eyes of the viewer. Each viewer has their own display set while
90 using head-mounted VR and AR, and this allows content producers to be confident on the
91 viewing distance and angle the viewer is sat at. However this technology is not without flaws.
92 It is still relatively expensive, although models such as the Oculus Rift are decreasing in price,

93 and there are issues with lag and conflict with vestibular cues (with VR), which is a
94 contributing factor to the regularly reported dizziness and motion sickness associated with
95 these technologies (Azuma et al., 2001).

96 The disparity information from the S3D display can be further combined with pictorial depth
97 cues such as, for example, perspective, shading and occlusion (Cavanagh, 1987) to generate
98 an immersive sense of depth, which is then experienced by the viewers of S3D.

99 With advancing technology, stereoscopic displays are becoming a part of everyday life. A
100 large portion of this is due to S3D TV becoming more readily available for home viewing
101 through cutting edge media systems, and the technology becoming cheaper and more
102 readily available (Noland & Truong, 2015). This technology has been used to excellent effect
103 in films to add an extra appeal to movie theatre visitors, in an attempt to increase revenue,
104 with films such as *Gravity* earning many plaudits for their intentional and pre planned use of
105 S3D technology.

106 While well considered S3D can generate a great depth percept for viewers, and add an extra
107 element to the enjoyment of the media being shown, the very concept that allows for the
108 introduction of depth to the image (manipulating the disparity and displaying to each eye
109 individually) can lead to distortions in perceived size and shape (Foley, 1968). This can occur
110 because previous assumptions based on 2D content may not necessarily hold true for S3D
111 content. The deviations from 2D content and the potential distortions that can arise in S3D
112 content perception is a large area of academic research.

113 1.2 Filming S3D content

114 S3D content production is a complicated procedure with a lot of specialist words and
115 concepts. In this introductory section some of these concepts are explained in detail to allow
116 the reader to understand more easily the subsequent chapters.

117 There is an important distinction between the terms parallax and disparity as used in this
118 thesis. Binocular disparity refers to the angular difference, on the retina, of the images of an
119 object seen by the left and right eyes, resulting from the eyes' horizontal separation. The
120 brain uses binocular disparity to extract depth information about the object in question from
121 the two-dimensional retinal images in stereopsis. Binocular parallax, as defined here, is the
122 actual horizontal offset between the left and right images on screen, as displayed together
123 on an S3D display using some form of filtering process.

124 S3D content is filmed on 2 separate cameras, using a complicated rig setup, to attempt to
125 mimic the two eyes' viewpoints of the scene. The rigs are typically either in a side by side or
126 mirror configuration, as displayed below in Fig. 1.2. The distance that the two cameras lenses
127 are apart is known as the interaxial distance, which can be considered analogous to the
128 interocular distance (IOD), or interpupillary distance in humans (the distance between the
129 centre of the two eyes' pupils).

130 In the side by side configuration the minimum value that the interaxial value can be is limited
131 by the size of the cameras, and is usually larger than the average IOD of 63mm (Dodgson,
132 2004), with a typical interaxial value of 7 inches. The mirror rig is more advanced, and more
133 expensive, positioning the cameras perpendicular from one another and using mirrored glass
134 at an angle of 45° to film the content. This configuration allows for an interaxial value of zero
135 by aligning the cameras up on top of one another, and hence the average human IOD can be
136 achieved with this configuration. To get around the lack of possible small interaxial values
137 for the side by side configurations, the cameras can be 'toed in'. That is, they can be rotated
138 slightly, towards one another, in an attempt to resolve the issue of the large interaxial value.

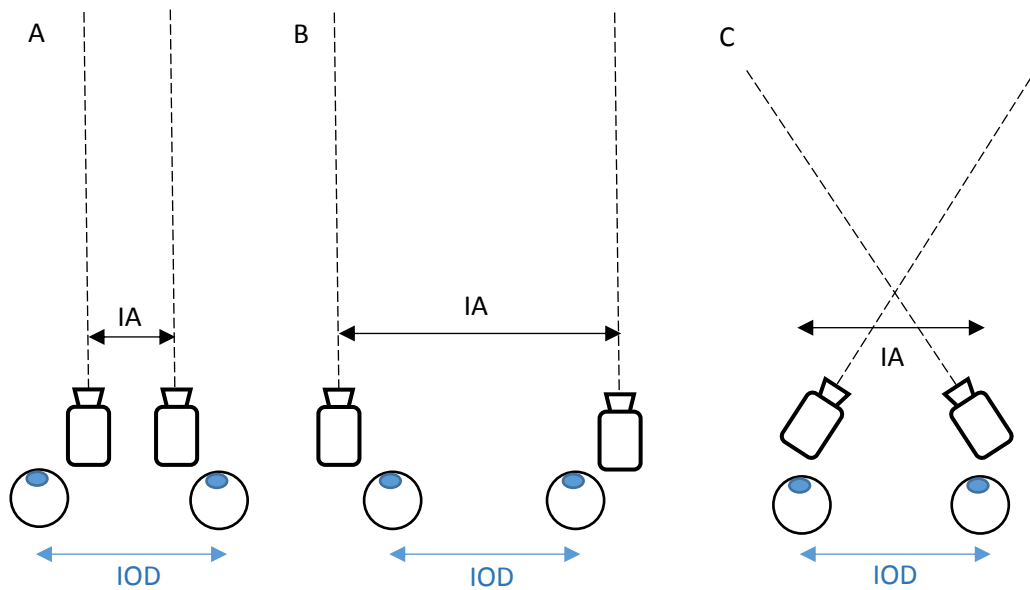


139

140 **Fig 1.2.** Side by side camera configuration (available at
141 [http://www.dashwood3d.com/blog/wp-content/uploads/2010/05/side-by-side-rig-on-](http://www.dashwood3d.com/blog/wp-content/uploads/2010/05/side-by-side-rig-on-white.png)
142 [white.png](http://www.dashwood3d.com/blog/wp-content/uploads/2010/05/side-by-side-rig-on-white.png)) and mirror rig configuration
143 (http://www.amplis.com/evolve/ca_images/genus/3D%20Rig/3D-camera-rig-1.jpg)

144 If the interaxial value of the setup is too large or too small then something can occur known
145 as the puppet theatre effect. This is an issue known in the industry, where the configuration
146 of the cameras can result in a too large or too small parallax between the displayed images
147 for the left and right eyes, and hence objects can appear either too large or too small,
148 resulting in them looking like puppets, attempting to mimic real life, rather than a realistic
149 S3D image. This is discussed in more detail in section 4.1

150 One possible future approach in filming and creating S3D content is to attempt to reproduce
151 the image orthostereoscopically or “orthostereo”. Orthostereo is showing the left and right
152 eye of the viewer the exact left and right image as if they were themselves in place of the
153 camera. This is a very complicated and difficult thing to do, as the interaxial values must
154 match the viewer’s IOD, and viewing distance, and viewing angle, need to also be identical
155 to that of the camera. This is discussed in more detail in section 3.1. However, if the content
156 is not rendered as orthostereo, the differences between the way the human visual system
157 views S3D and the way the content is created can cause distortions in shape, size, colour and
158 perception, leading to issues like the puppet theatre effect described above.



159

160 **Fig. 1.3.** Explanatory diagram explaining the different camera configurations possible and
 161 the most likely perceptual outcome. I show a standard interocular distance (IOD) and
 162 change the interaxial (IA) distance. A) IA too small for IOD, which typically results in
 163 gigantism. B) IA too large for IOD, which typically results in miniaturisation. C) Toed-in
 164 configuration, which can lead to miniaturisation.

165 1.3 The different cues to depth

166 It is important to note that in the commercial world it is very rare for content to be
 167 created where stereoscopic binocular disparity is the only depth cue. There are almost
 168 always other depth cues, e.g. such as perspective, shading, texture, and motion parallax.
 169 However this is enough to generate a vivid, clear impression of change in depth of a 2D
 170 image, as shown using the random dot stereograms designed by Julesz (Julesz, 1986). In the
 171 more clinical and controlled lab environment that most experiments are conducted in, it is
 172 important to control for anything that could also affect the perception of the stimulus. Hence
 173 most experimental stimuli used in research are less like natural environments and very much
 174 controlled images, such as sine wave gratings (Legge & Yuanchao, 1989) and random dot
 175 stereograms. In these experiments participants are regularly positioned using something to

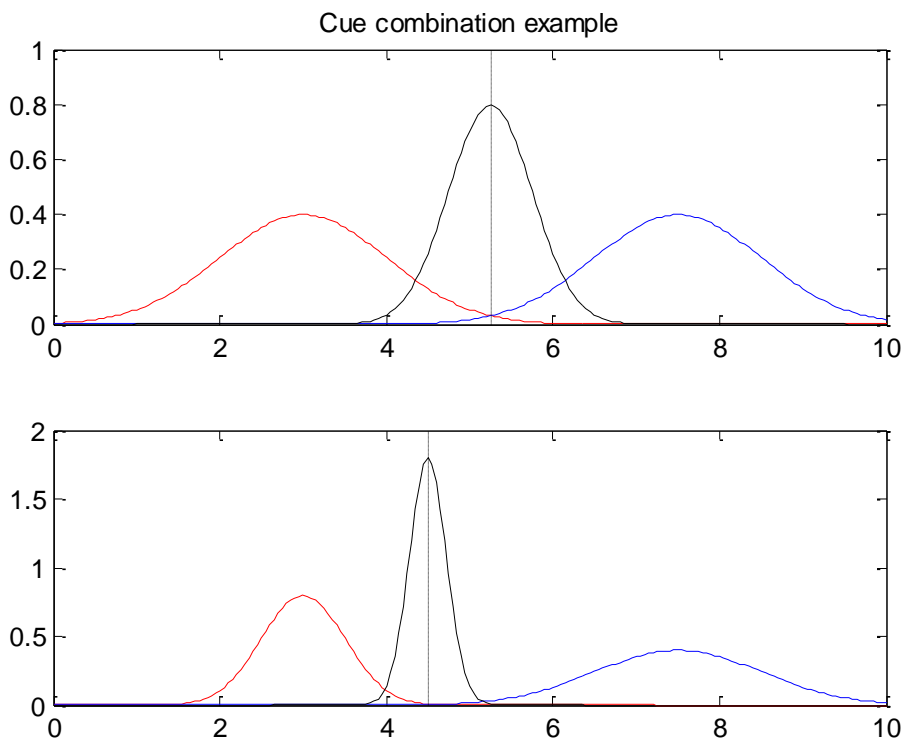
176 limit any movement, such as a bite bar or a chin rest. This also enables the experiment to
177 control for any motion parallax cues that might be available (Ames Jr, 1925). Pictorial cues,
178 such as shading, texture and motion, are affine cues to depth, whereas cues to depth from
179 binocular vision are metric, allowing the viewer (from the disparity information) to establish
180 definitely a distance between two points in the picture, relative to the rest of the depth in
181 the image. This isn't possible for affine cues from pictures, as the cues only provide a
182 suggestion as to relative depths in the image scene, allowing the viewer to estimate, for
183 example, local curvature of an object using shading (Di Luca, M., Domini, F. & Caudek, C.,
184 2006). It is an important consideration that while disparity driven depth is considered metric,
185 this is only in relation to the estimate of fixation depth that the user has at that time. Because
186 of this the actual depth information provided by disparity is necessarily relative, and
187 transformed into metric, based on this assumption of fixation depth (Foley, J. M., 1985).
188 Humans typically underestimate depth (Plumert, J. et al., 2005) which would reduce the
189 values of the estimated metric. It is because of the physical, measurable distances between
190 objects in depth, provided by the horizontal offset in both eyes images, that S3D works so
191 well in cinema and television. However this fixation distance judgement (and hence the
192 resulting disparity driven metrics being estimated) could potentially lead to problems, such
193 as incorrectly estimating the fixation distance. Hence the disparity driven metrics would also
194 be incorrect, leaving the viewer perceiving something that doesn't look as intended by the
195 producer of the S3D content (a good example of this is miniaturisation, which is studied in
196 more depth in chapter 4). Depth from motion parallax is only defined up to scale unless the
197 viewer knows for certain the distance that they (or the object) has moved, and is hence
198 another form of relative depth perception.

199 Like binocular disparity, information regarding the depth of different objects can be detected
200 directly using relative motion (Gibson, 1950). In this instance the geometry applied is exactly
201 the same as stereopsis; however rather than considering two views at once the visual system

202 considers two views at two different times. If these two cues to depth provide different
203 depth information about the scene, some people may perceive the scene with the disparity
204 defined depth; some may perceive the depth in the scene as defined by the relative motion
205 cue, and some may perceive the depth in the scene as some combination of the two cues.
206 This is a situation known as a cue conflict, where two elements provide different information
207 to the brain (in this case specifically, the visual cortex). The way that the different cues to
208 depth are combined in the visual cortex is an interesting area of research, and the
209 mathematics behind it can be quite complex. A key consideration in this area of research is
210 that conducted by Hillis et al. in which they establish that cues are combined in a statistically
211 optimal fashion, in so much that as the reliability of the information provided by the cue
212 decreases, the weighting of that cue in the combination also decreases (Hillis, J. M. et al.,
213 2004). This study provided further support to the theory that cues are combined using a
214 maximum likelihood estimating model, combining the information in an attempt to maximise
215 the accuracy of the signal detection from the cues. A confounding factor of this theory is that
216 reliability estimates of the cues are not constants, but vary according to the viewing
217 geometry the observer has at that time. This means that as the viewing distance increases,
218 the reliability of the different signals will not change in an identical fashion, resulting in a
219 change in viewing distance changing the weighting of the two cues. Mathematically, in the
220 context of depth, if you consider combining two different cues there are three outcomes that
221 are possible: Each one of the two cues can be preferred over the other cue explicitly, or a
222 weighted summation of the two cues occurs. In a weighted summation the explicit cue
223 preference case is still possible, as the weighting for that cue is reduced to zero. If the
224 discrepancies between the cues are large (known as robust cue integration) then the less
225 reliable cue is vetoed, and effectively ignored. In the case of smaller cue combinations a
226 weighted summation is optimised to get the most precise estimate to the depth from the
227 cues provided. Research by Knill (2007) considered robust cue integration with a maximum

228 likelihood estimator model and found that the while the less reliable cue's weighting does
229 decrease significantly, it never has a weight of zero in the sum. This would question whether
230 the cue can be completely ignored in combination of cues when the discrepancy is large
231 (Knill, 2007), or whether a weighted summation would be a more accurate consideration,
232 with the less reliable cues weight being considerably smaller than the more reliable cue.
233 Ernst and Banks (2002) consider a combination of cues in their study (in their case the cues
234 are visual and haptic cues). They found that the nervous system in general combines cues in
235 a statistically optimal fashion, very much like the maximum likelihood estimating models
236 considered above. The model attempts to minimise the variance by combining the variances
237 from the two cues. This is directly transferable to combining the reliability of the cues to get
238 the most precision in the estimate of the metric being considered. In cue combination usually
239 the two cues can be approximated as coming from separate Gaussian distributions, in which
240 the more reliable cue has a smaller variance (and hence standard deviation). The resulting
241 combination of the cues assigns more weight to the consideration of that respective signal,
242 so as to maximise the precision and minimise the variance. Fig. 1.4. below gives a brief
243 example of the cue combination concept. In the top figure both Gaussian curves, intended
244 to represent the different signals being detected, have equal variance and hence are equally
245 reliable on the precision of the cue, and so the mean value that the cue combination model
246 settles on (represented by the black dotted line) is directly in the middle of them. The
247 combined Gaussian distribution is shown for the optimal signal estimation. In the bottom
248 figure the means are the same as before but now the red curve has half the variance of the
249 blue curve, and is hence considered the more accurate and reliable cue. Because of this more
250 weight is assigned to the more reliable cue, and hence the estimate for the true value
251 perceived is more similar to this cue's value. This results in the black dotted line moving
252 closer to the more reliable signal. Landy et al. (1995) developed a model for cue combination
253 known as modified weak fusion, based on the Bayesian theories of cue combination used

254 elsewhere (Ernst and Banks, 2002). They consider three key issues of cue combination and
255 cue conflict: Robustness, weighting and promotion of different depth cues (Landy, Maloney,
256 Johnston & Young, 1995). The idea of combining cues in this manner, as well as the Bayesian
257 technique, is an interesting one and is considered in chapters of this paper.



258
259 **Fig 1.4.** Example figures to show cue combination optimisation. The red and blue curves are
260 example Gaussian curves representing two different signals that are being detected. The
261 width of the Gaussian (the variance associated with that Gaussian) is an indication of the
262 reliability of the signal - a smaller variance reflects a more reliable signal - and the peak of
263 each Gaussian is the most likely signal from that individual cue. In the top figure the signals
264 have equal variance (and hence are equally reliable), and in the bottom the variance of the
265 red curve is half of the blue curve (and is hence a more reliable cue). The black dotted line
266 represents how the cues would combine to choose a signal, with the combined Gaussian
267 distribution shown by the black curve.

268 Other cues to depth can interact with disparity to give a more vivid perception of depth, such
269 as motion, texture, and occlusion. Cavanagh (1987) conducted an experiment investigating
270 the interaction between binocular vision and occlusion, finding that the detection of
271 occlusion in line drawings appears to be analysed generally rather than specifically,
272 suggesting that detecting occlusion occurs somewhere in the visual cortex that has access to
273 all the different pathways of the visual system. It is therefore possible to use disparity to
274 reinforce depth perception from occlusion (and vice versa).

275 1.4 The vergence/accommodation conflict

276 Each eye uses the medial and lateral rectus muscles to perform fusional vergence.
277 This aligns the two eyes' images correctly onto a stimulus or object of interest. This process
278 is neurally connected to accommodation in the natural environment, a process in which the
279 ciliary muscles can contract or relax to alter the shape of the lens so that objects (at different
280 depths) stay clear and in focus (Howard & Rogers, 1995). Accommodation and vergence are
281 naturally connected such that a change in depth normally affects both the accommodation
282 and vergence (Banks, Read, Allison, & Watt, 2012; Hoffman, Girshick, Akeley, & Banks, 2008).
283 To recreate the perception of depth on a flat surface (i.e. a television monitor screen), the
284 mechanical connection of accommodation and vergence is necessarily broken by the
285 introduction of artificial disparity. Hence the eye accommodates to the depth of the screen
286 and verges to different depths depending on the disparities displayed in the scene on the
287 display screen. The decoupling of the accommodation and vergence cues is a depth cue
288 conflict. This results in S3D displays being slightly different to the natural environment, as
289 different depths (from disparity) all appear in focus on displays. In the natural world this isn't
290 the case, and objects at different depths to the focus appear blurry. This can be
291 approximated for in the displayed image by introducing blur artificially, but the object will
292 still be necessarily in focus, it is just a blurry image that is in focus. This also requires prior

293 knowledge of where the observer is looking in the scene, and is not feasible for content
294 generated in advance.

295 The vergence accommodation conflict is a good example of how S3D technology cannot
296 exactly reproduce reality. This inability to replicate the retinal image perfectly applies to 2D
297 content as well, in so much that while vergence, accommodation and disparity cues all
298 specify a flat surface displaying a completely flat image (and hence just a collection of lines
299 and colours), other depth cues such as perspective, shading and occlusion specify and
300 suggest 3D structure. An example is shown below in Fig. 1.5. where the image is displayed on
301 a flat surface (in this case drawn onto a piece of paper), but it appears, because of the lines,
302 that the drawn hand has 3D structure.



303
304 **Fig. 1.5.** An example of how 2D images can give an impression of 3D structure. Image
305 available at [http://www.yetanothermomblog.com/wp-content/uploads/2015/07/3D-hand-](http://www.yetanothermomblog.com/wp-content/uploads/2015/07/3D-hand-fb.jpg)
306 [fb.jpg](http://www.yetanothermomblog.com/wp-content/uploads/2015/07/3D-hand-fb.jpg)

307 In principle, when viewed with one eye through a pinhole from the centre of projection used
308 to generate the image, a 2D picture could recreate the same retinal image as a real scene.
309 However we never view 2D content in such a way. We are almost always viewing it from a

310 position other than the centre of projection. Based on this it is surprising that we do not
311 regularly perceive 2D pictures as distorted.

312 There is potential that the assumptions made about viewing S3D content based on previous
313 findings of viewing 2D content may not necessarily hold true, such as those associated with
314 viewing content from an oblique angle, which humans learn to account for in the perception
315 of the image on the retina (Vishwanath, Girshick, & Banks, 2005). During the young, plastic
316 years of the visual systems humans typically see in natural 3D (i.e. the environment they are
317 in) or in 2D (i.e. when viewing pictures and media in books, paintings, on televisions and
318 more recently on smartphones and tablets). This is in stark contrast to the time that young
319 people view S3D technology and displays, which is considerably smaller. The relatively novel
320 exposure of S3D technology to simulate 3-dimensional depth on a visual system most used
321 to natural 3-dimensional depth or 2-dimensional flat images may be the driving force behind
322 any perceived geometrical distortions or reported discomfort in viewing S3D content. This
323 may be particularly true if the content in question is viewed at a different distance or angle
324 to that which was desired upon creating the image in question, i.e. when the S3D content is
325 not viewed orthostereo. Almost all S3D content is designed for a normal viewing angle, i.e.
326 perpendicular to the screen, and each piece of content will have a desired viewing distance.
327 However little is known about the viewing distance typically desired for viewing S3D content,
328 or how the viewing angle can affect the perception of the content, in relation to potential
329 geometrical distortions and discomfort.

330 1.5 Aims

331 The overall aim of this thesis is to consider how different viewing conditions can
332 affect the perception of S3D content, and hence improve understanding of the different
333 factors that can affect viewer experience in S3D cinema and television. To achieve this,
334 Chapter 2 provides a report on a survey conducted on 559 UK members of the public

335 considering viewing habits in relation to viewing distance, and consumption of S3D content.
336 This was published as (Hands & Read, 2015). Chapter 3 considers an experiment designed to
337 assess the effect of oblique viewing of S3D content, and whether it can create geometrical
338 distortions in the structure of the images perceived. This work was published as (Hands,
339 Smulders & Read, 2015). Chapter 4 explores the interaction between vergence eye
340 movements from disparity and perceived size, with application to understanding the “puppet
341 theatre” effect that can be sometimes perceived in S3D content when the interaxial distance
342 (the distance between the left and right eye cameras) exceeds the interocular distance.
343 Producers sometimes attempt to ‘fake’ S3D by introducing a uniform parallax between
344 previously identical left and right images, hence shifting 2D content backwards. This is
345 because S3D content is expensive and difficult to generate, and producers therefore attempt
346 to substitute true S3D, and instead show the 2D clips as if through a window, in the hope
347 that this will be perceived as true S3D content. This work was presented at the Vision
348 Sciences Society meeting 2016. Finally, I am interested in the effect that showing content in
349 S3D may have on fixation. As an illustration it is possible that a particular region of the image
350 is not salient in 2D, however, in S3D, disparity in the image may cause that region to be more
351 likely to attract the viewer’s attention. In Chapter 5, I investigate viewer experience with
352 ‘fake S3D’ and show that while the impression of depth is marginally better than 2D content,
353 it is no substitute for genuine S3D content, and I use eyetracking technology to examine
354 whether viewers fixate on different video features when content is viewed in S3D as
355 compared to 2D.

356 2. Television viewing distance and S3D viewing habits in 357 British households

358 2.1. Introduction

359 Since the introduction of the first S3D movies in the 1920s, such as *The Power of*
360 *Love*, from Los Angeles (Zone, 2007), certain decades have shown considerably more S3D
361 movies than others, with some decades having little to no S3D releases. This wave pattern
362 of S3D movie popularity suggests that if the popularity of S3D movie viewing slows down at
363 any point that does not mean it will not resurface at a later time. Since the 1920s the
364 technology used to show S3D content has evolved considerably, with footage now able to be
365 viewed in full color with active shutter glasses technology. The technology has also become
366 easier to introduce to mainstream televisions, at a lower price, so people can view S3D at
367 home without needing to risk losing the impact of the original piece by viewing in colour
368 anaglyph. As televisions with active and polarised S3D technology became considerably
369 cheaper, it was assumed that the most recent upsurge in S3D movies, spearheaded by the
370 release of *Avatar* in 2009, would be a permanent increase. However, S3D viewing at home
371 remains the exception rather than the norm.

372 Another aspect of the evolving technology is that television sets have become much larger
373 and thinner. A 50 inch television set cost \$20,000USD in 2000 (Darlin, 2005), compared to
374 \$450USD in 2015. With this change the potential relationship of screen size and viewing
375 distance should be considered (Cooper, Piazza, & Banks, 2012; Tanton, 2004).

376 The industry standard recommendation is that viewing distance should be proportional to
377 screen size, i.e. that the screen should occupy a fixed angle in the field of view. Based on the
378 limitations above, The SMPTE guidelines justify the viewing angle of $\theta=30^\circ$ because viewing
379 from this angle “will result in a more immersive experience, and also lessen eye strain caused
380 by watching a smaller image in a dark room.” (Rushing, 2004) THX requires at least a 26°

381 viewing angle for the back row of a cinema theatre, but this, as well as its recommendation
382 for general viewing of $\theta=36^\circ$ (THX recommended viewing distances) appear to be arbitrarily
383 chosen.

384 Some basic trigonometry (Fig 2.1) results in the following relationships between viewing
385 distance V , angle subtended horizontally by the screen θ , screen width S , and screen diagonal
386 D , considering that the aspect ratio of widescreen televisions are 16:9:

387
$$V = \frac{S}{2 \tan \frac{\theta}{2}} \quad [\text{Eq 2.1.}]$$

388
$$S = 2V \tan \frac{\theta}{2} \quad [\text{Eq 2.2.}]$$

389
$$S = \frac{16D}{\sqrt{337}} \quad [\text{Eq 2.3.}]$$

390 Thus:

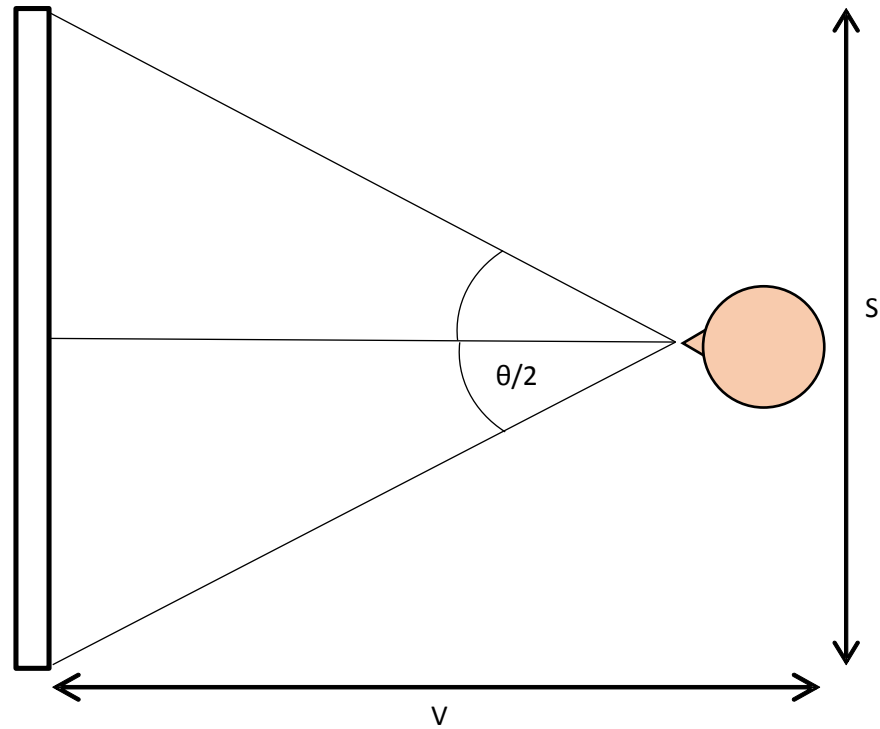
391
$$V = \frac{8D}{\sqrt{337} \tan \frac{\theta}{2}} \quad [\text{Eq 2.4.}]$$

392
$$D = \frac{V \sqrt{337} \tan \frac{\theta}{2}}{8} \quad [\text{Eq 2.5.}]$$

393 Different sources make different recommendations about what the angle θ should be. One
394 popular rule of thumb that doesn't consider viewing angle directly (Brady, 2009) is that
395 viewing distance should be equal to approximately 2.5 times the diagonal length of the
396 television screen,

397
$$V = 2.5 * D \quad [\text{Eq 2.6.}]$$

398 implying an angle $\theta=20^\circ$.



399

400 **Fig. 2.1.** *The relationship between viewing distance V and screen width S . The full visual angle*
 401 *subtended by the screen width is ϑ .*

402 Thus the three commonly encountered recommendations relating viewing distance (V) to
 403 diagonal screen size (D) are:

404 $D = 0.4 * V$ – rule of thumb (ROT) [Eq 2.7.]

405 $D = 0.615 * V$ – SMPTE [Eq 2.8.]

406 $D = 0.746 * V$ – THX [Eq 2.9.]

407 This implies that people should increase their viewing distance if they buy a larger television.

408 Lab studies indicate that people do indeed prefer to view larger images from further away

409 (Ardito, 1994; Cooper et al., 2012; Lund, 1993) and survey studies conducted by both Noland

410 and Truong in 2015 and Tanton in 2004 seems to echo this (Noland & Truong, 2015; Tanton,

411 2004). Ardito considered the relationship between viewing distance and picture height, as

412 reported subjectively by participants. He found that the viewing distance was preferred to

413 be between 4 and 5 times the height of the image. This is considerably larger than the

414 relationships now recommended, which could be a reflection of how much screen sizes have
415 grown since the study was conducted in 1994. Lund conducted a similar study, to establish a
416 relationship, and found similar results. Both results found relationships that were hyperbolic
417 and non-linear, which raises the question as to why the guidelines are still linear in structure.
418 It could possibly be in an attempt to make the advice more understandable for the average
419 consumer. Viewing distance is often constrained by the size of the living room, especially in
420 British houses which tend to be small compared to US ones (Footprint, 2013). Thus, it seems
421 unlikely that real viewers are following these recommendations.

422 This is particularly important for S3D. A distinctive feature of S3D is that viewing distance
423 potentially has a greater effect on viewer experience. In conventional 2D television, changes
424 in viewing distance effectively magnify or minify the image on the retina, without necessarily
425 implying anything about physical size. In S3D, changes in viewing distance can also alter the
426 perceived shape of objects (Welchman, Deubelius, Conrad, Bülthoff, & Kourtzi, 2005). Screen
427 size also has a more profound effect on S3D, not captured by the screen size angle θ . For
428 example, screen parallax exceeding the interocular separation is extremely uncomfortable
429 to view, as correctly verging becomes impossible, and hence the two retinal images cannot
430 be fused correctly (Mendiburu, 2009). S3D content created assuming one screen size which
431 is then viewed on a much larger screen could potentially exceed this limit, causing a lot of
432 discomfort to viewers. Thus, up-to-date information regarding both typical screen sizes and
433 typical viewing distances is essential to predict home viewer experience with S3D.

434 Accordingly, the aim of this chapter was twofold. First, to evaluate the amount of S3D
435 content viewing that people did, and any potential reasons for the amount of time spent
436 watching S3D. I compare true values of viewing distance and television size to see how they
437 compared to the commonly considered relationships above, and any potential reasons for
438 the relationship not being followed. This work has been published (Hands & Read, 2015).

439 2.2. Materials and methods

440 2.2.1. Participants

441 This study was approved by the Newcastle University Faculty of Medical Sciences
442 ethics committee. Participants were recruited on the basis they had at least one television in
443 their house. 559 people took part in the study, 452 of whom had data collected in person in
444 the North East of England. 107 of the responses were collected online via the free survey
445 website, SurveyMonkey, available at <https://www.surveymonkey.com/r/RZMNFHH>
446 (SurveyMonkey). Participants were given the opportunity to enter into a prize draw for a £10
447 gift voucher for completing the study.

448 2.2.2. Procedure

449 Participants were asked to complete a survey of 10 questions, based on their
450 general viewing habits and S3D consumption. These questions are shown below, with a brief
451 explanation, in table 2.1.

Question	Explanation
"How far away do you sit from your television?"	Participants were asked to give an approximation to the nearest 10cm as to how far away from the primary television they were sat the majority of the time.
"How large is your television?"	Participants were asked to give the diagonal size of their television in inches or cm. The answers were all converted into cm.
"Is there any reason you do not have a different sized television?"	Participants were asked to select from 5 options: Do not want another size; a different size would be too big or too small for the room; a different size would be too

	expensive to purchase; and the distance that I sit would have an adverse effect on picture quality if the size were different. An 'other' option was the fifth choice available but no participants wished to give a different reason.
"How many people do you typically watch television with?"	Participants were asked to give the number of people they typically watch television with (not including themselves, so zero was an option)
"Does your primary television have S3D technology?"	Participants were asked to confirm whether or not the primary television of the household had any S3D capacity.
"How much S3D content (including cinema) do you view per week?"	Participants were asked to quantify the number of hours of S3D content viewed on a weekly basis. They were given the following options: less than 1; 1 to less than 3; 3 to less than 5; 5 to less than 7; 7 to less than 9; or 9+ hours.
"On a scale of 1 to 7, where 1 is least and 7 is most, how much do you enjoy watching S3D content?"	Participants were asked to put a number to how much they enjoyed S3D content on a Likert scale of 1 to 7.

452 **Table 2.1.** Questions asked in the survey about viewing habits and an explanation of the
453 question.

454 Questions 8 and 9 were optional questions to find demographical information, asking for a
455 post code and age of the participants, and question 10 asked for an email address to use to
456 enter the participant into the prize draw.

457 The reason that the 7 questions were worded this way (those that were not demographical
458 or related to the prize draw) was to establish a large amount of information about viewing
459 habits without asking too many questions. The questions could be considered split into those
460 related to S3D viewing (questions 5, 6 and 7) and those related to general viewing habits
461 (questions 1 to 4).

462 Data was stored in Microsoft Excel and participant information was anonymized by removing
463 the email address from the results. Participants were given an identification number from 1
464 to 559 based on the order they completed the survey in. Analysis of the results was
465 conducted in Matlab and Microsoft Excel.

466 2.3. Results

467 One participant's data was deemed to be irregular enough that they were removed
468 from the study, based on the grounds that the responses were nonsensical in that the
469 participant reported sitting only 20cm away from the television screen, which was 150cm in
470 diagonal screen size. Results from the remaining 558 participants are presented below.

471 I considered separating the data into two subsets: those that filled in the survey online and
472 those that filled the survey in using pen and paper. The results were very similar in range and
473 mean of responses (mean of 236cm and 249cm viewing distance for online and paper
474 respectively, 97cm and 107cm screen size). Because of this I opted to combine the results
475 for analysis.

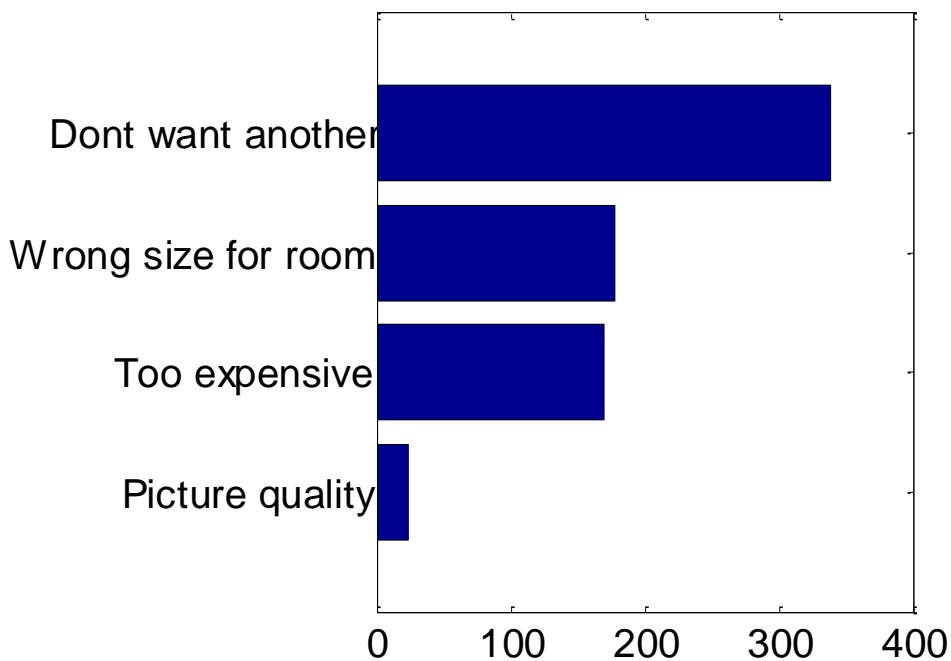
476 2.3.1 S3D viewing

477 All participants owned at least one television. Only 15.2% (85/558) owned an S3D
478 ready television.

479 Only 4 of the 85 participants who owned an S3D television reported watching S3D more than
480 1 hour a week (3 responses of '3-5 hrs' and 1 response of '1-3 hrs'). One additional participant
481 who did not own an S3D television reported viewing S3D '3-5 hrs' a week. Thus, of the entire
482 population sampled, only 0.9% (5/558) responded that they watched S3D more than 1 hour
483 a week. This suggests that despite S3D content being a lot more easily accessible in the home,
484 with the technology cheaper and allowing for a higher quality experience compared to past
485 surges in popularity of S3D content, people still do not very often watch S3D.

486 The mean score given to participants' enjoyment of S3D content was 4.13. There was no
487 significant difference between the 85 participants who reported having S3D capacity on their
488 primary television compared to the 473 who did not (two sample t-test, $p = 0.517$). The five
489 participants that said they watched more than 1 hour of S3D television a week reported
490 scores of 7,7,7,3 and 2 respectively. Thus, 60% of these 5 participants gave the highest
491 possible score for S3D enjoyment, compared to 17.9% (99/553) in the participants who
492 watched less than 1 hour a week.

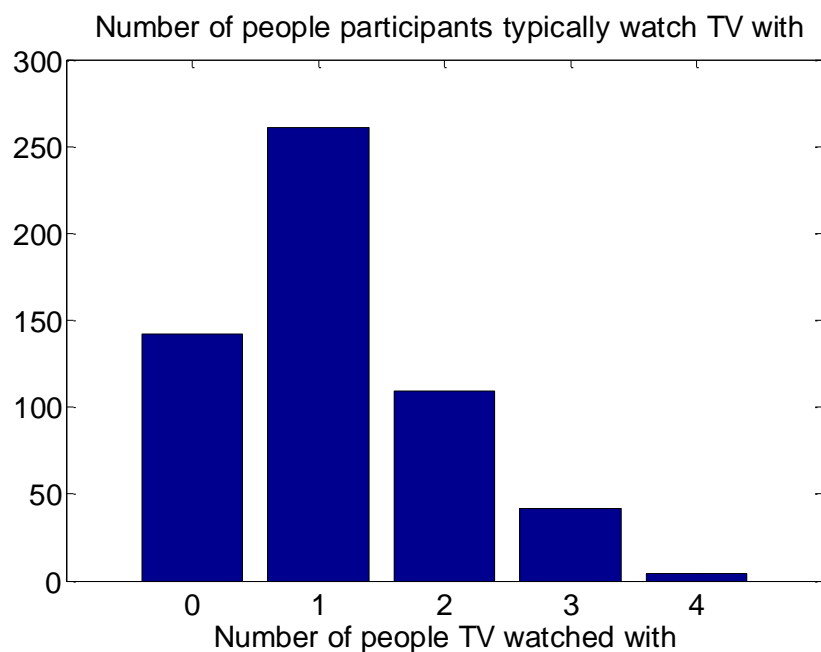
Reason for choosing TV size



493

494 **Fig. 2.2.** Histogram of responses to question number 3.

495 Question 3 asked people the reason for the size of their television. The most common
496 response was simply that they did not want another sized television (61%, 339/558), while
497 30% (170/558) answered that they could not afford a bigger one. Only 4% (24/558) agreed
498 that “the distance that I sit would have an adverse effect on picture quality if the size were
499 different,”, labelled above in Fig. 2.2 as ‘Picture quality’, but 32% (178/558) replied that a
500 bigger television would be too big for the room that it was in, potentially expressing an
501 awareness that a bigger television would need to be viewed from further away.



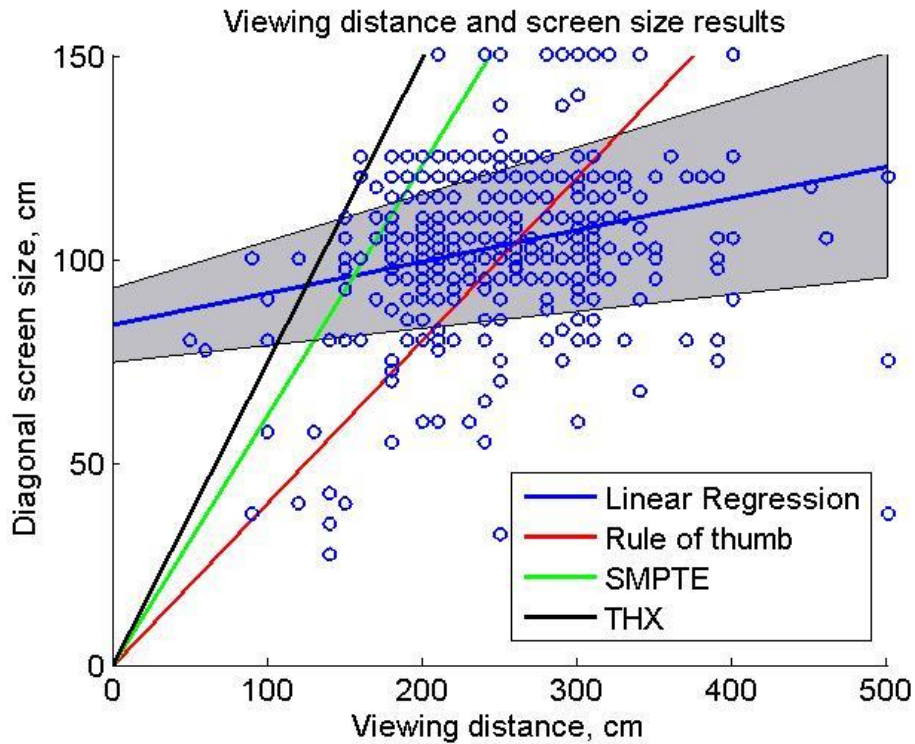
502

503 **Fig. 2.3.** Histogram to show the number of people participants typically watched television
504 with. As can be seen the most responses are from the lower numbers, with 1 other person
505 being the most common number to watch television with.

506 The answers to question 4 (Fig. 2.3.) show that typically people watch television with 1 other
507 person (48%, 261/558 participants) with 2 other people or watching by themselves also
508 popular (20% (109/558) and 25% (142/558) respectively).

509 2.3.2 Viewing distance and screen size

510 Fig. 2.4. shows the scatter of results from reported viewing distances and screen
511 sizes. I also show a linear regression generated from my data, and include the lines from the
512 3 recommended guidelines for comparison.



513

514

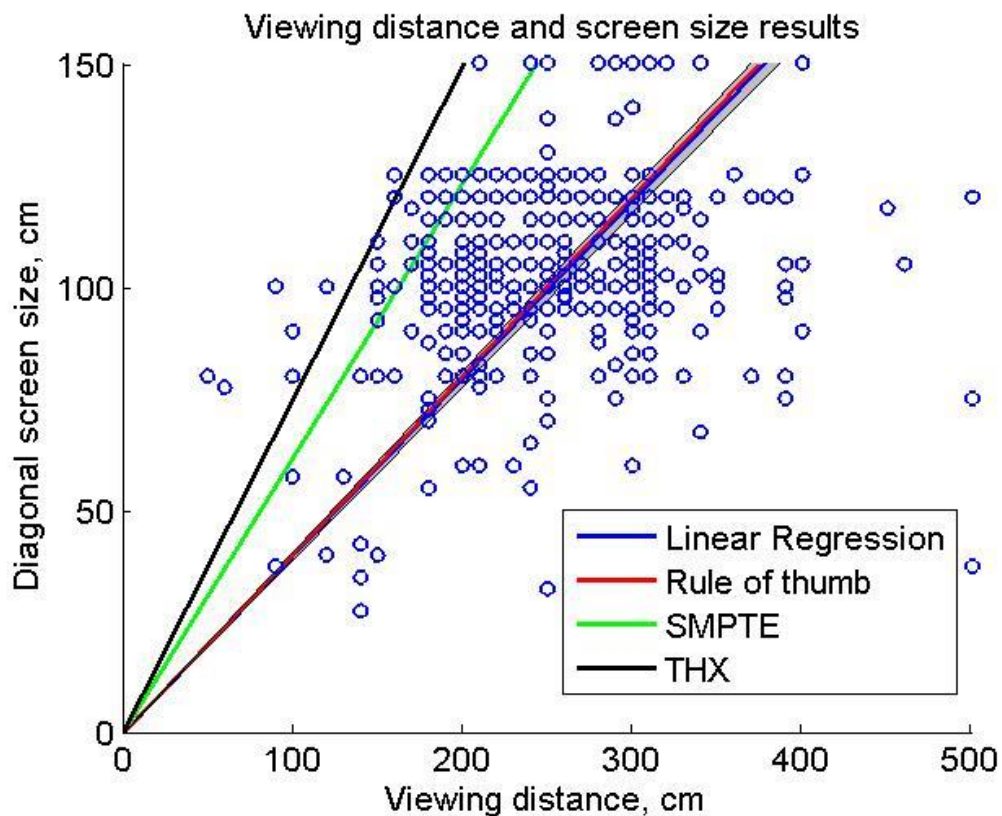
515 **Fig. 2.4.** Scatter graph of responses from $n = 558$ participants. The horizontal axis is the
516 reported distance from the television that participants sit, and the vertical axis is the reported
517 diagonal screen size of the television, both in cm. Also shown are the regression line of best
518 fit, and the three recommended relationships between viewing distance and screen size. The
519 shaded area is the 95% confidence interval of the regression fit.

520 The fitted regression line is:

521
$$D = 0.077 * V + 83.705 \quad [Eq 2.10.]$$

522 This curve has a significantly shallower slope than the 3 recommended relationships (the 95%
523 confidence interval for the slope were 0.038 to 0.114, much lower than the lowest
524 recommended slope of 0.4 from ROT) and also has an intercept, which none of the other

525 three guidelines have. It is possible that the guidelines don't have an intercept because the
526 guidelines would then be more complicated for consumers to implement themselves. Fig.
527 2.5. shows the same fitted regression forced to have no intercept. As can be seen the
528 regression line is very similar to the rule of thumb, and the confidence interval is considerably
529 smaller.

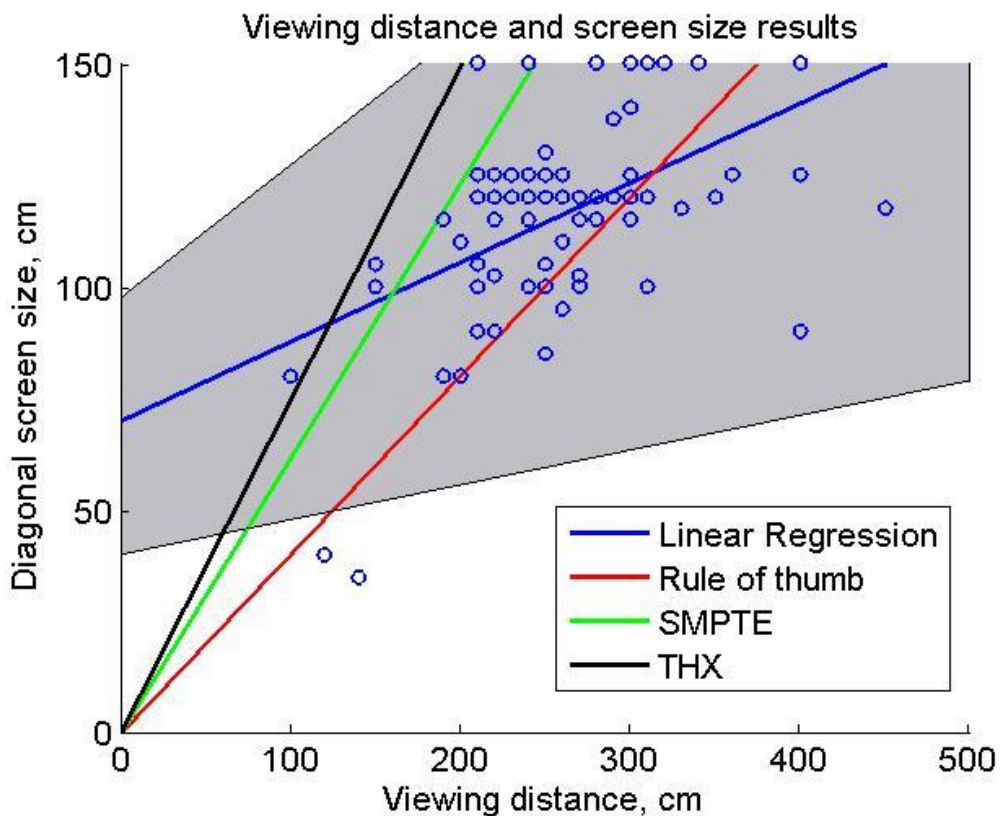


530

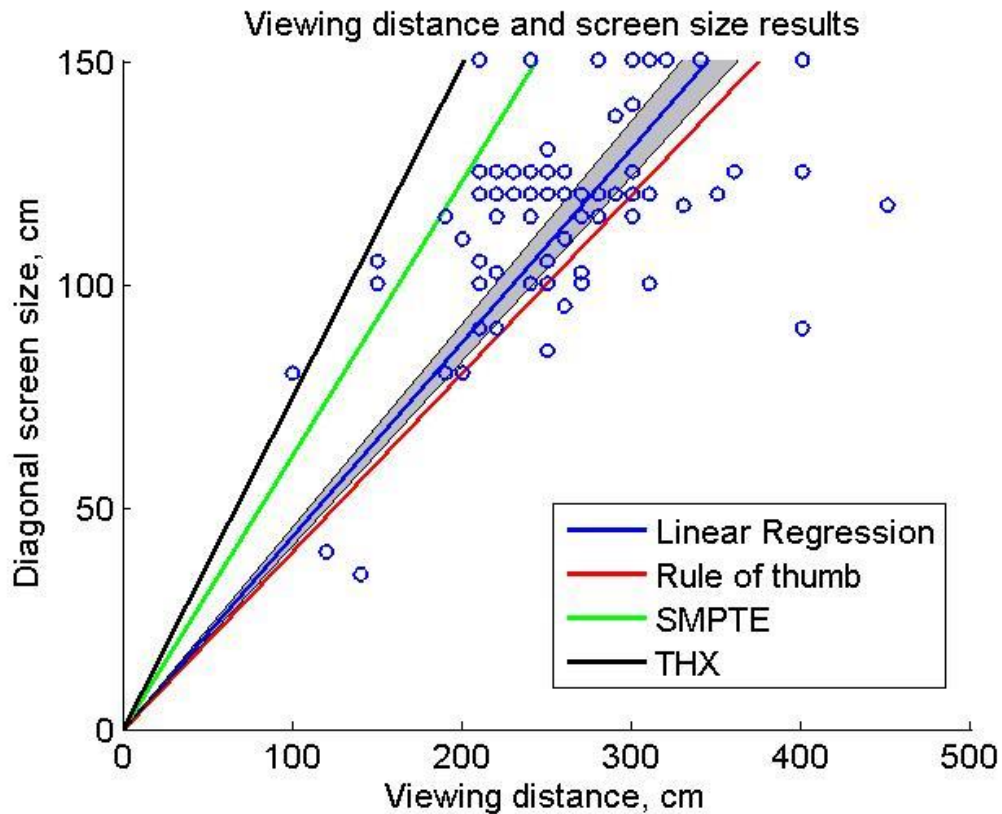
531 **Fig. 2.5.** Results plotted as for Fig. 2.4. but now the regression line is forced to have no
532 intercept. As can be seen the linear regression and the rule of thumb are very similar in slope.

533 Viewing distance was only weakly correlated with screen size ($R = 0.2483$, $P < 10^{-8}$, Pearson's
534 correlation coefficient), suggesting that the relationship between viewing distance and
535 screen size is not an important consideration when buying a television. This correlation
536 increased when only the subset of participants who owned an S3D television was considered
537 ($R = 0.4576$, $P < 10^{-4}$). This subset is shown below in Figs. 2.6. and 2.7. As can be seen the
538 subset has not got as many larger outliers, and hence the correlation is more positive. The
539 confidence intervals for the regression lines are larger, due to a smaller population. The

540 values selected suggest that those with S3D televisions purchase larger televisions and sit
 541 further away than those with televisions that do not have S3D capacity. One way this could
 542 be interpreted is that people who are more affluent can afford to purchase the larger, 3D
 543 capable televisions to put in their houses.
 544



545
 546 **Fig. 2.6.** Scatter graph of responses from $n = 86$ participants whom owned an S3D capable
 547 television. The horizontal axis is the reported distance from the television that participants
 548 sit, and the vertical axis is the reported diagonal screen size of the television, both in cm. Also
 549 shown are the regression line of best fit, and the three recommended relationships between
 550 viewing distance and screen size. The shaded area is the 95% confidence interval of the
 551 regression fit.



552

553 **Fig. 2.7.** Results plotted as for Fig. 2.6. but now the regression line is forced to have no
 554 intercept. As can be seen the linear regression and the rule of thumb are still very similar in
 555 slope.

556 There was a significant correlation between viewing distance and screen size. This further
 557 increased for the subset of participants in the study that had an S3D capable television. This
 558 suggests that viewers, whether consciously or subconsciously, consider the space available
 559 for them when they choose what size television to purchase, i.e. they either buy a television
 560 that is suitable for the size of the room, or possibly rearrange furniture to accommodate as
 561 larger a viewing distance as possible, or both. While the correlation values are only weak,
 562 there is clearly a relationship between screen size and viewing distance. If the intercept is
 563 forced to be zero then the rule of thumb appears to be the guideline most consistently
 564 followed by participants in this experiment. However the fit of the data is much better if a
 565 non-zero intercept is permitted, suggesting that in reality the increase of the screen size with
 566 viewing distance is much less steep than any of the guidelines specify. The limited range of

567 television sizes (mainly 80-120cm) and viewing distances (mainly 150-300cm) mean that
568 most viewers end up sitting at viewing distances broadly consistent with all the guidelines
569 (Figs. 2.4 and 2.5.). Indeed in feedback once the survey was completed participants regularly
570 stated they did not know the guidelines even existed. This could be a reflection on the fact
571 that it appears the recommendations made by THX and SMPTE are successful in capturing
572 people's instinctive preference about where is the best place to sit with regards to screen
573 size.

574 2.4. Discussion

575 Our study suggests that despite S3D technology in the home being a lot more
576 accessible than previously, most people in the North East of England are still not opting to
577 purchase an S3D ready television. My study also suggests that having a television with S3D
578 capacity is no indication of the preference for or enjoyment of S3D content. Even where
579 people possess an S3D-ready television, they still are not regularly watching S3D content.

580 One reason for not viewing more S3D in the home could be discomfort. Viewers do report
581 more adverse effects, such as headache, when viewing S3D TV, possibly due in part to
582 negative expectations of watching S3D content (Read, 2014; Read & Bohr, 2014). Potential
583 theoretical explanations of discomfort include the vergence/accommodation conflict
584 (Hoffman et al., 2008) and other, more subtle violations introduced in post-production
585 editing, such as hue or colour saturation being slightly different in each respective eye's
586 image, or timing issues. Another reason may be lack of available content, especially since the
587 UK's only S3D channel (on BSkyB) was withdrawn last year. Very little of the UK's most
588 popular television programming is available in S3D.

589 A further reason could in principle be that it is harder to watch S3D in groups. For an optimal
590 experience, the viewer must be seated at correct positions for both viewing distance and
591 viewing angle (Hands & Read, 2013), and the need to wear S3D technology glasses means
592 only a set number of people can watch the S3D content. However, in my survey most people

593 reported that they usually viewed television with just 1 or 2 other people. This suggests that
594 the number of glasses that come with S3D televisions should not be a limiting factor.
595 Additionally, it should not be hard for two or even three people to find a good viewing angle
596 to watch S3D together. The need to wear S3D glasses, however, may be off-putting in itself,
597 compared to sitting down to watch regular 2D television content. Further study is needed to
598 establish what factors limit the popularity of S3D content.

599 There is a possibility that participants didn't understand the wording of the adverse picture
600 quality response to question 3. I was trying to ascertain whether the participants understood
601 explicitly the relationship between viewing distance and screen size and the effect it can have
602 on the content viewed. However I was concerned about leading the participant to an answer,
603 or introducing any bias into my results, so left the wording as it was.

604 The primary consideration assigned to the recommended viewing distances appears to be
605 self-serving, and economic. Creators of televisions want consumers to purchase the largest
606 (and hence most expensive) televisions possible. Even removing economic pressures the
607 'best' viewing distance and screen size is not straightforward to calculate. Limitations in the
608 human visual system, the content displayed and the limitations of the technology are all
609 factors that could have an impact on the calculation.

610 An interesting observation from Ardito and Lund was that the relationship between screen
611 size and viewing distance was not a linear one, as I have modelled with regards to my data
612 and the guidelines recommended by THX, SMPTE and the rule of thumb. Indeed, Ardito
613 suggests the relationship is hyperbolic, while Lund simply states the relationship is decidedly
614 not linear. I wished to consider the guidelines in place already, and because they are linear I
615 opted to force my results to be linear also. There is an important distinction in the studies
616 also, as the experiments by Lund and Ardito were constructed in a lab and participants
617 changed the screen size until happy, compared to the method of survey to record already
618 established viewing habits. It would be interesting in a further study to consider if a more

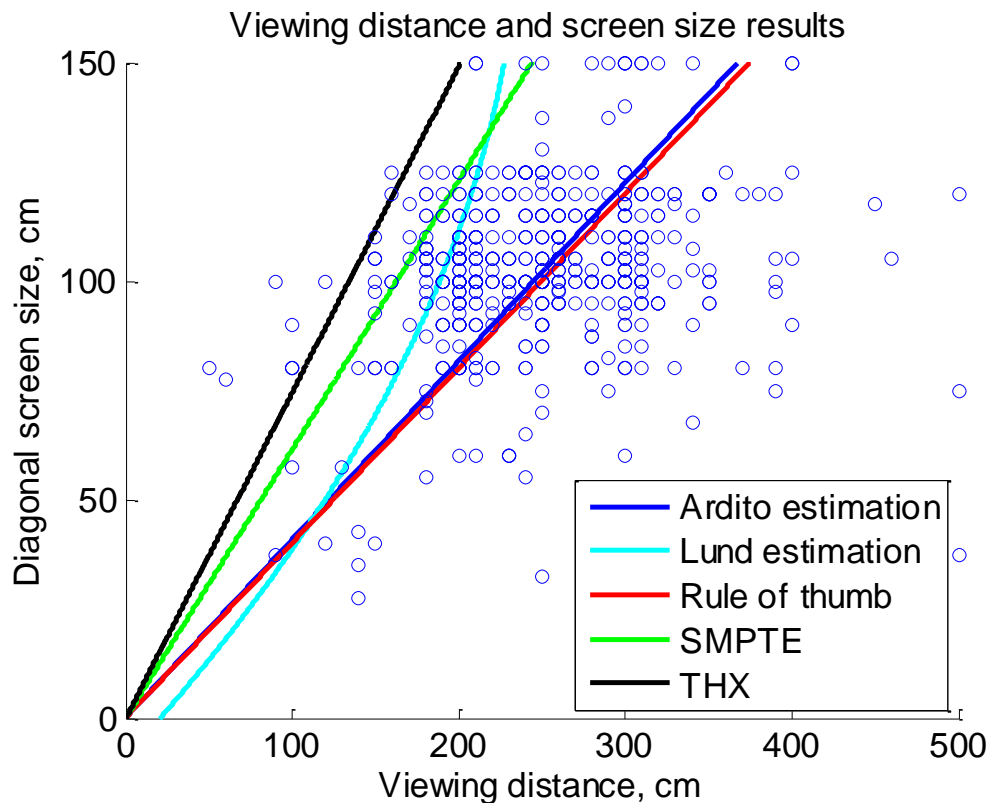
619 complicated model could better encapsulate the data, while still being accessible to
620 consumers. Fig. 2.8. shows my data along with the fits from both Ardito and Lund, instead of
621 from my own regression. The equation Ardito derived was

$$622 \quad D = 0.4079 * V, \quad [\text{Eq. 2.11.}]$$

623 which is strikingly similar to the rule of thumb (and my non intercept regression). This is a
624 guideline settled on despite stating that the equation wouldn't be linear. Lund's equation is
625 considerably more complicated, and relates the screen height (rather than diagonal screen
626 size) to viewing distance by

$$627 \quad V = 20.739 + 4.647 * H - 0.025 * H^2, \quad [\text{Eq. 2.12.}]$$

628 which is also shown on fig. 2.8. Both of these lines capture my data quite well. However I
629 believe that the Lund equation (Eq. 2.12.) would be too complicated to use with the general
630 public, and the Ardito equation was forced to not have an intercept. This resulted in a very
631 similar fitted line to my own regression, but I believe allowing for an intercept better
632 encapsulates all the data. It is worth noting that Lund considered televisions that were not
633 high definition, and estimated that for HDTVs the distance could be as much as 15% closer
634 than standard televisions.



635

636 **Fig. 2.8.** Datapoints from $N = 558$ participants along with the fitted equations from Ardito
 637 (1994) and Lund (1993), and the recommended guidelines. Ardito has a linear relationship
 638 very similar to that of the rule of thumb (blue and red lines respectively), while Lund's
 639 equation is hyperbolic (cyan line).

640 A further potential issue that this paper faces is that while the overall number of participants
 641 was high and there were many surveys collected, only a very small subsection ($n = 5$) watched
 642 S3D content regularly. This might question the confidence in the results based on this
 643 subsection. A potential way to avoid this would be to target S3D viewers specifically (e.g. to
 644 wait outside an S3D movie theatre to ask the questions). However this would put bias into
 645 the survey, and I wanted to assess the prevalence of S3D viewing in a general population. A
 646 potential other solution would be to increase the total survey size to try to get an increased
 647 population of S3D viewers. Unfortunately, if only 0.9% of the population regularly watch S3D
 648 content, to get a larger subset I would need a considerably larger population. As an

649 illustration to get 20 participants reporting regular S3D viewing I would need to question
650 over 2200 participants.

651 One final consideration for this study is that the data was provided in a subjective manner
652 by the participant, so is subject to potential error or misunderstanding of the questions. I
653 note that similar limitations apply to the study done by Noland and Truong (Noland & Truong,
654 2015) and Tanton (Tanton, 2004). This is an inevitable limitation of such surveys and is hard
655 to avoid without visiting people's homes to take accurate measurements, which would be
656 both arduous and costly.

657 2.5. Conclusion

658 I conducted a study to assess whether people had S3D capable televisions, whether they
659 liked S3D content and considered the relationship between the viewing distance and screen
660 size of televisions. I found that very few (0.9%) participants regularly watched S3D content,
661 and that participants generally did not like nor dislike S3D content (4.125 on a 7 point score
662 scale). There was also only a weak relationship between viewing distance and screen size,
663 contrary to advice from SMPTE and THX, with most people viewing their television from a
664 distance of 150-300cm. Most viewers watch television content with one other person, hence
665 the typical procedure of providing two pairs of S3D glasses with S3D capable televisions is
666 not a limiting factor. Despite improvements in technology and being available at a more
667 affordable price, S3D content still accounts for a very small proportion of television viewed
668 in the UK.

669 3. Stereoscopic 3D content appears relatively veridical when 670 viewed from an oblique angle.

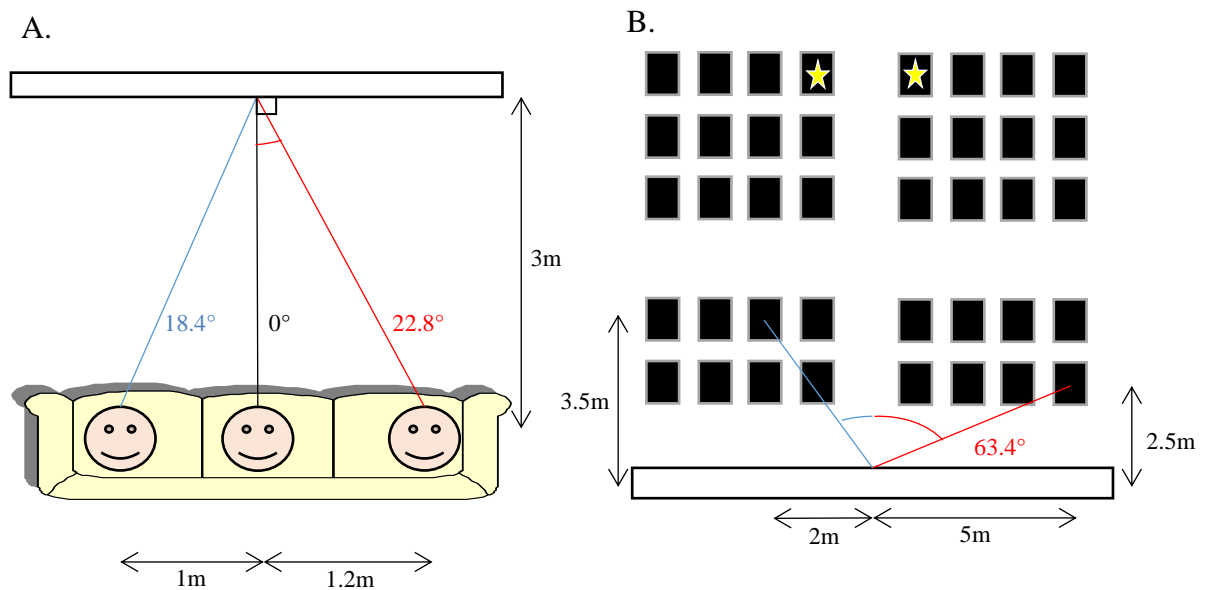
671 3.1 Introduction

672 S3D displays make it possible to recreate the different retinal images caused by a
673 real object in space. This exact recreation is often referred to as “orthostereoscopic” or
674 “orthostereo” (Kurtz, 1937). However, almost no commercial S3D content is
675 orthostereoscopic. To display in S3D orthostereoscopically, it is necessary to control and
676 coordinate all the aspects of the content production, from capture to display.
677 Mathematically, S3D displays produce an orthostereo image only when the viewer is
678 positioned with each eye exactly at the centre of projection for which that eye’s image was
679 filmed or rendered (Held & Banks, 2008; Woods, 1993). If the viewer moves away from this
680 specified position, the object depicted by the retinal stimulus will alter. Indeed, the retinal
681 disparities will in general be non-epipolar, i.e. not consistent with any physical object, given
682 the position of the eyes (Held & Banks, 2008; Read, Phillipson, & Glennerster, 2009; Woods,
683 1993).

684 One can distinguish two main ways in which viewers can move away from the centre of
685 projection. First, they may view content from the wrong distance. Second, they may view
686 content from the wrong angle. Previous studies have shown that incorrect viewing distance
687 can lead to distortions in perceived depth and shape (Held & Banks, 2008; Woods, 1993).
688 Woods discusses the different perceptual distortions that can occur based on camera
689 configuration, including depth non-linearity and size magnification. Woods shows that a
690 number of these distortions can be corrected for with a more precise camera configuration,
691 but that due to the limitations of S3D displays, some of these distortions are impossible to
692 correct for (Woods, 1993). Sometimes there is no correct viewing distance. Commercial S3D
693 content is often filmed with the cameras “toe-in”, i.e. converged on the object of interest.

694 This can produce a “keystone” distortion in the images. A keystone distortion causes objects
695 which should appear square to appear narrower at the right and wider at the left, and hence
696 appear as a keystone. To be orthostereo, such content either has to be corrected for the
697 distortion, or viewed on two screens: one for each eye, orthogonal to the line of sight from
698 the respective eye. While this can be arranged in a laboratory haploscope, it is almost never
699 the case for commercial S3D. If uncorrected content filmed with converged cameras is
700 presented on a single screen, the pattern of vertical disparities could only occur in reality if
701 the viewers’ eyes were more converged than is the case when they view the content (Banks,
702 Read, Allison, & Watt, 2012). Thus, there is no viewing position for which the content is
703 orthostereo. In any case, the correct viewing distance will typically vary during a feature. In
704 a mass-viewing venue like a cinema, viewing distance will vary greatly for different audience
705 members.

706 Viewing angle is more straightforward in that there is a clear “correct” viewing angle: almost
707 all S3D content is created to be viewed on a screen frontoparallel to the viewer. More
708 specifically, the eyes should be positioned such that the plane bisecting the interocular axis
709 should be normal to the screen and pass through the center of the screen. However, both in
710 cinemas and at home, many viewers will be viewing the screen obliquely. Even if they turn
711 their head towards the centre of the screen, such that the plane bisecting the interocular
712 axis passes through the screen midline, this plane will not be normal to the screen (Fig 3.1A).
713 Similarly in a cinema theatre, viewers seated at the extreme front and side of the auditorium
714 will be subject to a very large deviation away from the perpendicular viewpoint (Fig. 3.1B).
715 This is problematic since geometrically, the shape specified by a 3D display changes with the
716 viewing angle (Held & Banks, 2008; Woods, Docherty, & Koch, 1993). Thus if human depth
717 perception were based on the geometry of the retinal images, content created to be viewed
718 perpendicularly should look distorted from any other viewing angle.



719

720 **Fig. 3.1.** (A) A diagram to illustrate how viewing angles can change. If three people
 721 sit in the home watching television together, 3 metres away from the screen and between 1
 722 and 1.2 metres apart, the viewing angle can be up to 23°. Angles are measured from
 723 perpendicular viewing (here I show a top down diagram). (B) Viewing angles for different
 724 seats in a hypothetical cinema theatre auditorium. Stars indicate the seats with viewing
 725 angles closest to ideal (0° = frontoparallel).

726 Of course, these problems also apply to 2D images, in the sense that the image projected
 727 onto the retina varies as a function of viewing angle. The problem of why, nevertheless,
 728 images appear veridical from a range of viewing angles has fascinated researchers since the
 729 Renaissance (Kubovy, 1988; Pirenne, 1970). Several factors seem to contribute. One is that
 730 humans are simply not very sensitive to the distortion introduced by oblique viewing
 731 (Cutting, 1987; Gombrich, 1972). Additionally, images usually depict familiar objects, so the
 732 viewer's perceptions can be influenced by their expectations (Thouless, 1931). However, it
 733 is also clear that observers are capable of compensating for the oblique viewing, so that
 734 perception is based not on the image actually projected onto the retina, but on the image
 735 which would have been seen if viewed perpendicularly (Perkins, 1973; Rosinski, Mulholland,

736 Degelman, & Farber, 1980; Vishwanath, Girshick, & Banks, 2005). This compensation could
737 work by recovering the true centre of projection and reinterpreting the retinal image
738 accordingly. The true centre of projection could be estimated from cues present within the
739 depicted scene (De La Gournerie, 1859; Saunders & Backus, 2007), such as the location of
740 vanishing points, and/or from external cues regarding the orientation of the picture plane
741 combined with simplifying assumptions such that the true centre of projection lies on the
742 central surface normal. Presumably, such a mechanism would have to reflect experience
743 with 2D pictures (Deregowski, 1969; Jahoda & McGurk, 1974a, 1974b; Olson & Boswell,
744 1976). Vishwanath et al. (2005) have recently argued for a simpler heuristic, whereby the
745 retinal image is reinterpreted locally based on local surface slant. They argue that this may
746 reflect a more general heuristic which is useful when interacting with real objects viewed
747 obliquely, not a specific mechanism for interpreting pictures. External cues to local surface
748 slant include binocular disparity, vergence, accommodation, the position of specular
749 highlights relative to external light sources, and perspective cues provided by a frame
750 surrounding the screen plane. If you consider the perceptual mechanisms involved it would
751 make sense that the initial slant and orientation of the picture impact heavily on the
752 compensation mechanism and are a cue used. These lines are the initial deciphering of the
753 visual scene in the visual cortex of V1, and hence it would be consistent with the deeper
754 areas of the neural pathway deciding 'what' the image was showing after the angular
755 orientation of the viewing medium has been calculated and compensated for. Accordingly,
756 occluding the frame of the display, viewing monocularly or viewing through a pinhole all tend
757 to make the compensation less effective, so that images appear warped when viewed at
758 oblique angles (Bereby-Meyer, Leiser, & Meyer, 1999; Perkins, 1973; Vishwanath et al.,
759 2005).

760 There is a widespread belief that this compensation process is less effective for S3D stimuli
761 (Banks, Held, & Girshick, 2009; Bereby-Meyer et al., 1999; Perkins, 1973; Pirenne, 1970; Zorin

762 & Barr, 1995). There are several reasons why this should be so. In 2D displays, disparity and
763 vergence are powerful cues which specify that the picture lies on a flat plane, and also
764 indicate the orientation of this plane. Critically, these binocular cues are unaffected by the
765 contents of the picture, and therefore allow the viewer to estimate screen slant without
766 confounds. In S3D, both these cues now indicate that the scene is not planar but consists of
767 objects at different depths (Bereby-Meyer et al., 1999). In the words of Pirenne (1970), “in
768 the case of [stereoscopic images], the observer is hardly aware of the surface of the picture,
769 as a surface.” Ironically, therefore, the very thing that makes S3D a powerful visual
770 experience, namely the use of binocular disparity to depict 3D objects in space rather than
771 lying on a flat picture plane, might make viewers less able to correct for oblique viewing.
772 Additionally, despite recurrent upsurges of interest in S3D displays since the nineteenth
773 century, viewers will have had far less exposure to S3D pictures than to 2D. If experience
774 with 2D pictures plays a role in compensating for oblique viewing, these mechanisms may
775 not have developed to the same extent for S3D.

776 Surprisingly, however, this widespread belief has been little tested. I am aware of only three
777 previous studies, other than my own, which have considered perceptual distortions in
778 stereoscopic 3D due to oblique viewing (Banks et al., 2009; Bereby-Meyer et al., 1999;
779 Perkins, 1973). Banks et al. (2009) is the only study to compare perception of 2D and S3D
780 stimuli, although only one observer viewed both. They concluded that, as predicted, percepts
781 from stereo pictures are significantly more affected by oblique viewing angle than percepts
782 from conventional, 2D pictures.

783 All three previous studies used static content. This is a potentially serious limitation given
784 that commercial S3D usually consists of video content, which contains powerful internal
785 structure-from-motion cues. There are good theoretical reasons for expecting that these
786 cues could affect viewers’ ability to compensate for oblique viewing angle (Cutting, 1987).

787 The interpretations of 3D shapes based on motion are under-determined: the sequence of
788 images is consistent with many possible movements of objects in the world. Thus, humans
789 need to apply additional constraints, such as the rigidity assumption: “Any set of elements
790 undergoing a 2D transformation which has a unique interpretation as a rigid body moving in
791 space, should be interpreted as such” (Ullman, 1979). Humans are very good at
792 reconstructing this interpretation when they view a series of such 2D images. However, when
793 the same series of frames is viewed obliquely, the successive retinal images will not in
794 general be geometrically consistent with a rigid body in motion. Mathematically, this is the
795 same phenomenon discussed above in stereographic 3D: a stereogram designed to be
796 orthostereographic for frontoparallel viewing becomes non-epipolar – geometrically
797 inconsistent with any object – when viewed obliquely (Held & Banks, 2008). In stereo, the
798 visual system is capable of extracting the non-epipolar component of disparity and using it
799 to change the interpretation of the epipolar component, effectively interpreting the scene
800 as if it were being viewed with a different eye position (Mayhew & Longuet-Higgins, 1982;
801 Ogle, 1938; Rogers & Bradshaw, 1993). Conceivably, a related computation might be present
802 in the motion domain: the visual system might be able to use the rigidity assumption to
803 estimate the angle at which a projected image is being viewed from, as well as the shape of
804 the object and its motion relative to the eye. As I have shown, in picture perception, the brain
805 has to decide whether it is viewing a projection of Shape 1 from the correct angle, or a
806 projection of Shape 2 from an incorrect viewing angle. I have already shown some ways the
807 visual system might in principle choose between these, e.g. by using disparity cues from the
808 picture surface to deduce that the viewing angle is incorrect. However, with a dynamic
809 stimulus, the brain has to decide whether it is viewing a projection of a moving, deforming
810 Shape 1 from the correct angle, or a projection of a moving, rigid Shape 2 from an incorrect
811 angle. An assumption that objects are generally rigid would tend to result in the latter choice.
812 Since the rigidity assumption would apply equally to 2D and S3D content, this would tend to

813 reduce the difference between S3D and 2D content otherwise expected from the disrupted
814 binocular cues in S3D.

815 In the present study, I address this question using a canonical form task in which subjects are
816 asked to report their perception of cubes rendered for perpendicular and oblique viewing.
817 Cubes are a familiar object which have been used in many previous studies of picture
818 perception (Perkins, 1971; Hagen & Elliott, 1976; Hagen, Elliott, & Jones, 1978; Cutting,
819 1987). In a previous study (Hands & Read, 2013), I used static wireframe cubes. These
820 displayed the well-known Necker illusion (Necker, 1832), i.e. they could be perceived in one
821 of two different orientations. Because my cubes were rendered using perspective projection
822 and presented fairly close to the observer, only one of the two orientations appeared as a
823 cube; the other appeared as a warped frustum. To avoid distortions caused by this effect, in
824 the present study I used solid cubes (Fig 3.2B) whose orientation was unambiguous. The
825 cubes were covered with a checkerboard pattern. Thus, the stimuli contained several cues
826 which could potentially be used to judge whether the objects were perfect cubes with
827 parallel equal-length sides and right-angle corners (e.g. perspective, shading, texture). I
828 examined the effect of three factors on perceptual compensation. To determine the effect
829 of the picture frame, I compared results when the edges of the screen were occluded versus
830 when they were visible. To examine motion, I interleaved static objects with objects depicted
831 as rigidly rotating. Using an S3D display, I interleaved monocular, binocular 2D, and
832 stereoscopic 3D stimuli to test whether the visual system is less able to compensate for
833 oblique viewing in S3D than in 2D.

834 3.2 Material and methods

835 3.2.1 Participants

836 Participants were recruited via an internal volunteer scheme at Newcastle University
837 Institute of Neuroscience, on the basis that they had no visual problems other than wearing

838 glasses or contact lenses. The work was approved by Newcastle University Faculty of Medical
839 Sciences Ethics Committee. Ten participants (9 female, 8 naïve; 1 male, PH) were used in the
840 study. Only one voluntary participant and the author took part in both the previous study
841 (Hands & Read, 2013) and the current one due to availability of the other participants. Naïve
842 participants were not informed of the experimental aims or hypotheses, but due to the
843 random order of blocks will have been able to work out that the viewing angle was changing.
844 Participants were paid £10 for completing the study.

845 3.2.2 Apparatus

846 Stimuli were presented on a 50inch stereoscopic 3D monitor (LG 47LD920-ZA,
847 www.lg.com) using passive stereo technology. The resolution of the monitor was 1920 pixels
848 wide x 1080 high, and left/right eye images are presented on alternate pixel rows, so that
849 each image has a vertical resolution of 540 pixels. As described below, the monitor was used
850 in 2D mode to avoid artefacts due to the vertical averaging performed by the monitor in 3D
851 mode. The maximum luminance of the display was 20 cd/m² as measured through the 3D
852 glasses with a Minolta LS100 photometer. Interocular crosstalk was 1.4% when measured
853 with the screen frontoparallel to the photometer, rising to 2.0% for a viewing angle of 20°
854 and 7.1% for a viewing angle of 45°.

855 Participants sat at a viewing distance of 120cm, measured perpendicularly from the center
856 of the screen to the midpoint of the eyes, with their eyes at the same height as the center of
857 the screen. They wore passive 3D glasses throughout the experiment, enabling us to
858 interleave 3D, 2D and monocular stimuli. The monitor sat on a turntable, which allowed it
859 to be accurately rotated between ±45° about a vertical axis passing through the midline of
860 the screen. I define the viewing angle, θ_{view} , to be the angle between the plane normal to the
861 screen and the viewer's line of sight to the center of the screen (Fig 3.2A). In different
862 experimental blocks, the turntable was rotated so that angle θ_{view} was either 0°, -45° (closer
863 to the viewer on their right) or +20°. It was convenient to alter the viewing angle by moving

864 the display screen rather than the participant (see Fig 3.2A). A chinrest was used to ensure
865 the subject's eyes were at the correct position, and the chair was adjustable to ensure the
866 participant was comfortable. In some experimental blocks, a fabric curtain with a hole was
867 pulled across which occluded all four screen edges from the participant's view, while allowing
868 them to see the stimuli.

869 3.2.3 Stimulus generation

870 Stimuli were generated and the experiments run using the computer programming
871 environment Matlab (The Mathworks, www.mathworks.com) and the Psychtoolbox
872 extension (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997;). For each frame of
873 the stimulus, I generated separate left and right images of resolution 1920×540 pixels,
874 treating each pixel as being effectively a rectangle twice as high as broad (e.g. a frame 100
875 pixels wide by 50 pixels high would appear square on the screen). I used the interleaved line
876 stereomode of Psychtoolbox to combine these images on alternate pixel rows, and displayed
877 the result as a single image with the monitor in 2D mode.

878 In all my experiments, virtual cubes were rendered onto the screen via central perspective
879 projection. The center of each cube lay in the screen plane. Usually when one renders a
880 scene, the projection plane is perpendicular to the line from the center of projection to the
881 center of the scene. In my experiments, the projection plane was sometimes rotated away
882 from this position (Fig 3.2A). To find where to render a point on this rotated projection plane,
883 I imagine drawing a straight line from the center of projection through the point in question.
884 The point where this line intersects the projection plane is where the point should be
885 rendered. For a monocular viewer, whose eye is a pinhole at the center of projection, this
886 should produce exactly the same retinal image as the real object.

887 In a previous study (Hands & Read, 2013) using wire-frame cubes, I wrote my own Matlab
888 software to calculate where to render the vertices of each cube. I checked my calculations
889 by drawing a square onto a sheet of acetate and mounted it on a sheet of Perspex in front of

890 the screen, representing one face of the virtual cube. I supplied my code with the physical
891 coordinates of this square, and rendered it for different viewing angles. I verified that, in
892 each case, the image drawn on the screen lined up with the physical square drawn on the
893 acetate, confirming that my code was rendering the virtual code correctly, whether the
894 screen was perpendicular to the viewer or viewed obliquely at the specified angle. In the
895 experiments reported here, I used Psychtoolbox with the OpenGL library to draw solid,
896 textured cubes. I confirmed that this produced the same vertex positions by using
897 Psychtoolbox to draw dots on top of the rendered cubes at the locations of the vertices as
898 calculated by my own code, and checking that these dots lay on the vertices of the rendered
899 cubes.

900 Fig 3.3A shows the same wire frame cube rendered for render angles of 0° (red) and 45°
901 (blue). In the S3D condition, stimuli were rendered separately for left and right eyes. In the
902 M2D (monocular) condition, one eye saw the same stimulus as in the S3D condition while
903 the other eye saw a black screen (except for any crosstalk). In the B2D (binocular) condition,
904 the stimulus was rendered as if for a single cyclopean eye in the middle of the two actual
905 eyes. I used a standard value for interocular distance of 6.3cm, close to the average for adult
906 humans (Dodgson, 2004). Commercial S3D content is necessarily generated for a standard
907 viewer, and I was interested in measuring the effect of oblique viewing under these
908 conditions. Additionally, my data indicates viewers are insensitive to large errors in the angle
909 at which they view the screen ($>10^\circ$), so it seems unlikely they were very sensitive to errors
910 caused by the small variation in interocular distance.

911 3.2.4 Experimental design

912 In each trial, the participant viewed two cube-like objects, one rendered into the top
913 half of the screen and one onto the bottom. The participant was asked to choose which cube
914 looked the “most cube-like” in the sense of having equal length sides and all right-angle
915 vertices. They indicated their answer by pressing the up or down arrow on the keyboard.

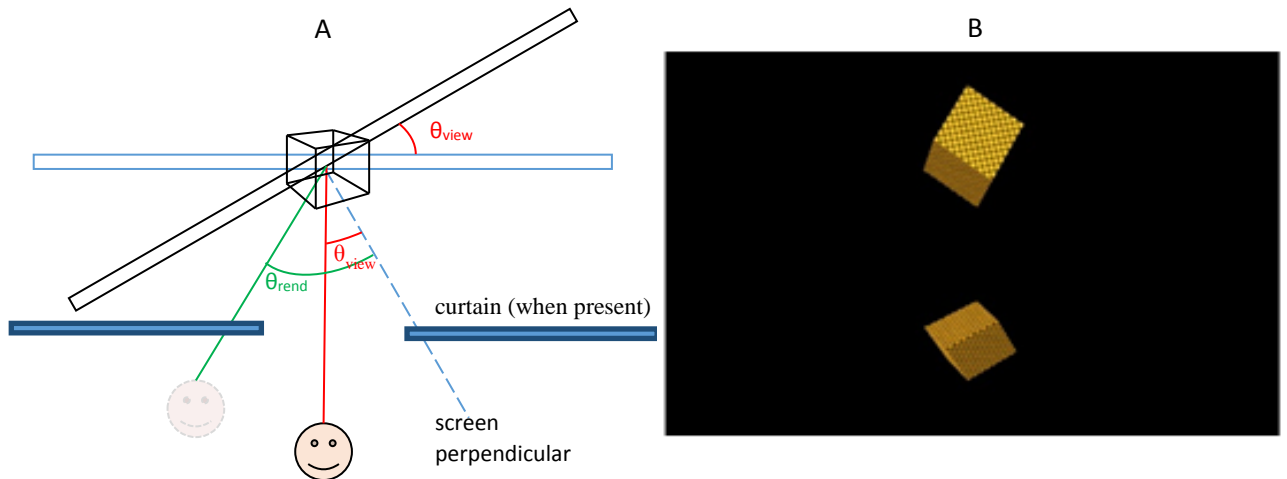
916 The objects were perspective projections of virtual cubes in space. The center of the virtual
917 cube was always in the screen plane, one-quarter screen-height either above or below the
918 center of the screen. This was unaffected by the screen orientation, since the midline of the
919 screen was the axis of rotation.

920 In each trial, one of the two cubes was rendered for frontoparallel viewing in the normal
921 way, i.e. for a line of sight perpendicular to the screen. The other was rendered for an oblique
922 viewing angle that varied between $\theta_{\text{rend}}=-45^\circ$ and $\theta_{\text{rend}}=+45^\circ$. I will refer to these as the
923 “normal-rendered” and “obliquely-rendered” cube, respectively. When $\theta_{\text{rend}}=\theta_{\text{view}}$, the
924 obliquely-rendered cube was rendered for the actual viewing angle of the participant. I will
925 refer to this as “geometrically correct”. In the S3D condition, the geometrically-correct
926 stimulus is orthostereoscopic, i.e. each eye ideally saw the retinal image which would have
927 been projected by a physical cube in front of the viewer, apart from accommodation effects.
928 On each trial, the orientation of each cube was random: each virtual cube was rotated
929 through a random angle about all three axes in succession before being rendered.

930 Fig 3.3B shows a cube rendered for 5 different values of θ_{rend} . The apparent distortion
931 increases monotonically as the rendering angle departs from frontoparallel. Additionally, a
932 given cube has a wider horizontal extent on the screen when rendered for oblique viewing
933 (Fig 3.3). To help ensure that participants did not simply judge the “more cube-like” object
934 to be the one with the smallest extent on-screen, the size of the virtual cubes was chosen
935 randomly on each trial. The side-length of one cube, L , was picked from a uniform
936 distribution between 6cm and 14cm, and the side-length of the other cube was then set to
937 $20\text{cm}-L$. The sum of the two side-lengths was therefore always 20cm, ensuring that the
938 rendered cubes never overlapped on the screen. This manipulation meant that the obliquely-
939 rendered cube could be either larger or smaller than the normal-rendered one.

940 In static trials both objects remained stationary on the screen; in motion trials, both
941 objects rotated at a constant speed of 18 deg/s about all three axes. This rotation speed was

942 chosen as being slow enough to be comfortable for the participant to follow, while fast
 943 enough to produce rapid changes in the on-screen image and thus powerful structure-from-
 944 motion cues (Ullman, 1979). In both types of trials, the objects remained on screen until the
 945 participant indicated whether the top or bottom object appeared most cube-like.
 946



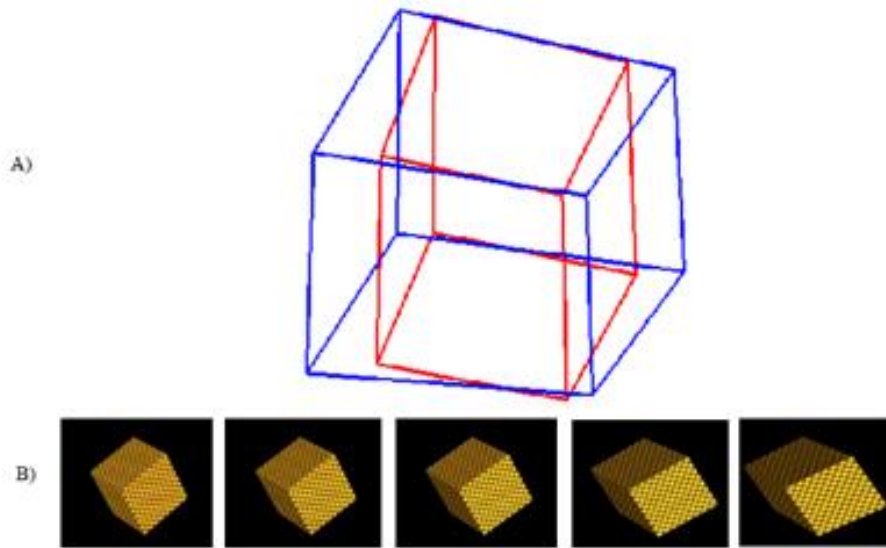
947
 948 **Fig. 3.2.** Experimental setup seen from above (A) and example stimulus (B). A: The
 949 heads of two possible viewers are sketched. The head drawn with solid lines represents the
 950 actual position of the participant, whose line of sight to the screen (red line) is at angle θ_{view}
 951 to the perpendicular (blue dashed line). The head drawn with dotted lines represents the
 952 position of a hypothetical viewer whose line of sight is at θ_{rend} (green line). The obliquely-
 953 rendered cube is projected correctly for this hypothetical viewer. It is correct for the
 954 participant only when $\theta_{\text{view}} = \theta_{\text{rend}}$. The normal-rendered cube is projected correctly for a
 955 second hypothetical viewer whose line of sight is perpendicular to the screen. In some blocks,
 956 a curtain was pulled across so as to occlude the edges of the screen from the participant's
 957 view. B: Stimulus drawn on the display screen. Here, the top cube is the "normal-rendered
 958 cube" (rendered for viewing perpendicular to the screen) and the bottom cube is the
 959 "obliquely-rendered cube", here rendered for a viewing angle of 45° . The reader may find that
 960 the bottom cube appears less distorted when viewed from 45° to the right.

961 3.2.5 Experimental parameters

962 The experiment was composed of six blocks. In each block the participant sat at one
963 of three viewing angles, $\theta_{\text{view}} = -45^\circ, 20^\circ$ or 0° , and had the curtain occluder either present
964 or absent. In blocks where the occluder was present, it was always pulled across before the
965 television's orientation was changed, so the participant had no prior knowledge of the screen
966 orientation. Each participant did the six blocks in a random order chosen with a random
967 number generator. In each block, the following 4 parameters were manipulated:

- 968 1. The angle θ_{rend} used to project the obliquely-rendered cube (8 possible values: $\pm 45^\circ, \pm 35^\circ,$
969 $\pm 20^\circ$ and $\pm 10^\circ$; see Fig 3.3B)
- 970 2. Whether the normal-rendered cube was at top or bottom of the screen (2 possible values)
- 971 3. Object motion (2 possible values: static or rotating)
- 972 4. Binocularity (4 possible values: S3D (binocular; each eye sees a different image), B2D
973 (binocular; each eye sees the same image on the screen) or M2D (monocular, left or right)).

974 For each combination of the first three parameters, the S3D and 2D conditions were
975 presented four times in each block, while the monocular-left and monocular-right trials were
976 presented twice. Thus, each block contained $8 \times 2 \times 2 \times (4+4+2+2) = 384$ trials, in a random
977 order chosen by the computer. No difference in results was apparent between the left-
978 monocular and right-monocular trials, so these were pooled for analysis along with cube
979 location (top or bottom of the screen). Thus, each block effectively contained 8 repetitions
980 of each of 48 combinations of experimental parameters ($8 \theta_{\text{rend}} \times 3$ binocularity,
981 S3D/B2D/M2D $\times 2$ object motion, static/rotating). Altering the viewing and rendering angles
982 enables us to assess the effectiveness of perceptual compensation for oblique viewing.
983 Binocularity, object motion and frame occlusion are the three "viewing factors" whose effect
984 on compensation I wish to assess.



985

986 **Fig. 3.3.** (A) Rendered image on screen for two different render angles. The two cubes are
 987 projections from exactly the same virtual cube in space, but the red cube is rendered for
 988 perpendicular viewing while the blue cube is rendered for a viewing angle of 45°. If you view
 989 the image with one eye from 45° to the right, the blue cube should appear as the red cube
 990 does when viewed normally. (B) Example cube rendered for five different angles used in my
 991 experiments. From left to right $\theta_{\text{rend}}=0^\circ, 10^\circ, 20^\circ, 35^\circ, 45^\circ$.

992 **3.2.6 Modelling**

993 To explain my data, I developed a mathematical model which assumes that object
 994 appearance is influenced by two competing mechanisms. First, I postulate that objects
 995 appear more veridical (in this case, cube-like) when the image on the retina is consistent with
 996 a perspective projection of a real cube (geometrically correct). In my experiments, this is the
 997 case $\theta_{\text{rend}} = \theta_{\text{view}}$. However, both my data and the existing literature indicate a second
 998 mechanism: objects also appear more veridical when rendered for frontoparallel viewing,
 999 $\theta_{\text{rend}} = 0^\circ$, even if the screen is in fact viewed obliquely. I assume that the “perceived
 1000 veridicality” due to each mechanism declines according to a Gaussian function as the value
 1001 of θ_{rend} moves away from the optimum, and I further assume that the “perceived veridicality”

1002 of the object is simply the sum of contributions from each factor. Accordingly, I model the
 1003 “perceived veridicality”, V , of each object as

$$1004 \quad V = A \exp\left(-\frac{\theta_{rend}^2}{2s^2}\right) + B \exp\left(-\frac{[\theta_{rend}-\theta_{view}]^2}{2r^2}\right), \quad [\text{Eq3.1}]$$

1005 where the free parameters s and r determine each factor’s sensitivity to θ_{rend} , and A and B
 1006 determine the relative weight of each factor. A is the weight given to normal rendering, and
 1007 B the weight given to geometrical correctness. In my experiments, one of the cubes was
 1008 always rendered for perpendicular viewing, $\theta_{rend}=0^\circ$. The difference in “perceived
 1009 veridicality” between this normal-rendered cube and the obliquely-rendered cube is
 1010 therefore

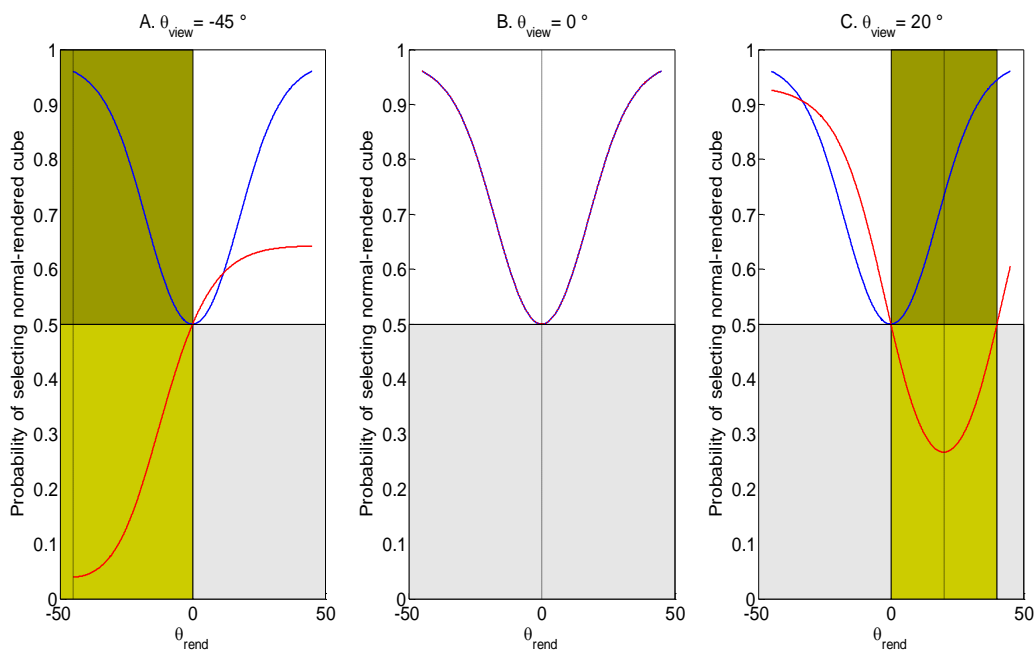
$$1011 \quad \Delta V = A - A \exp\left(-\frac{\theta_{rend}^2}{2s^2}\right) + B \exp\left(-\frac{\theta_{view}^2}{2r^2}\right) - B \exp\left(-\frac{[\theta_{rend}-\theta_{view}]^2}{2r^2}\right). \quad [\text{Eq3.2}]$$

1013 When this difference is positive, the viewer perceives the normal-rendered object as most
 1014 cube-like. To account for the graded chance in performance as a function of θ_{rend} and θ_{view} ,
 1015 as well as trial-to-trial variation, I make the usual assumption that this signal is subject to
 1016 internal noise, which I model as Gaussian. Without loss of generality, I set the standard
 1017 deviation of the noise to 1, since this degree of freedom is already accounted for by the
 1018 weights A and B . I assume that the viewer selects the normal-rendered object as most
 1019 resembling a cube whenever their noisy internal estimate of ΔV is greater than zero. The
 1020 probability that the viewer will select the normal-rendered object as most resembling a cube
 1021 is then given by

$$1022 \quad P = 0.5 + 0.5 \times \text{erf}\left(\frac{\Delta V}{2}\right). \quad [\text{Eq3.3}]$$

1023 At $\theta_{rend} = \theta_{view} = 0^\circ$, the model returns a probability of one-half for selecting either cube, which
 1024 is correct since at this point both cubes are rendered for the same viewing angle (they would
 1025 not be identical on the screen due to the randomization of size and orientation described
 1026 above).

1027 To illustrate the effect of the two mechanisms, Fig 3.4 shows model results for two different
 1028 extreme cases: perfect compensation (blue, $B = 0$) and no compensation (red, $A = 0$). With
 1029 perfect compensation, the results are unaffected by viewing angle: the model always selects
 1030 the normal-rendered cube when the obliquely-rendered cube is rendered with a perceptibly
 1031 different rendering angle. With no compensation, the model selects the obliquely-rendered
 1032 cube when this is closer to geometrically correct.



1033
 1034 **Fig. 3.4.** Model predictions with perfect compensation for oblique viewing (blue) and
 1035 no compensation (red). Curves show probability that the model selects the normal-rendered
 1036 cube as being more veridical, plotted as a function of render angle θ_{rend} for 3 different viewing
 1037 angles θ_{view} . Model parameters were $r=s=24^\circ$; for the blue curves, $A=3$ and $B=0$ (perfect
 1038 compensation for oblique viewing); for the red curves, $A=0$ and $B=3$ (no compensation).
 1039 White/gray regions show where the normal-rendered cube is also closer to geometrically
 1040 correct than the oblique cube ($|\theta_{\text{rend}}-\theta_{\text{view}}|>|\theta_{\text{view}}|$); yellow-shaded regions show where the
 1041 obliquely-rendered cube is closest to geometrically correct ($|\theta_{\text{rend}}-\theta_{\text{view}}|<|\theta_{\text{view}}|$). Light
 1042 shading (white or light yellow) is used to show the direction of preference expected under the
 1043 no-compensation model, i.e. below 0.5 where $|\theta_{\text{rend}}-\theta_{\text{view}}|<|\theta_{\text{view}}|$ and above 0.5 otherwise.

1044 3.2.7 Fitting

1045 Our high level model to fit a curve to the data assumes that the four model
1046 parameters A , B , r and s , do not change with viewing angle, θ_{view} . However, I allowed the
1047 model parameters to vary for the different viewing factors, i.e. frame occlusion, binocularity
1048 and object motion, to account for the effect these may have on perceptual compensation. I
1049 used maximum likelihood fitting assuming simple binomial statistics, as follows. Suppose that
1050 on the j^{th} set of stimulus parameters, my subjects chose the normal-rendered object on M_j
1051 out of N_j trials. Then the log-likelihood of the data-set is, apart from a constant which has no
1052 effect on the fitting,

$$1053 \log L = \sum_j \{M_j P_j + (N_j - M_j)(1 - P_j)\}, \quad [\text{Eq3.4}]$$

1054 where P_j is the model probability for the j^{th} data-point, which in turn depends on the stimulus
1055 parameters θ_{view} , θ_{rend} and the 4 model parameters as described by Eq2 and Eq3. I adjusted
1056 the model parameters so as to maximize this likelihood. The mathematical properties of the
1057 model meant that many different sets of model parameters gave virtually the same value for
1058 ΔV and were thus indistinguishable. To avoid this degeneracy, I set the value of the
1059 parameter A to 3 and allowed B to vary. I thus fitted sets of 3 model parameters (B, r, s) to
1060 sets of 24 data-points (8 values of θ_{rend} x 3 values of θ_{view}).

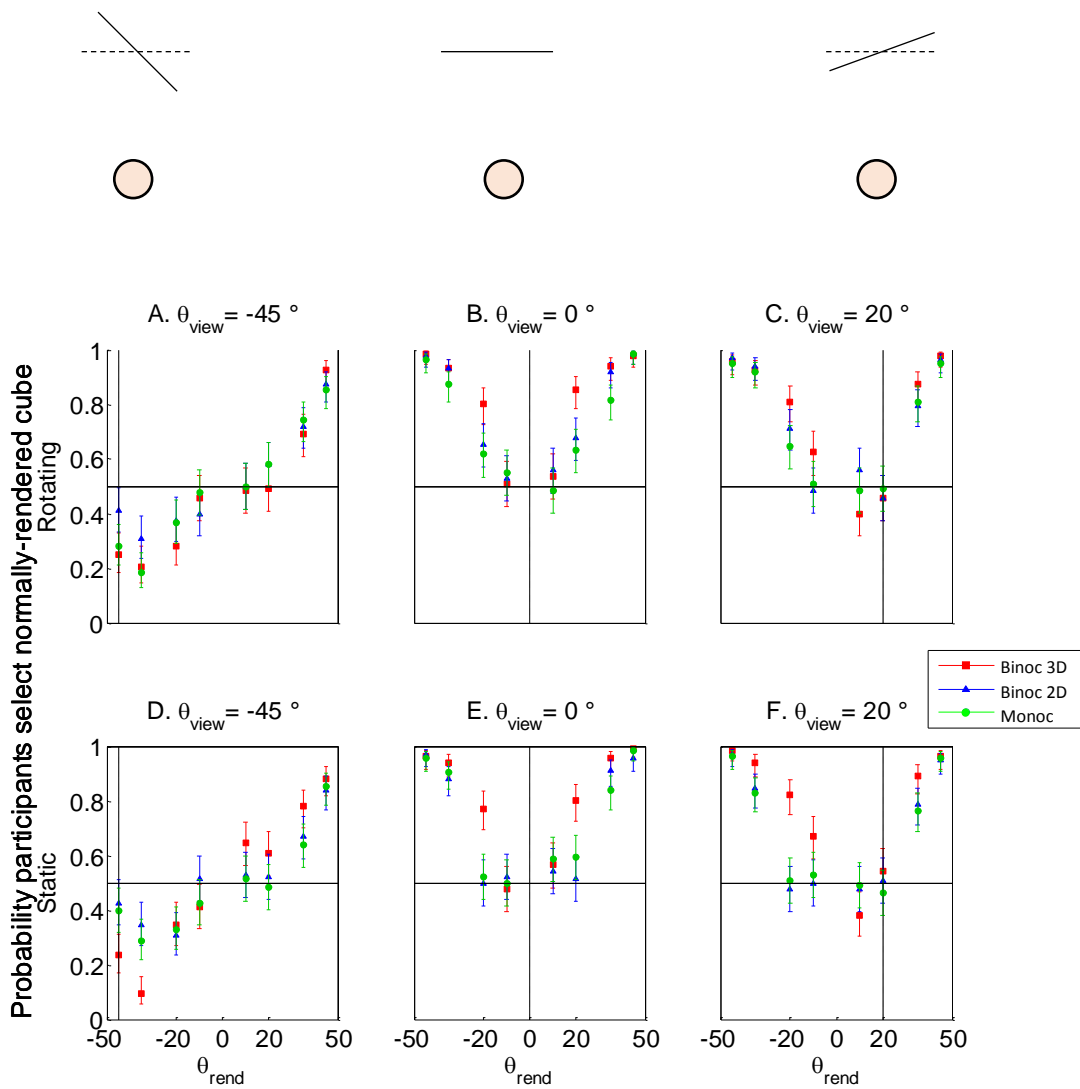
1061 3.3 Results

1062 Figs 3.5 and 3.6 show the proportion of trials on which the normal-rendered cube
1063 was selected as being “more cube-like”, pooled over all observers. I plot this as a function of
1064 θ_{rend} , the viewing angle for which the obliquely-rendered cube was drawn (Fig 3.2A). For
1065 $\theta_{\text{rend}}=0$, both cubes would be rendered for perpendicular viewing, so performance would
1066 necessarily be at chance. Figs 3.5 and 3.6 show results for the frame-visible and frame-
1067 occluded conditions, respectively. The three panels in each row of Figs 3.5 and 3.6 show
1068 results for the 3 different viewing angles, θ_{view} . The different colours and symbols show
1069 different binocularity conditions: red squares = binocular viewing in stereoscopic 3D; blue

1070 triangles = binocular viewing in 2D (same image on screen for both left and right eyes); green
 1071 disks = monocular viewing (pooled left and right monocular results). The upper panels (ABC)
 1072 show data for rotating stimuli and the lower (DEF) for static.

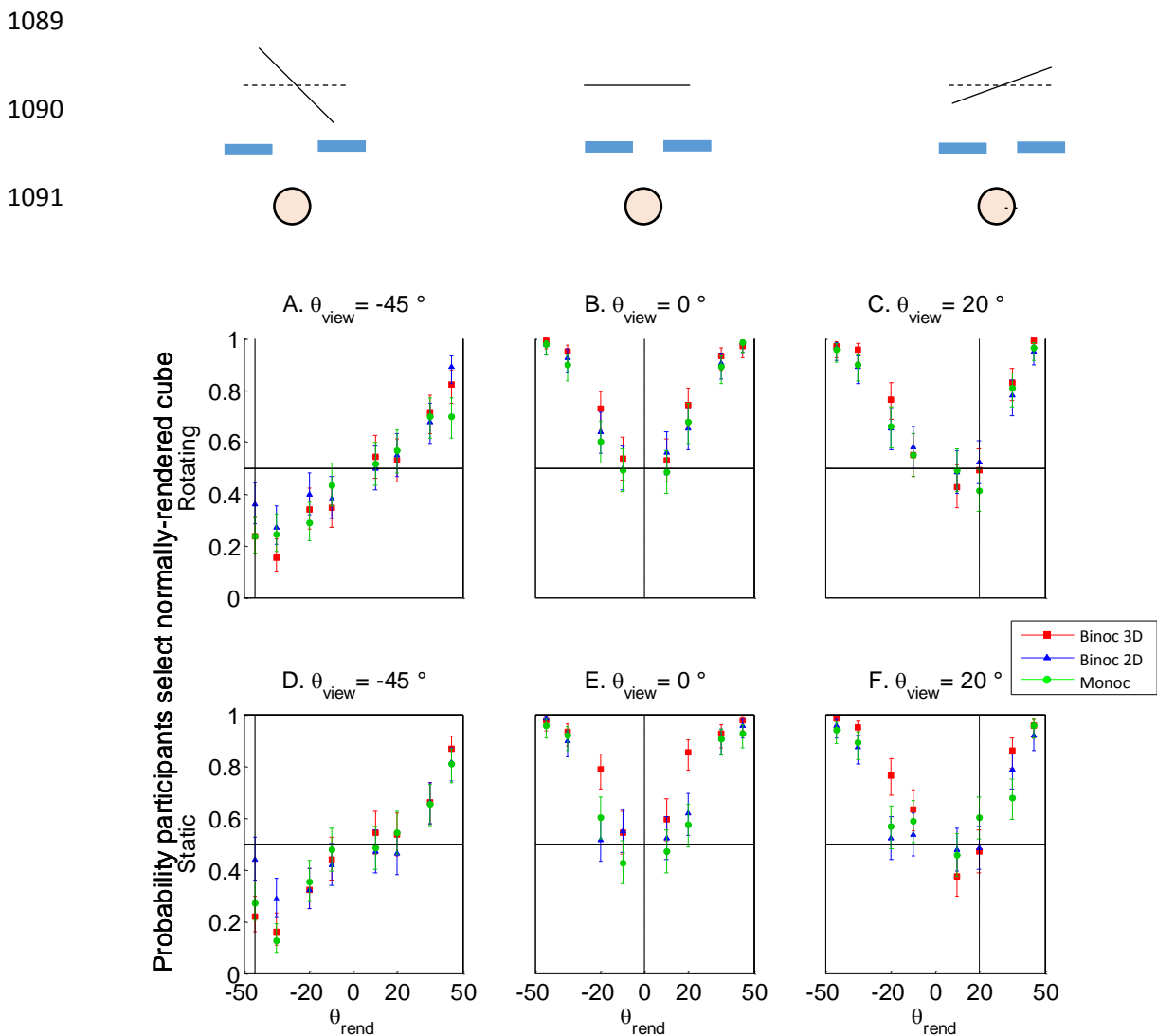
1073 Fig 3.5 shows results for the frame-visible condition, where subjects could see the television
 1074 screen and thus were aware when they were viewing it obliquely; Fig 3.6 shows results for
 1075 the frame-occluded condition, where the edges of the screen were not visible. Above each
 1076 figure a schematic is drawn of how the television was orientated and whether a curtain was
 1077 present or not, as an aid to the reader.

1078
 1079
 1080



1081

1082 **Fig. 3.5.** Data from the frame-visible condition. The vertical axis displays the
 1083 proportion of trials on which subjects reported the normal-rendered cube as appearing “more
 1084 cube-like”. Results are plotted as a function of θ_{rend} , the viewing angle for which the obliquely-
 1085 rendered cube was drawn. Subjects would therefore necessarily be at chance at $\theta_{\text{rend}}=0$. Data
 1086 was pooled across subjects; each data-point represents 176 trials from 10 subjects. Error-bars
 1087 show 95% confidence intervals using simple binomial statistics. The top row are for trials with
 1088 rotating cubes, the bottom row are for trials with static cubes.



1092

1093 **Fig. 3.6.** Frame-occluded data, presented as described in Fig. 3.5.

1094 Figs 3.7 and 3.8 are the same data as shown in Figs 3.5 and 3.6, however in these figures I
1095 use my model to fit the displayed curves to the data as described in the Methods. I discuss
1096 this in more detail below.

1097 The vertical dashed lines mark the case $\theta_{\text{rend}}=\theta_{\text{view}}$. In this case, for the S3D condition, the
1098 obliquely-rendered cube should project the same image onto each retina as a real cube
1099 (geometrically correct stimulus). The horizontal line at 0.5 marks chance (i.e. both cubes
1100 looked equally 'cube-like' to the participant and they selected one at random).

1101 Qualitatively, the results suggest that viewing angle has a direct influence on how veridical
1102 objects appear on a viewing medium. In panels A and D for both figs 3.5. and 3.6. as θ_{red} tends
1103 to the same angle of θ_{sit} (in this case, -45°) the probability of selecting the perpendicularly
1104 rendered cube decreases, suggesting a preference for geometrically correct rendering. This
1105 is somewhat echoed in panels C and F, although the decrease in probability is not as large.
1106 This could be a reflection of the smaller change in viewing angle from perpendicular (20°).
1107 The central panels (B and E) suggest that the introduction of S3D technology results in
1108 participants being more sensitive to rendering angle, as the red squares increase in
1109 probability of selecting the normally rendered cube the quickest. I now fit curves to our data
1110 using the high level model described in the methods in an attempt to discuss the results in a
1111 mathematical language.

1112 Figs 3.7. and 3.8. have the model fitted over the data. The model is fitted to all participants'
1113 results pooled together. If objects look veridical when rendered for normal viewing, even
1114 when viewed obliquely, data-points should lie above this line. If objects look veridical when
1115 they are geometrically correct on the retina, where data-points should lie depends on
1116 rendering and viewing angle. The white regions in each panel show where the normal-
1117 rendered cube is closer than the obliquely-rendered cube to being geometrically correct for
1118 the particular viewing angle. Here the normal-rendered cube should look more veridical, so

1119 subjects should select it whenever they can detect a difference between the two render
1120 angles (probability ≥ 0.5). The fact that data-points do lie in the white regions, rather than in
1121 the gray regions below them, confirms this, but does not enable us to distinguish between a
1122 preference for normal rendering and a preference for geometrical correctness.

1123 Conversely, the yellow regions show where the obliquely-rendered cube is closer to
1124 geometrically correct. The fact that data lies predominantly in the bright yellow regions
1125 below 0.5, rather than in the dark regions above 0.5, indicates that the preference for
1126 geometrical correctness usually wins out over that for normal rendering. However, the fact
1127 that datapoints never go as far below 0.5 as above it reveals that viewers are also affected
1128 by a preference for normal rendering. This agrees with previous work suggesting that, there
1129 are two factors which make a virtual object viewed on a screen appear “correct” to an
1130 observer: First, if it creates the same image on the retina as a real object would; but second,
1131 if the virtual object *would* create the same image on the retina as a real object *if* the observer
1132 were viewing the screen perpendicularly. In the next two sections, I discuss in more detail
1133 several aspects of my data which confirm this conclusion.

1134 3.3.1 Statistical analysis

1135 Figs 3.5 and 3.6 present data with different viewing factors, varying in frame visibility
1136 vs occlusion, binocularity, and object motion. I carried out several analyses to assess the
1137 effect of these different factors. First, I analysed the raw data (proportion of normal-
1138 rendered selections), which are independent of the assumptions made in my fitted model. I
1139 evaluated statistical significance using a generalised estimating equation in SPSS, using inter-
1140 subject and global comparisons of the raw data with edge occlusion, object motion,
1141 binocularity, angle of projection (θ_{rend}) and viewing angle (θ_{view}) as variables. We do this,
1142 rather than considering an ANOVA to determine the effects of variables and their
1143 interactions, as the data collected could not be assumed to be from a Gaussian distribution.
1144 The responses were either correct or incorrect, i.e. the distribution was binomial, not

1145 Gaussian, and the generalised estimating equation in SPSS can consider data from a binomial
1146 distribution. The 5 way interaction yielded significant results ($P < 0.0005$, Table 3.1), but this
1147 could be simply due to one specific set of factors yielding a significant result, rather than
1148 significance of the factors themselves. Thus I evaluate the main factors and the different
1149 possible interactions between the factors in Table 3.1. I discuss the nature and size of these
1150 differences in the following sections. The statistical significance of all main effects and
1151 interactions are reported in Table 3.1. I report χ^2 values with the degrees of freedom
1152 specified. If a factor or interaction is significant at the 0.05 level, the row in the table is
1153 highlighted green; if it is not, then the row is highlighted red.

1154 As can be seen from Table 3.1, all factors except for edge occlusion had a significant main
1155 effect on my results. This implies that, as one would expect, the perceived distortion of the
1156 cubes is affected by the angle at which they are viewed and the angle for which they are
1157 rendered, as well as by whether they are viewed in S3D, or binocularly or monocularly in
1158 2D. However, perhaps surprisingly, whether or not the edges of the TV screen are occluded
1159 with the curtain does not appear to be important. Most interactions, including all 4-way
1160 and 5-way interactions are also significant.

1161 In the above statistical analysis I considered all the data collected. This makes it difficult to
1162 assess the effect of different factors on the two different components identified in my
1163 model. As argued above, my data imply that two factors affect whether an object appears
1164 distorted: whether it is geometrically correct on the retina ($\theta_{\text{rend}} = \theta_{\text{view}}$), but also whether it
1165 would be correct if viewed perpendicularly ($\theta_{\text{rend}} = 0^\circ$). Much of my data confound these two
1166 effects, because often, both factors imply that the user should select the normal-rendered
1167 cube. This situation corresponds to the white regions in Figs 3.5 and 3.6. To assess how the
1168 different experimental conditions (occlusion, binocularity, rotation) affected the
1169 competition between the two model components, I also repeated this statistical analysis

1170 using only data where the two components pulled in opposite directions, i.e. the yellow
1171 regions in Figs 3.5 and 3.6. Here there is no overlap in the values of θ_{view} and θ_{rend} so the
1172 statistical significance of θ_{view} cannot be determined. I therefore only consider the main
1173 factor influences and the interactions between frame occlusion/visibility, binocularity,
1174 rotation and θ_{rend} . Table 3.2 shows the main effects and interaction terms for these 4
1175 factors.

1176 Within this more limited data-set, frame occlusion now has a highly significant main effect
1177 on the results, as well as the other factors which did so previously. Considering the 2, 3 and
1178 4-way interactions in Table 3.2, I see that all the interactions including θ_{rend} return significant
1179 results whereas any interactions not including θ_{rend} are not significant. This makes sense,
1180 because clearly the rendering angle θ_{rend} is key to whether the object appears distorted. All
1181 analysis up to this point is independent of my model. My statistical analysis implies that
1182 frame occlusion, binocularity and object motion all affect the balance between the
1183 competing preferences for the “geometrically correct” vs “normal” rendering angle.

Factor or interaction (*)	χ^2	DF	P
Occlusion	2.723	1	0.099
Binocularity	53.290	2	<0.0005
Motion	6.391	1	0.011
ϑ_{view}	105.692	2	<0.0005
ϑ_{rend}	4814.267	7	<0.0005
Occlusion*Binocularity	6.461	2	0.040
Occlusion*Motion	1.758	1	0.185
Occlusion* ϑ_{view}	50.411	2	<0.0005
Occlusion* ϑ_{rend}	4018.044	7	<0.0005
Binocularity*Motion	114.703	2	<0.0005
Binocularity* ϑ_{view}	1602.966	4	<0.0005
Binocularity* ϑ_{rend}	>10 ¹⁵	11	<0.0005
Motion* ϑ_{view}	10.122	2	0.006
Motion* ϑ_{rend}	1182.191	7	<0.0005
ϑ_{view} * ϑ_{rend}	>10 ¹⁴	10	<0.0005
Occlusion*Binocularity*Motion	0.624	2	0.732
Occlusion*Binocularity* ϑ_{view}	8.495	4	0.075
Occlusion*Binocularity* ϑ_{rend}	>10 ¹²	10	<0.0005
Occlusion*Motion* ϑ_{view}	0.564	2	0.754
Occlusion*Motion* ϑ_{rend}	3270.183	7	<0.0005
Occlusion* ϑ_{view} * ϑ_{rend}	>10 ¹⁰	9	<0.0005
Binocularity*Motion* ϑ_{view}	10.000	4	0.040
Binocularity*Motion* ϑ_{rend}	>10 ¹⁴	9	<0.0005

Binocularity* ϑ_{view} * ϑ_{rend}	166597.996	8	<0.0005
Motion* ϑ_{view} * ϑ_{rend}	>10 ¹²	9	<0.0005
Occlusion*Binocularity*Motion* ϑ_{view}	54.382	4	<0.0005
Occlusion*Binocularity*Motion* ϑ_{rend}	1768.396	8	<0.0005
Occlusion*Binocularity* ϑ_{view} * ϑ_{rend}	>10 ¹²	10	<0.0005
Occlusion*Motion* ϑ_{view} * ϑ_{rend}	>10 ¹²	9	<0.0005
Binocularity*Motion* ϑ_{view} * ϑ_{rend}	>10 ¹²	11	<0.0005
Occlusion*Binocularity*Motion* ϑ_{view} * ϑ_{rend}	>10 ¹⁴	12	<0.0005

1184 **Table 3.1.** How the individual factors affected my results (main effects) and the interactions
1185 between the factors, denoted by *. Results are from a generalized estimating equation done
1186 in SPSS, and return the χ^2 value, along with the degrees of freedom and associated P-value.
1187 Significant effects are highlighted green, non-significant results in red.

Factor or interaction (*)	χ^2	DF	P
Occlusion	32.800	1	<0.0005
Binocularity	354.095	2	<0.0005
Motion	5.174	1	0.023
ϑ_{rend}	42080.094	6	<0.0005
Occlusion*Binocularity	4.183	2	0.123
Occlusion*Motion	0.042	1	0.837
Occlusion* ϑ_{rend}	35.415	6	<0.0005
Binocularity*Motion	1.921	2	0.383
Binocularity* ϑ_{rend}	816.705	8	<0.0005
Motion* ϑ_{rend}	155.478	6	<0.0005
Occlusion*Binocularity*Motion	1.718	2	0.424
Occlusion*Binocularity* ϑ_{rend}	>10 ¹⁴	9	<0.0005
Occlusion*Motion* ϑ_{rend}	12.864	6	0.045
Binocularity*Motion* ϑ_{rend}	228.869	8	<0.0005
Occlusion*Binocularity*Motion* ϑ_{rend}	>10 ¹⁴	10	<0.0005

1188 **Table 3.2.** How the individual factors and interactions (denoted by *) affected the results that
1189 had a different geometrically correct cube to the perpendicularly projected cube. Results are
1190 from a generalized estimating equation done in SPSS, and return the χ^2 value, along with the
1191 degrees of freedom and associated P-value. Significant effects are highlighted green, non-
1192 significant results in red.

1193 3.3.2 Sensitivity to rendering angle θ_{rend}

1194 I first consider the central panel, Fig 3.7B, where $\theta_{view}=0^\circ$, i.e. the screen was
1195 frontoparallel in the usual way. If $\theta_{rend}=0^\circ$, both cubes would have the same projection, so
1196 performance would be at chance. As the obliquely-rendered cube is drawn at ever more

1197 extreme angles, it appears progressively more distorted and subjects become more likely to
1198 choose the normal-rendered cube. The rendering angle θ_{rend} is significant when considering
1199 only this subset of the data ($\chi^2 = 42080.1, P < 0.0005$). In agreement with previous studies
1200 (Cutting, 1987), subjects are fairly insensitive to incorrect rendering. At $|\theta_{\text{rend}}|=10^\circ$, results
1201 do not differ significantly from chance for any binocularity conditions (95% confidence
1202 intervals in Fig 3.7 overlap chance). Even when θ_{view} is as large as 20° , the results are not
1203 significantly different from chance for a static cube viewed without S3D. For a rotating cube,
1204 or a static cube viewed in S3D, subjects are significantly more likely to choose the normal-
1205 rendered cube, but do so only about 75% of the time. Even when the obliquely-rendered
1206 cube is drawn for a viewing angle as extreme as 45° , subjects still choose it as being “more
1207 cube-like” on nearly 10% of trials when viewing a static cube in 2D. This is surprising, given
1208 that a rendering angle of $\theta_{\text{rend}}=45^\circ$ produces a very different image on the screen from one
1209 of 0° (Fig 3.3A).

1210 3.3.3 Effects of oblique viewing angle, $\theta_{\text{view}} \neq 0$

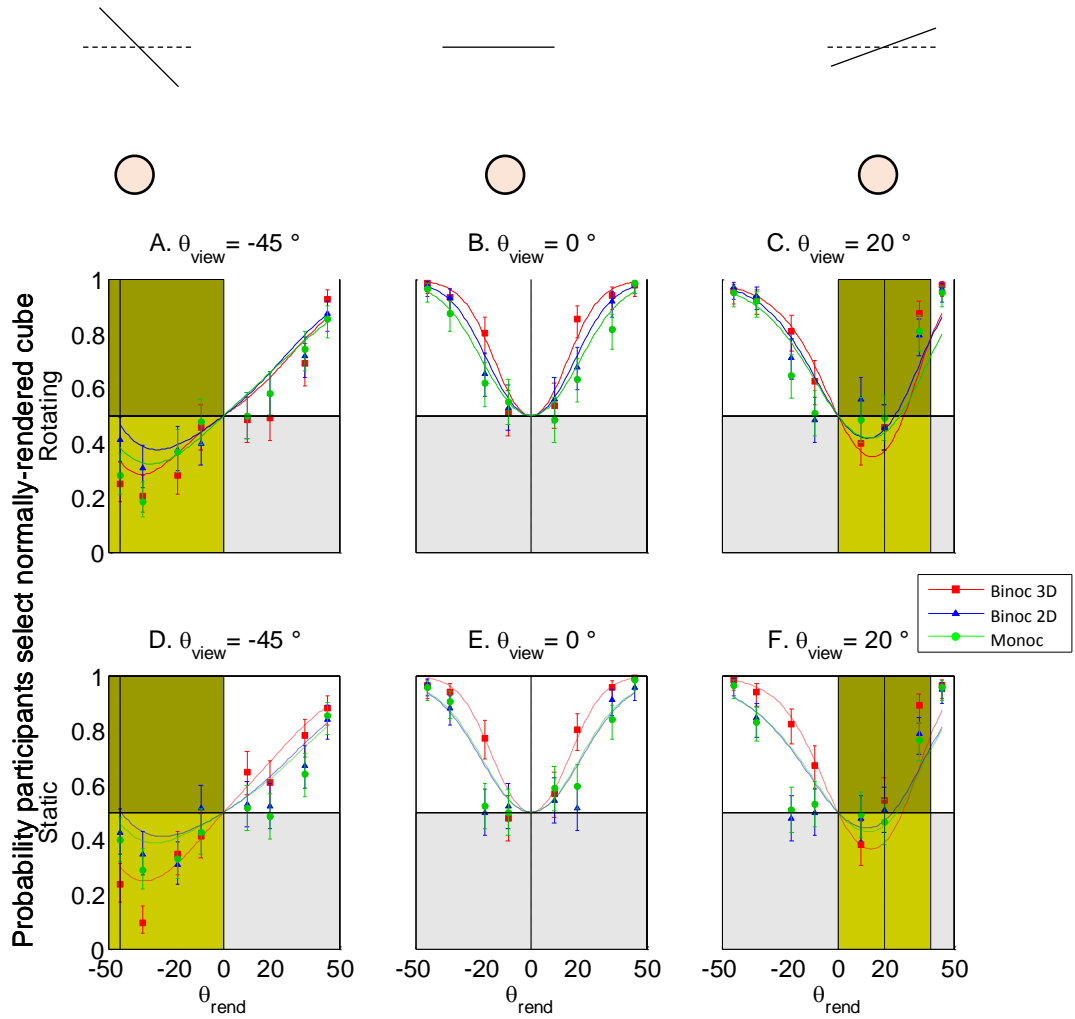
1211 Fig 3.7A and C show results where subjects were viewing the screen obliquely.
1212 Clearly, the results are very different. At almost every value of θ_{rend} , participants are less likely
1213 to select the normal-rendered cube than when the screen was frontoparallel to them. In the
1214 yellow-shaded regions, where a preference for normal rendering conflicts with a preference
1215 for geometrical correctness, data lies in the bright region below chance rather than the
1216 shaded region, i.e. participants were more likely to select the object which was closer to
1217 geometrically correct. This indicates that they were not able to compensate completely for
1218 the oblique viewing angle.

1219 However, oblique viewing clearly had a strong effect on perception, even when the retinal
1220 image had been designed to take oblique viewing into account. For example in Fig 3.7A, the
1221 viewing angle was $\theta_{\text{view}}=-45^\circ$. Thus at $\theta_{\text{rend}}=-45^\circ$, the obliquely-rendered cube produced the

1222 geometrically correct image of a cube on the retina, whereas the normal-rendered cube was
1223 distorted. Fig 3.5B shows subjects are quite capable of detecting a 45° error in rendering
1224 angle when the display is frontoparallel: they reject the erroneous rendering over 80% of the
1225 time. However, when viewing obliquely at $\theta_{\text{view}}=-45^\circ$ (Fig 3.5A), subjects did not show a
1226 comparably strong preference for the geometrically correct cube: they chose it only 25% of
1227 the time for the S3D stimulus at $\theta_{\text{rend}}=-45^\circ$, while for the 2D stimuli, they picked both cubes
1228 equally often. This cannot be explained simply by a lack of sensitivity to distortion (Cutting,
1229 1987; Gombrich, 1972), but must reflect a mechanism favoring normal rendering.

1230 A similar conclusion is indicated by the asymmetry about the line $\theta_{\text{rend}}=\theta_{\text{view}}$ in Fig 3.7C.
1231 Geometrically, the obliquely-rendered cube should appear equally distorted for viewing
1232 angle discrepancies of equal magnitude, $|\theta_{\text{view}}-\theta_{\text{rend}}|$. Thus it should appear more distorted
1233 for $\theta_{\text{rend}}=-10^\circ$ (a discrepancy of 30° with the true viewing angle, $\theta_{\text{view}}=20^\circ$) than for $\theta_{\text{rend}}=35^\circ$
1234 (a discrepancy of only 15°). Yet Fig 3.5C shows that in fact, for 2D stimuli, subjects cannot
1235 perceive the distortion at all for $\theta_{\text{rend}}=-10^\circ$ (they pick the obliquely-rendered cube as often as
1236 the normal-rendered cube), whereas it is fairly obvious to them at $\theta_{\text{rend}}=35^\circ$ (they pick the
1237 normal-rendered cube on 75% of trials). This asymmetry, along with the lack of a clear
1238 preference for the geometrically correct rendering, is another indication of a compensation
1239 mechanism which corrects for oblique viewing and makes objects rendered for normal,
1240 perpendicular viewing tend to appear “correct”, even if the retinal image is in fact distorted.
1241 However, this compensation works only up to a point. If the compensation were perfect,
1242 then Figs 3.7A and C would be identical to Fig 3.7B (cf Fig 3.4).

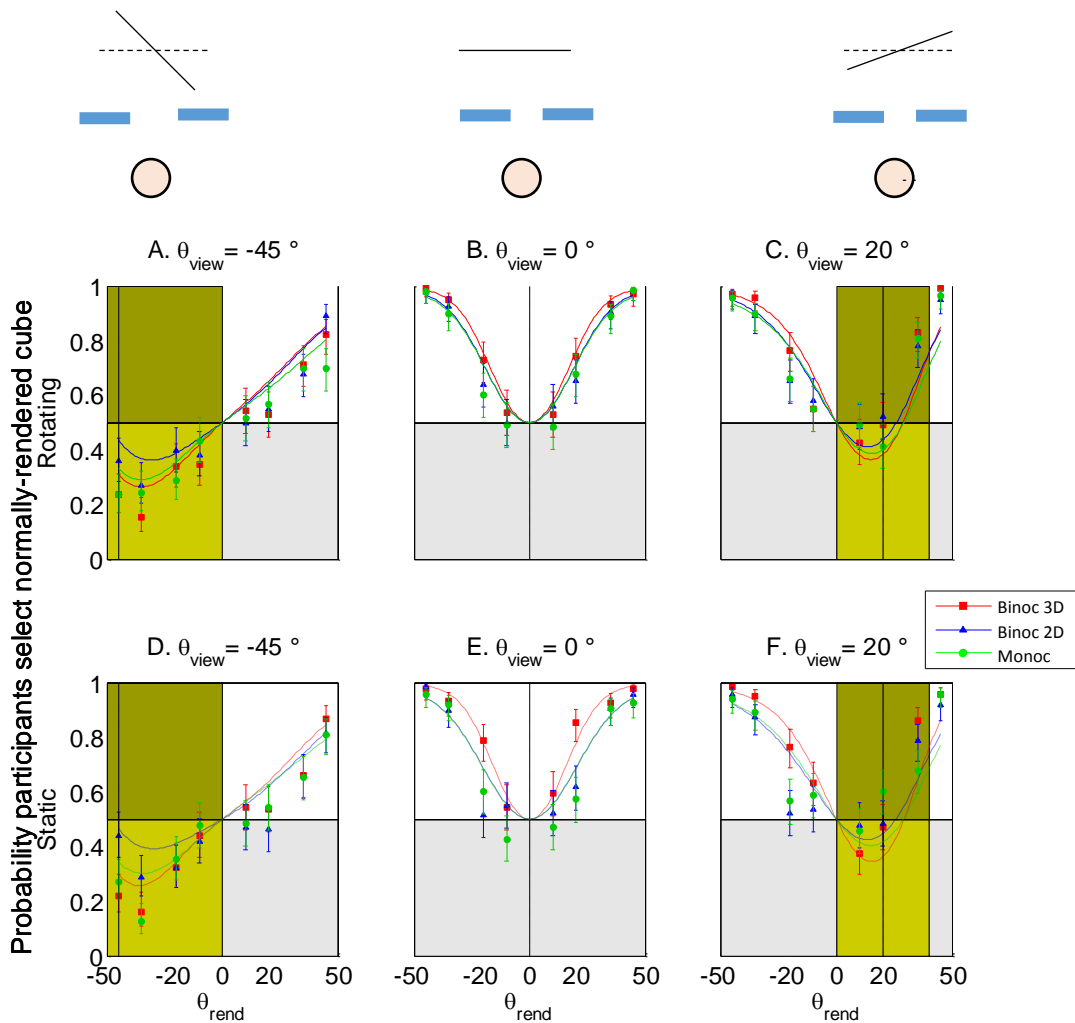
1243
1244
1245



1246

1247 **Fig. 3.7.** Results from the frame-visible condition as in Fig. 3.5, with the model fitted
1248 to the data. The model is fitted to all participants as a pooled group. The top row (solid lines)
1249 are for trials with rotating cubes, the bottom row (dashed lines) are for trials with static
1250 cubes. Shading as in Fig 3.4.

1251
1252
1253



1254

1255 **Fig. 3.8.** Results from the frame-occluded condition. As for Fig 3.7, except here a curtain
1256 prevented the subject from seeing the edges of the screen.

1257

3.3.4 Model fitting

1258

1259

I made this intuitive description quantitative in my two-factor model of “perceived veridicality” (Eq 3.1). As Figs 3.5 and 3.6 show, it gives a fairly good account of my results.

1260

Table 3.3 gives fitted model parameters and percentage variance explained for the different

1261

conditions. The parameters were fitted simultaneously to all data in a given object

1262

motion/binocularity/frame-occlusion condition, i.e. across all 3 panels in each row, the same

1263

parameters are used for all curves of a given colour. Table 3.3 gives the values of these

1264

parameters along with the percentage of variance explained. In every case, the model

1265

explains >80% of the variance. Interestingly the fits are generally better for the binocular S3D

1266 and B2D conditions, where they explain >93% and >85% of the variance respectively, than
 1267 for the monocular conditions, even though the fit parameters are fitted independently for
 1268 each binocularity condition. In the monocular conditions, subjects tend to choose the
 1269 obliquely-rendered cube slightly more often than my model can capture, especially when the
 1270 obliquely-rendered cube is close to being geometrically correct. However, the generally
 1271 successful performance of the model confirms the qualitative argument developed above,
 1272 that objects tend to look less distorted if they are rendered *either* for the geometrically
 1273 correct viewing angle *or* for normal, perpendicular viewing. An advantage of the model is
 1274 that it also allows us to make quantitative comparisons between the two mechanisms, as
 1275 follows.

		Fitted model parameters			Compensation index, $C=A/(A+B)$	% variance explained
		Weights	Sensitivity			
		Geometrically-rendering weight, B	For geometrically-correct rendering, r (in deg)	For normal-rendering, s (in deg)		
Frame-visible (Fig 3.5) Rotating	Monocular	1.84	26.58	50.52	0.62	87.52%
	Binocular 2D	1.72	23.92	42.04	0.64	91.20%
	Binocular S3D	2.12	20.59	41.81	0.59	90.98%
Static	Monocular	1.40	23.76	49.15	0.68	83.84%
	Binocular 2D	1.33	24.79	47.40	0.69	81.10%
	Binocular S3D	2.47	24.00	42.16	0.55	91.31%
Frame-occluded (Fig 3.6) Rotating	Monocular	1.82	23.00	51.93	0.62	88.60%
	Binocular 2D	1.68	23.41	44.83	0.64	90.91%
	Binocular S3D	2.18	22.65	45.01	0.58	91.23%
Static	Monocular	1.74	24.53	55.46	0.63	83.75%

Binocular 2D	1.40	23.01	47.87	0.68	85.13%
Binocular S3D	2.20	21.12	44.01	0.58	91.91%

1276 **Table 3.3.** Fitted model parameters for weights and sensitivity, including the implied
1277 effectiveness of compensation, as well as the percentage variance explained, for all
1278 conditions [Eq 3.2]. Table rows are colour coded the same as in figures 3.5 and 3.6. Note that
1279 the normally-correct weight, parameter A , was constrained to be equal to 3, and so is not
1280 included in the table (see section 3.3, Methods)

1281

1282 3.3.5 Quantifying the preference for normal rendering vs geometrical 1283 correctness

1284 Our model suggests that the mechanism favouring geometrical correctness is much
1285 more sensitive to incorrect rendering angle than that favouring normal rendering. The
1286 standard deviations fitted for the Gaussians are 23° and 47° respectively (means across
1287 conditions for data pooled across subjects; Table 3.2). However, the model suggests that the
1288 preference for normal rendering is generally stronger than that for geometrical correctness.
1289 The parameter A , representing the weight given to normal rendering, is generally larger than
1290 B , the weight given to geometrically correct images. To quantify this, I define the
1291 compensation index as the ratio $C=A/(A+B)$ (Table 3.3). $C=0$ would indicate no compensation,
1292 such that perception reflects only the geometrical correctness of the image on the retina,
1293 without regard for whether the on-screen image would appear correct when viewed
1294 normally. $C=1$ would indicate perfect compensation, such that viewing angle has no effect
1295 on perceived veridicality, and no preference for geometrical correctness. Another
1296 interpretation for compensation index becomes apparent when I consider how the perceived
1297 veridicality of an object rendered for frontoparallel viewing declines monotonically with
1298 viewing angle, relative to its veridicality at frontoparallel viewing. From Eq3.1. I have

1299
$$\text{relative veridicality} = \frac{A + B \exp\left(-\frac{\theta_{\text{view}}^2}{2r^2}\right)}{A + B} = C + (1 - C) \exp\left(-\frac{\theta_{\text{view}}^2}{2r^2}\right)$$

1300 [Eq3.5]

1301 At large viewing angles, this reduces to C. Thus, in my model, the compensation index C
 1302 describes how good a normally-rendered picture looks when viewed at the most extreme
 1303 viewing angles.

1304 Fig 3.9 plots the compensation index C for the different viewing conditions in my experiment,
 1305 pooled across participants. All 12 data-points in Fig 3.9 lie well above 0.5, indicating that the
 1306 preference for normal rendering dominates. This may seem surprising given that in the
 1307 yellow regions of Figs 3.7 and 3.8 where the two preferences conflict, data and model fits
 1308 both lie below 0.5, i.e. the geometrically correct cube is chosen preferentially. To see why
 1309 this occurs, it is helpful to consider how the model compares cubes rendered for $\theta_{\text{rend}}=0^\circ$
 1310 (normal) and $\theta_{\text{rend}}=30^\circ$, when the viewing angle is 45° . To the “normal rendering” mechanism
 1311 (A-term in Eq3.1), the normal-rendered cube is perfect, and the other cube is less veridical
 1312 because it is 30° away from the peak of the Gaussian. However, because the Gaussian is
 1313 broad, the difference is not extreme, so the “normal rendering” mechanism has only a weak
 1314 preference for the normal-rendered cube. Conversely, to the “geometrically correctness”
 1315 mechanism (B-term in Eq3.1), the obliquely-rendered cube looks acceptable - the 15° error
 1316 in render angle is less than one standard deviation – but the normal-rendered cube looks
 1317 very poor, with a 45° error of two standard deviations. This mechanism therefore has a
 1318 strong preference for the obliquely-rendered cube. When the preferences of both
 1319 mechanisms are summed, the strong preference for the obliquely-rendered cube wins out
 1320 over the weak preference for the normal-rendered cube.

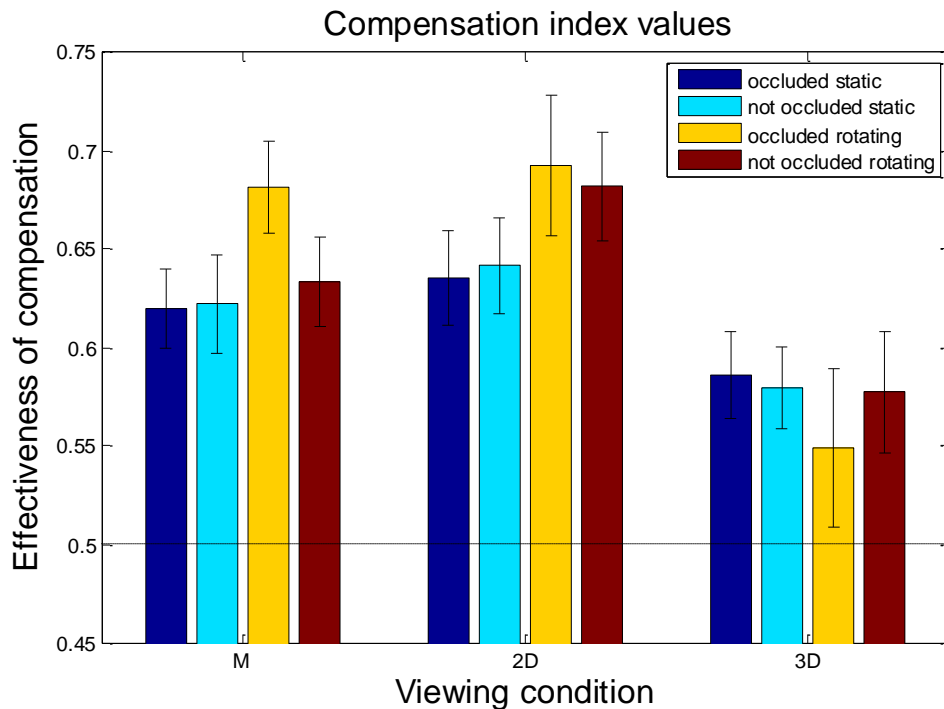
1321 When I consider results for individuals they are all very similar in their structure and pattern,
 1322 in that they all have compensation indices over 0.5 (suggesting that compensation was
 1323 similarly high, and relied on more than a preference for geometrically correct rendering).

1324 I can use the compensation index to assess how the different viewing factors influence the
 1325 relative weight of the competing preferences. To evaluate the significance of the different
 1326 factors, I generated compensation indices for each subject individually, and used a linear
 1327 general estimating equation, implemented in SPSS. I found a significant three-way
 1328 interaction between binocularity, object motion and frame occlusion ($P < 0.0005$). The full
 1329 results are shown below in table 3.4.

Factor or interaction (*)	χ^2	DF	P
Binocularity	50.583	2	<0.0005
Occlusion	0.333	1	0.564
Motion	0.851	1	0.356
Binocularity*Occlusion	6.823	2	0.033
Binocularity*Motion	27.717	2	<0.0005
Occlusion*Motion	1.287	1	0.257
Binocularity*Motion*Occlusion	42.914	2	<0.0005

1330 **Table 3.4.** How the individual factors affected the compensation index (main effects) and the
 1331 interactions between the factors, denoted by *. Results are from a linear generalized
 1332 estimating equation done in SPSS, and return the χ^2 value, along with the degrees of freedom
 1333 and associated P-value. Significant effects are highlighted green, non-significant results in
 1334 red.

1335 In summary, then, my statistical analyses both of the raw data and of the fitted model
 1336 parameters imply that binocularity, frame-occlusion and object motion all affect the balance
 1337 between the preferences for geometrical correctness vs normal rendering. In the next
 1338 sections, I consider each factor in turn.



1339

1340 **Fig. 3.9.** Compensation index, defined as the ratio $A/(A+B)$. A and B are model
 1341 parameters modelling the strength of the preference for normal rendering and for
 1342 geometrical correctness, respectively. Errorbars show standard error values generated during
 1343 the linear general estimating equation. Higher values of C indicate more compensation for
 1344 oblique viewing. The dashed horizontal line marks where both weights are equal. This figure
 1345 shows the values derived from fits to data pooled across all subjects. To carry out the analysis
 1346 of significance, I derived compensation indices for individual subjects from fits pooled to data
 1347 from that subject only.

1348 3.3.6 Effect of S3D

1349 Fig 3.9 implies that S3D weakens the compensation mechanism and gives more
 1350 weight to whether rendered objects create the correct image on the retina. This agrees with
 1351 Banks et al (2009). The compensation index drops from $C=0.66$ for binocular 2D viewing to
 1352 0.575 for stereoscopic S3D viewing (averaged over other viewing conditions): a small but
 1353 statistically significant difference. This effect can be seen in the raw data, when I compare
 1354 the S3D results in Fig 3.7 to the 2D (red squares vs blue triangles). This is particularly clear in

1355 Fig 3.5D, where the viewing angle is extreme ($\theta_{\text{view}}=-45^\circ$). When the obliquely-rendered cube
1356 is close to the correct retinal image (θ_{rend} close to θ_{view}), subjects perceive it as more cube-
1357 like than the normal-rendered cube when it is viewed in S3D, and select it >75% of the time.
1358 However, when viewed in 2D, it appears nearly as distorted as the normal-rendered cube
1359 and is selected only slightly more than half the time. The effect of S3D is also apparent in Fig
1360 3.5E, where the screen is viewed perpendicularly. Viewers are more sensitive to errors in
1361 rendering angle with S3D than with 2D or monocular content. In 2D, a rendering angle error
1362 as large as 20° cannot be distinguished from the correct rendering angle of 0° . In S3D,
1363 performance at $\pm 20^\circ$ is around 75%, suggesting that the error is detected on about half of
1364 trials.

1365 The difference between the S3D and other conditions is less pronounced with the
1366 solid cubes than with the wireframe cubes used in my previous study (Hands & Read, 2013).
1367 This suggests that a major effect of S3D in that study may simply have been its ability to
1368 disambiguate the Necker illusion. Once this illusion is removed through the use of solid
1369 cubes, S3D makes less difference to the perceptual compensation mechanisms which lead
1370 viewers to select mainly the normal-rendered cube. However, even with solid objects,
1371 viewing in S3D does tend to enhance the preference for geometrical correctness.

1372

1373 3.3.7 Effect of frame occlusion

1374 As my statistical tables show, the effect of frame occlusion was one of the weakest
1375 in my statistical analysis. Comparing Figs 3.7 (frame-visible) and 3.8 (frame-occluded), little
1376 difference is apparent. In Fig 3.9, the compensation index is barely affected by frame
1377 occlusion, moving from $C=0.63$ when the frame was visible to 0.62 when it was occluded
1378 (averaged over other viewing conditions), which is not statistically significant. However,
1379 occluding the frame does produce a substantial – and significant – drop in compensation for
1380 the monocular static condition (Fig 3.9). This is in qualitative agreement with Vishwanath et

1381 al (2005). These authors found some compensation with monocular viewing when the
1382 picture frame was visible, but none for monocular viewing through an aperture. I also saw a
1383 significant effect of occlusion when restricting my analysis to data where the normal-
1384 rendering and geometrical-correctness preferences make opposite predictions ($P < 0.0005$,
1385 Table 3.3).

1386 The small effect of frame occlusion overall is surprising since the occlusion did appear to be
1387 very effective in removing conscious awareness of screen orientation; even the authors could
1388 not reliably say which side of the screen was closer when viewing it through the occluder.
1389 Yet this does not seem to have produced a substantial tendency to select the geometrically-
1390 correct cube over the normal-rendered one. For binocularly-viewed stimuli, frame occlusion
1391 may have had little effect because disparity and vergence cues to screen orientation
1392 remained available to viewers and may have been used unconsciously to compensate for
1393 screen slant (Rogers & Bradshaw, 1993; Vishwanath et al., 2005). For monocular stimuli,
1394 frame occlusion has more of an effect (Fig 3.9), and indeed I see a significant difference in
1395 the results for $\theta_{\text{view}}=\theta_{\text{rend}}=-45^\circ$. Pooling static and rotating stimuli in Figs 3.7 and 3.8, viewers
1396 were closer to chance when they could see the screen edges (chose the obliquely-rendered
1397 cube on 120 out of 352 trials) and preferentially chose the obliquely rendered cube when
1398 the edges were occluded (90 out of 352 trials). A similar effect persists at $\theta_{\text{rend}}=-35^\circ$.
1399 Elsewhere, the lack of an effect seems to be because my participants were relatively
1400 insensitive to the distortions caused by rendering angle, and thus did not notice when these
1401 distortions were corrected.

1402 3.3.8 Effect of object motion

1403 Our statistical analysis of the raw data indicates that object motion is a significant
1404 factor in both the full data set and the important subset where the preference for
1405 geometrical correctness is pitted against the preference for normal rendering ($P = 0.011$ and
1406 $P = 0.023$ respectively). However, object motion did not have the effect I had expected. I had

1407 speculated that structure-from-motion cues might contribute to the compensation
1408 mechanism, increasing the preference for normal-rendered objects. In fact, object motion
1409 decreased the compensation index for both monocular and binocular 2D cubes (Fig 3.9); this
1410 was significant in the monocular condition. In stereoscopic 3D, object motion did tend to
1411 increase compensation index, but the increase was not significant. As table 3.2 shows, when
1412 considering the full set of raw data, there is a significant interaction between object motion
1413 and binocularity ($P < 0.0005$). Pairwise comparison shows that object motion has a significant
1414 effect even when considering the individual binocularity conditions ($P = 0.011$ for all three
1415 conditions of S3D, B2D and monocular) However, this interaction was not significant when I
1416 restricted my analysis to the subset of data in table 3.3. I conclude that overall, object motion
1417 has little consistent effect on perception.

1418 3.4 Discussion

1419 Still photographs and movies are generally designed to be presented on a surface
1420 which is frontoparallel to the viewer. Despite this, they continue to look veridical when
1421 viewed from an oblique angle. As discussed in section 3.2 (Introduction), this is partly
1422 because humans are fairly insensitive to the image distortions produced by oblique viewing,
1423 but also because the visual system actively compensates for oblique viewing. This
1424 compensation mechanism ensures that an image viewed on a screen is perceived as if the
1425 screen is frontoparallel to the observer, even if it is in fact viewed obliquely. Stereoscopic 3D
1426 brings its own complications, e.g. filming with converged camera axes, but has always
1427 implicitly relied on the same compensation mechanism previously shown to exist for 2D
1428 displays. However, this assumption has not yet been adequately tested for S3D content.

1429 There are good reasons to imagine that this compensation mechanism might be weaker for
1430 stereoscopic 3D content, mainly because disparity is now not a reliable cue to the location
1431 and orientation of the screen plane. Informally, one can experience this by moving from left
1432 to right in front of an S3D image. The image appears to move in synchrony with you, as when

1433 a portrait's eyes appear to follow the viewer around the room (Koenderink, van Doom,
1434 Kappers, & Todd, 2004; Perkins, 1973) but now extended to the whole depicted object. S3D
1435 content is often already affected by a number of distortions, such as the puppet theater
1436 effect or cardboard cut-out effect (Banks et al., 2012; Yamanoue, Okui, & Okano, 2006) . If
1437 oblique viewing produces further distortions in perceived depth or shape, this would be a
1438 further problem for creators of S3D content. It would be particularly difficult to address in
1439 applications such as 3D cinema, where content must be viewed by large numbers of people
1440 simultaneously.

1441 I examined this issue by comparing images rendered for a range of oblique viewing angles
1442 with those rendered for a frontoparallel screen, in both 2D and S3D. I confirm that the human
1443 visual system compensates to some extent for oblique viewing angles (Bereby-Meyer et al.,
1444 1999; Perkins, 1973; Vishwanath et al., 2005). Due to this compensation, images tend to
1445 appear veridical if they are rendered for normal (orthogonal) viewing, even if actually viewed
1446 from an oblique angle. However, I also find that a competing factor affects appearance:
1447 images also tend to appear veridical if they are rendered for the geometrically correct
1448 viewing angle. This effect predominates for viewing angles more oblique than about 20°. I
1449 have produced a quantitative model which well describes viewers' perceptual judgments on
1450 this task across a wide range of viewing and rendering angles.

1451 [3.4.1 Sensitivity to viewing angle](#)

1452 Our results confirm viewers are relatively insensitive to distortions caused by
1453 inappropriate viewing angles. In 2D, most viewers cannot tell the difference between a
1454 stimulus rendered for perpendicular viewing and a stimulus rendered with up to 20° error in
1455 viewing angle (Cutting, 1987; Perkins, 1973). My modeling also suggests that viewers are
1456 much less sensitive to oblique viewing angle in content that was rendered to be viewed
1457 normally, than they are to deviations from the geometrically correct viewing angle in content
1458 that was rendered for oblique viewing.

1459 3.4.2 Range over which compensation operates

1460 It seems reasonable to expect that viewers should compensate better for small
1461 oblique viewing angles than for large ones. My data appear to support this. For example,
1462 when $\theta_{\text{view}}=20^\circ$, viewers show only a weak preference for the geometrically correct cube
1463 ($\theta_{\text{rend}}=\theta_{\text{view}}$), suggesting that compensation makes the normally-rendered cube appear nearly
1464 as veridical, whereas when $\theta_{\text{view}}=-45^\circ$, they show a stronger preference for the geometrically
1465 correct cube (Figs 3.5 and 3.6, panels AD vs CF). According to my model, pictures appear
1466 more veridical for small oblique viewing angles than for large ones [Eq3.5]. My model
1467 assumes that compensation works equally well for all viewing angles (blue curves in Fig 3.4).
1468 The decline in veridicality comes from the preference for geometrical correctness against
1469 which the compensation mechanism is pitted. In my model, taking $C=0.62$ and $r=23^\circ$ as
1470 representative values, veridicality never drops below 62% of optimal even at the most
1471 extreme angles, and remains above 80% even out to viewing angles of 28° .

1472 3.4.3 Regression to expected shape

1473 Some previous authors have suggested that people have a tendency to “regress”
1474 distorted images of familiar objects to their expected form (Gombrich, 1972; Thouless, 1931).
1475 Presumably, regression is imagined as operating on retinal images to make them appear
1476 more geometrically correct. If the regression operated perfectly no matter what the
1477 distortion, both objects in my experiment would appear equally cube-like and performance
1478 would be at 50% throughout. The “geometric” term in my model [Eq3.1, terms in B] is
1479 effectively an implementation of regression which allows for the possibility that regression
1480 is more effective for small departures from geometrical correctness. The parameter r
1481 describes the range over which regression operates, with perfect regression corresponding
1482 to the case $r \rightarrow \infty$ and $A=0$.

1483 3.4.4 Differences between stereoscopic 3D and 2D

1484 In line with expectations, I found that compensation for oblique viewing works
1485 better in 2D images than for stereoscopic 3D. A plausible reason is that, in binocular 2D
1486 viewing, the true orientation of the screen can be deduced from binocular cues such as
1487 disparity or vergence, even when the edges of the screen are occluded from view. This makes
1488 it possible to apply the appropriate compensation (Vishwanath et al., 2005).

1489 Banks et al (2009) also compared shape distortions in oblique viewing for 2D and S3D stimuli
1490 and found that viewing in S3D abolished compensation almost completely. In contrast, I find
1491 that compensation still dominates even in S3D, though it is less effective than in 2D. I
1492 highlight three differences in protocol which may contribute to this difference. First,
1493 perceptual invariance depends on the stimulus, and particularly the depth variation in the
1494 stimulus (Banks et al., 2009). My stimuli were solid cubes, with a sidelength from 6 to 14cm,
1495 viewed from a distance of 120cm. The stimuli of Banks et al were hinged wireframe squares
1496 with a sidelength of 30cm, viewed from 45cm. my stimuli thus contained relatively less depth
1497 variation. I opted for this stimulus rather than one which would be more closely related to
1498 Banks et al. as in this thesis I wish to consider stimuli and setups as close to natural viewing
1499 in the home as possible. In that respect sitting participants further away made more sense.
1500 Second, the longer viewing distance used in my study may have enhanced the preference for
1501 normal rendering. Artists since the Renaissance have discussed the recommended distance
1502 at which to capture the perspective projection of an object, in order for it to look pleasing
1503 and natural. Hagen and colleagues have argued for a distance at least ten times the mean
1504 object size along its various dimensions, very close to that used in my experiments (Hagen &
1505 Elliott, 1976; Hagen, Elliott, & Jones, 1978); Leonardo da Vinci recommends a smaller
1506 distance of 3 times the height of the object (Da Vinci, 2012). The longer viewing distance
1507 reduces the amount of perspective convergence, making the projection closer to
1508 orthographic. Viewers report such projections as appearing more veridical, even when they

1509 are geometrically incorrect for the given viewing distance (Hagen & Elliott, 1976; Hagen et
1510 al., 1978). This reflects the fact that viewers do not compensate for wrong viewing distance
1511 as they do for oblique viewing (Cooper, Piazza, & Banks, 2012). Thus, when viewed and
1512 rendered normally, my cubes should have looked veridical, whereas the hinge stimuli of
1513 Banks et al may still have looked “wrong” because of the short viewing distance relative to
1514 the size of the object. This may have weakened the effectiveness of the compensation for
1515 oblique viewing. Finally, my stimuli were renderings of cubes, where there is a clear canonical
1516 form which may have influenced perception, whereas those of Banks et al were wireframe
1517 hinges, with no clear expectation regarding hinge angle. One might expect the “regression”
1518 mechanism of Thouless (1931) and Gombrich (1972) to operate more strongly on cubes than
1519 on hinges. In the terms of my model, this would be expected to boost the parameter r , i.e.
1520 make subjects more tolerant of departures from geometrical correctness. It might also boost
1521 the weight of B relative to A , thus reducing the compensation index C . If so, this could
1522 potentially be one reason I found less compensation with cubes than Banks et al did with
1523 hinges.

1524 Our longer viewing distance and use of familiar objects makes my study more relevant to
1525 typical applications of S3D displays in entertainment. The S3D entertainment industry can
1526 therefore be reassured by the lack of difference I found between S3D and 2D content, even
1527 at a viewing angle as large as 20° . The differences only became apparent at the largest
1528 viewing angle used, 45° . For most S3D display systems, such an extreme viewing angle
1529 already causes other problems such as increased cross-talk or contrast changes.

1530 [3.4.5 Effect of frame visibility and object motion](#)

1531 A new contribution of my study was that I investigated the effect of object motion.
1532 This is particularly relevant for entertainment applications of S3D, where content is generally
1533 dynamic. I had speculated that structure-from-motion cues, together with the rigidity
1534 heuristic, might enable the visual system to compensate more effectively for oblique

1535 viewing. In fact, object motion had little effect in S3D, and tended to weaken compensation
1536 in the 2D and monocular conditions. This suggests that the difference between S3D and 2D
1537 TV and movies may be even less than that for S3D and 2D static images.

1538 [3.4.6 Limitations](#)

1539 My experiment suffered from high levels of crosstalk. Crosstalk (or ghosting) refers
1540 to any ‘leaking’ of the left eye’s image into the right eye and right eye images into the left.
1541 This can disrupt the perceived depth of the image and lead to double vision, as both eyes see
1542 some part of both the stereoscopic images displayed on the screen. High levels of crosstalk
1543 can lead to the image seen being perceived incorrectly and affect image quality, so it is
1544 essential to minimize crosstalk to achieve high-impact, impressive 3D images (Woods, 2011).
1545 Since the 3D television was manufactured for perpendicular viewing, the amount of crosstalk
1546 between the images increased substantially with oblique viewing: up to 7%. The fact that the
1547 experiments were conducted in darkness also tended to make any crosstalk more visible to
1548 observers. I could have reduced the contrast of the images to attempt to reduce the crosstalk
1549 observed by the viewers. I chose not to do this, as in conventional viewing at home this is
1550 possibly not something viewers would opt for, and instead they would see the image with
1551 the crosstalk present from the oblique angle. The high levels of crosstalk could mean that my
1552 “monocular” images are in fact 2D binocular images in which one image is much lower
1553 contrast than the other. Thus, my experiments may underestimate the difference between
1554 the 2D and monocular stimuli, especially for oblique viewing. However, I did repeat some of
1555 the monocular conditions with one eye covered, instead of using the 3D glasses to interleave
1556 “monocular” and binocular stimuli, and obtained broadly similar results. When it comes to
1557 S3D TV, the crosstalk increases the ecological relevance of my study, since these levels of
1558 crosstalk are those which would be experienced by S3D TV viewers in a normal home
1559 environment.

1560 The condition of monocular viewing when the frame was occluded from view (green data in
1561 Fig 3.6ACDF) was intended to remove all information about screen orientation, by removing
1562 disparity and perspective cues. However, I was evidently not successful in this, since there
1563 was evidence of active compensation even in this monocular, frame-occluded condition. For
1564 example, at a viewing angle of $\theta_{view}=20^\circ$, my data show an asymmetry in the effect of render
1565 angle. This indicates that objects looked more cube-like when rendered for a viewing angle
1566 closer to frontoparallel than the actual viewing angle, than when the render angle was
1567 equally distant from the actual viewing angle, but in the opposite direction. This must mean
1568 that subjects had access to some source of information about screen orientation. Possible
1569 sources of information include accommodation, motion parallax from small head
1570 movements within the headrest, gradients in luminance across the screen and so on.
1571 However, this limitation does not affect my main conclusion, which relates to the difference
1572 between binocular 2D and S3D viewing. Less surprisingly, in this impoverished viewing
1573 condition, subjects had greater uncertainty and were less able to perceive any differences
1574 between the two cubes. my model fits indicate lower sensitivity under monocular viewing in
1575 almost all cases,

1576 In debriefing after the experiment, several participants stated that they tended to choose
1577 the smaller cube when the task was difficult, presumably because any deviations from
1578 cubeness are harder to detect in smaller objects. Since the size of my cubes was chosen at
1579 random, this strategy would push performance towards chance, making it harder for us to
1580 detect effects of my experimental parameters.

1581 This study has only considered one effect of oblique viewing: distortions in perceived shape.
1582 Another approach would be to consider whether viewing stereoscopic content from
1583 inappropriate viewing angles is a source of viewer discomfort (Howarth, 2011; Lambooi,
1584 Ijsselsteijn, Fortuin, & Heynderickx, 2009). It would be interesting to look at the various

1585 different definitions of a “zone of comfort”, the range of depth allowed in 3D content before
1586 discomfort begins to adversely affect the viewing, and see if changing the viewing angle has
1587 any effect on the zone of comfort (Shibata, Kim, Hoffman, & Banks, 2011).

1588 Finally, my study only asked viewers to consider which of two objects most resembled a
1589 perfect cube. I did not assess what the viewers were using to make this distinction, nor how
1590 they perceived the objects. Accordingly, my model also only predicts perceptual judgments
1591 in this comparison task, rather than directly predicting perceived shape.

1592 3.5 Conclusion

1593 When viewing a familiar object, especially one in motion, viewers are very nearly as
1594 tolerant to oblique viewing in S3D as in 2D. This is partly because viewers are fairly
1595 insensitive to detecting when they are sat at an incorrect viewing angle, and partly because
1596 of a compensation mechanism which makes content rendered or filmed for a frontoparallel
1597 screen continue to appear veridical even when viewed obliquely. Contrary to previous
1598 literature suggesting that this compensation is substantially impaired for S3D content, I find
1599 little difference. This helps explain why S3D content is popular and effective even though it
1600 is usually viewed from the “wrong” position.

1601 4. The interaction between familiar size and vergence depth 1602 cues in stereoscopic three-dimensional displays

1603 4.1 Introduction

1604 A great deal is known about the information concerning depth from different cues,
1605 such as texture, stereoscopic vision, perspective, viewing distance, shading, motion parallax,
1606 occlusion and also haptic and auditory cues (Cavanagh, 1987; Cullen, Galperin, Collins,
1607 Kapralos, & Hogue, 2012; Ernst & Banks, 2002; Hoffman et al., 2008; Snowden, Snowden,
1608 Thompson, & Troscianko, 2012). The information is integrated together in the visual cortex
1609 of the brain (Banks et al., 2012) and the natural scene formed. However, cues are judged
1610 independently and often inconsistently to measure distance, shape and size (Brenner & van
1611 Damme, 1999). This can lead to cues conflicting and problems arising such as depth sign,
1612 depth magnitude and slant (Banks et al., 2012). Problems can also arise from the
1613 misinterpreting of information by the brain, such as Alice in Wonderland syndrome (Brumm
1614 et al., 2010; Golden, 1979; Kuo, Chiu, Shen, Ho, & Wu, 1998; Todd, 1955), where typically
1615 parts of the natural scene (including on occasion parts of the patient themselves) appear
1616 distorted and either too large or too small (Kuo et al., 1998).

1617 One clearly recognised conflict that can potentially arise from viewing S3D content is that
1618 between the information given from the apparent size of an object and vergence
1619 (stereoscopic information about the depth) required by eyes to correctly perceive the object
1620 (Foley, 1968), particularly that of a very familiar object (Gregory, 2015). This relationship, not
1621 including the concept of stereopsis, has been noted since the ancient Greeks (Euclid) and is
1622 well explained by Emmert: '*for a given retinal image size, perceived size is proportional to*
1623 *perceived distance*' (Snowden et al., 2012). It is well known in research that the angular sizes
1624 of objects can give strong cues to their depth (Wallach, Frey, & Bode, 1972; Walsh &
1625 Kulikowski, 1998). In 2D television and cinema, the distance to the object is undefined.

1626 Viewers will know approximately the distance to the screen plane, but they also know that
1627 the depicted objects are not intended to lie on the screen plane in the image shown.
1628 Therefore they are free to assume that the object's distance is such as to make the retinal
1629 size correct for the known physical size. I.e. a large image of a mouse appears close, a small
1630 image of a mouse appears far, but both appear mouse-sized on a 2D display. This can fail in
1631 S3D, since now the disparity potentially tells you the distance to the object, and raises
1632 questions about the familiar size assumption. Thus, in S3D, which now gives vergence and
1633 disparity cues to depth, there is more depth information available compared to 2D or
1634 monocular images. This allows for both the absolute and relative depth in the image to be
1635 absolutely assessed. This information can then be integrated and compared with the angular
1636 size that the object occupies on the retina. These two pieces of information can then be
1637 compared to the memory of previous objects of the same type and the field of view and
1638 depth information they held. If this information is not the same as previous memories, then
1639 a conflict occurs. This results in the visual cortex needing to combine the information to get
1640 a most reliable guess as to the correct size and depth of the familiar object. In some cases
1641 this combination could be to completely disregard one depth cue, e.g. the vergence
1642 information, and focus instead on only the familiar size information.

1643 The conflict between the two different pieces of depth information can arise due to the
1644 producers of the content increasing the inter-axial values of the cameras or using different
1645 camera configurations to enhance the sense of depth the viewer perceives, such as a toed-
1646 in configuration. With the perceived distance on the S3D display to the object now
1647 potentially being different to the actual distance of the screen where the object is displayed,
1648 the perceived size could also be altered. Consequently, the relationship between familiar size
1649 and depth is affected, resulting in the cue conflict (Yamanoue et al., 2006).

1650 This change in the appearance of the scene is known as miniaturisation (or the ‘puppet-
1651 theatre’ effect) (Banks et al., 2012). The effect makes the 3D object in question appear
1652 noticeably smaller compared to what the viewer would expect to see naturally (Hopf, 2000;
1653 Meesters, Ijsselsteijn, & Seuntiens, 2004; Yamanoue et al., 2006). The effect can be in the
1654 other direction (gigantism) where the object of focus appears too large. This occurs when
1655 the interaxial value of the two camera lenses is too small, compared to too large for
1656 miniaturisation. However gigantism occurs considerably less often than miniaturisation. The
1657 effect cannot be physically measured as it is subjectively assessed, therefore each person
1658 may perceive the amount of miniaturisation differently (Yamanoue et al., 2006). The level of
1659 effect is heightened when using the ‘toed-in’ camera configuration (Banks et al., 2012).
1660 Furthermore, the probability of the effect occurring increases with larger viewing distances
1661 (Yamanoue et al., 2006). Some of these concepts were shown in Fig 1.2. in the introduction.
1662 Solutions have been studied to minimise the effect, with some techniques allowing greater
1663 stereoscopic distances to be presented (Hopf, 2000). The size-stereo miniaturisation conflict
1664 is often reported in the viewing of football matches in S3D. As the matches are filmed a great
1665 distance away from the players, there is little sense of depth for the viewer. As mentioned
1666 previously, to enhance this sense of depth the interaxial values are vastly increased and
1667 ‘toed-in’ configuration often used, which in turn enhances the depth of the images
1668 presented. This can lead to the players appearing unnaturally small compared to the pitch
1669 and stadium, for this reason losing the sensation of realism (Yamanoue et al., 2006). Another
1670 good example of this is in the S3D feature film ‘Gravity’, which set the interaxial distance of
1671 the two ‘virtual cameras’ in creating the CGI footage of space to be at a distance of ‘infinity’,
1672 so as to give the impression of vastness to space and ‘smallness’ to the characters in the
1673 movie. I intend to assess this conflict as although it is well known, it has not been studied in
1674 great detail in a research environment.

1675 However, setting the interaxial values of the camera larger than the standard interocular
1676 distance (IOD) is not the only way to generate a sense of miniaturisation, as argued by Smith
1677 and Malia. In their paper, they considered miniaturisation and gigantism as a result of width
1678 magnification, the calculation of which includes the interaxial values (as has been discussed)
1679 (Smith & Malia, 2015). In my study I will be focussing on the change in the vergence
1680 information, and the warping due to an increase (or decrease) in size, which could be
1681 considered as a change in the magnification of the image.

1682 Experiments have been completed considering the effect of similar size on estimated size of
1683 objects, determining that an object that has familiarity associated with it will still influence
1684 impression, in spite of other cues present (such as the observed size in that instance of
1685 viewing and the respective binocular information). The belief that a familiar size cue
1686 influences perception is supported well by McIntosh (2008), who showed in his experiment
1687 that the motion to grasp an object is influenced by the familiarity of it, despite binocular cues
1688 being present (McIntosh, R., Lashley, G., 2008). In my experiment I am not testing familiarity,
1689 as all the experimental parameters were based on familiar objects. However the fact that
1690 known familiar size does have an influence on perception helps to validate this experiment.

1691 Many people also associate roundness as an important factor in the perception of S3D
1692 content (Devernay & Beardsley, 2010). If an object has too much depth or not enough then
1693 an object can appear to have incorrect depth within itself. That is, for example, a sphere
1694 could appear as an ellipse whose depth is either too large for its height or too small. I do not
1695 consider this depth factor here as my stimulus was chosen specifically to remove any
1696 roundness cue.

1697 Familiar size as a cue to depth is inherently tied to the stereoscopic cues to depth. Wallach
1698 and O'Leary considered these two with physically created stimuli on cardboard and deemed
1699 that the interaction between the two cues was of interest. In their study, they did not

1700 separate the familiar size and vergence cues, but recorded mathematically that both must
1701 be involved in the perception of depth (O’leary & Wallach, 1980).

1702 In this chapter I intend to look at the cue conflict that arises between familiar size and stereo
1703 information in S3D content and whether one cue is preferable to the other in visual
1704 perception. Without the combination of cues, images can appear unrealistic for the viewer
1705 (Scarfe & Hibbard, 2011). However, when two cues conflict with one another in S3D content,
1706 one of them will more often than not violate the perception of what one expects the original
1707 natural scene to be viewed as and the other one will take preference (Cavanagh, 1987).
1708 Studies have demonstrated that different weightings can be put on cues depending on the
1709 environment or situation the subject was in (Rushton & Wann, 1999). There also appears to
1710 be some prior bias depending on how the cues have been used during developmental years,
1711 when the system is still considered plastic (Rushton & Riddell, 1999). It is known that the
1712 human visual system is better at detecting relative disparity over absolute disparity (Parker,
1713 2007). This could somewhat explain some of the reactions to familiar size and binocular
1714 disparity cues in the scenes seen, as most of the S3D scenes will be displaying relative
1715 disparity, along with various other cues to depth such as occlusion, shading, and texture. It
1716 could be that the introduction of disparity in S3D media means that different considerations
1717 need to be given to the setting of the scenes in, production and filming of S3D content. I
1718 conducted an experiment to consider whether the chosen structure of my experiment would
1719 allow us to compare the familiar size and vergence cues. Once this was confirmed I continued
1720 with my full experiment. Additionally, the amount by which each cue is weighted is thought
1721 to be judged on the reliability of the cue itself in the given situation (Rosas, Wagemans, Ernst,
1722 & Wichmann, 2005). Therefore, in the main experiment, I also intend to numerically assess
1723 the conflict between size information and disparity to see if it is possible to assign weighting
1724 bias to the size and stereo information cues via signal detection modelling. I ran three
1725 experiments to do this. Work in this chapter has been presented at the Vision Science Society

1726 conferences 2015 (Hands, Khushu & Read, 2015), and the other experiment study has also
1727 been published (Hands et al. 2014)

1728 4.2. Initial experiment

1729 4.2.1 Material and methods - Participants

1730 Participants were recruited via an internal volunteer scheme and were recruited on
1731 the basis they had no visual problems other than wearing glasses or contact lenses. The work
1732 was approved by Newcastle University Faculty of Medical Sciences Ethics Committee. 10
1733 participants (8 F, 2 M, all naïve to the study) were used in the initial experiment. Participants
1734 were paid with a £10 gift voucher for their participation.

1735 4.2.2 Material and methods - Equipment

1736 Subjects were shown a computer generated image of a standard credit card (ISO/IEC 7810
1737 identity card), as this object is well known and of a particular size (8.56cm x 5.398cm). The
1738 images were presented on an LG passive 3D TV (LG 47LD920-ZA) using the computer
1739 programming environment Matlab ("www.mathworks.com,") and the Psychtoolbox
1740 extension (Brainard, 1997; Kleiner et al., 2007). Participants sat at different viewing
1741 distances of 50cm, 100cm and 200cm in different blocks of the experiment, measured
1742 perpendicularly from the centre of the screen to the midpoint of the eyes, with their eyes
1743 at the same height as the centre of the screen. They wore passive 3D glasses throughout
1744 the experiment, so as to not be able to tell when there was no disparity between the left
1745 and right eye images. The three different viewing distances were used to attempt to assess
1746 how different viewing distances might affect the responses. I believe that viewing over
1747 these three distances will allow us to discuss the vergence-accommodation conflict with
1748 regards to this cue combination, as at the 200cm distance, accommodation cues are
1749 somewhat eliminated (Banks et al., 2012; Hoffman et al., 2008). Therefore, only vergence
1750 should influence the perceived depth of the stimuli at that distance. The subject was sat on

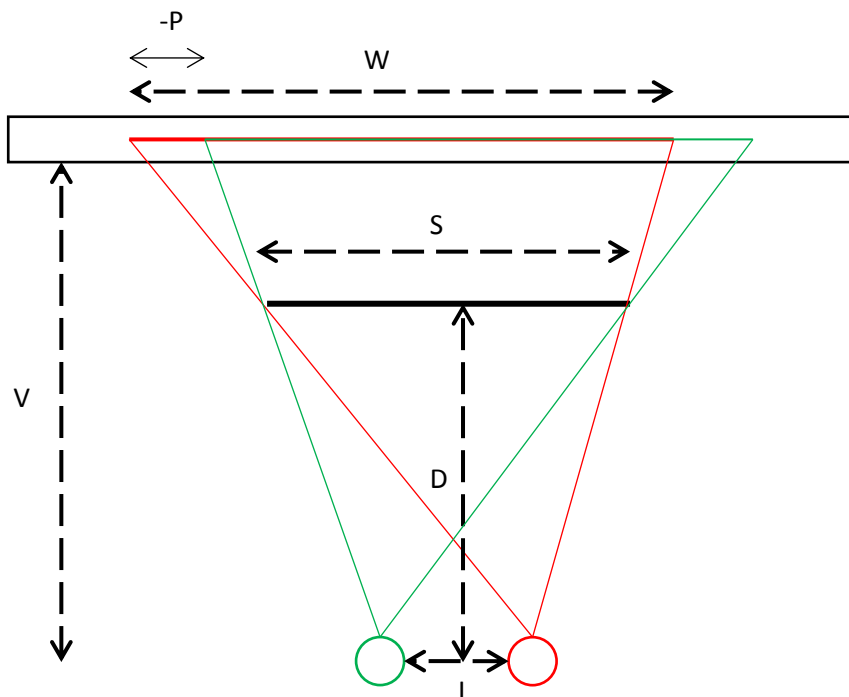
1751 an adjustable chair to allow for height change. This meant each subject had their eyes in
1752 the correct position both horizontally and vertically. Participants were given a reference
1753 credit card that they could hold and look down at, but were asked to not raise the card up
1754 to the screen to aid in choosing whether the card was bigger or smaller, as this would have
1755 enabled participants to make relative comparison judgements purely on the basis of retinal
1756 size, independent of perception, which I wanted to avoid. Finally, a keyboard was used for
1757 the subjects to make their choices using the up and down arrow keys.

1758 4.2.3 Material and methods - Procedure

1759 The main aim of the experiment was to assess the interactions between the two
1760 different depth cues, to see if one cue was preferably used to assess depth information and
1761 to see if conflicting information resulted in any strange conclusions about the depth or size
1762 of the familiar object. Thus in the experiment each trial consisted of a credit card displayed
1763 at a set physical size W and set parallax P (Fig. 4.1. below and the mathematical modelling
1764 section, 4.3.4). To compare absolute and relative disparity I opted to show the stimulus card
1765 either on a completely black background (to measure absolute depth decisions) and a
1766 background made up of Gaussian noise (to measure relative depth decisions). Participants
1767 were asked to select whether the card displayed was “in front of” or “behind” the screen
1768 plane.

1769 For the black background, 7 sizes and 7 parallaxes gave a total of 49 different parameter
1770 options as each size was shown with each disparity. The noisy background only had 28
1771 parameter options as only negative and zero parallax disparities from the above 7 were used
1772 (4 values). This was to ensure that the disparity of the noisy background could be kept
1773 constant on the screen plane, hence having zero disparity itself. Therefore, a total of 77
1774 parameter options were shown to the subject. At times the card would appear the exact
1775 same size as the reference card and in the screen plane; in these instances the subject was
1776 forced to make a decision as no ‘same’ option was available, which I would expect to be

1777 simply a chance selection. Each parameter permutation was repeated 20 times. The 1540
 1778 (77 x 20) trials were displayed over 5 blocks to allow the subject to have a few minutes rest
 1779 between each block. Each subject saw a random order of the different trial parameters.



1780
 1781 **Fig. 4.1.** How the on-screen size of the card would change depending on its depth.
 1782 Geometrically, in order to depict a virtual object of physical size S , I must draw the image with
 1783 width W on screen. I will refer to W as “on-screen size” and to S as “physical size”. Here $I =$
 1784 Interocular distance, $D =$ distance from the viewer to the stimulus’s virtual depth, $V =$ physical
 1785 distance from the participant to the monitor screen, $S =$ geometrically implied width of the
 1786 stimulus at its virtual depth, $W =$ physically displayed width of the stimulus on the monitor, P
 1787 = the parallax between the left and right images (which gives the stimulus its virtual depth).
 1788 In the diagram the parallax is denoted $-P$ to ensure it fits with the convention of negative
 1789 parallax for objects in front of the screen.

1790 Stimuli were generated by PH and the experiments run by Aniketa Khushu (please see
 1791 Acknowledgements) using the computer programming environment Matlab and the
 1792 Psychtoolbox extension (D. H. Brainard, 1997; M. Kleiner et al., 2007; Pelli, 1997)

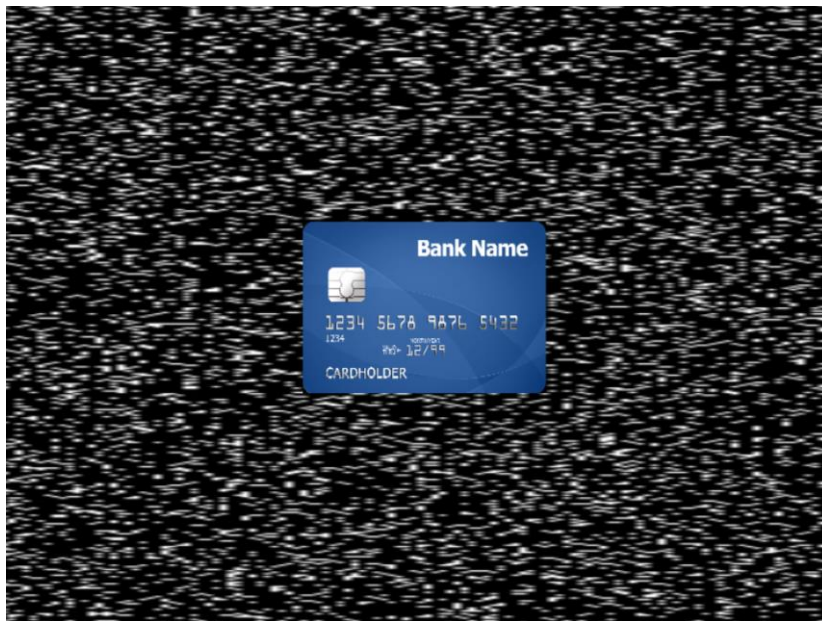


1793

1794 **Fig. 4.2.** The stimulus used was a credit card, shown above, at different widths and heights.

1795 Here it is displayed at a correct width and height (width 8.56cm, height 5.398cm). Image

1796 taken from <http://www.psdgraphics.com/psd/credit-card-template/>.



1797

1798 **Fig. 4.3.** Example stimulus of card (at zero disparity) on noisy background.

1799 4.2.4 Material and methods - Parameters

1800 7 widths W were chosen to be displayed as the width values to be measured. The
 1801 ‘true’ correct size of a credit card is 8.56cm. The stimulus used was therefore chosen to be
 1802 $8.56\text{cm} \pm 1, 2 \text{ and } 3\text{cm}$ respectively. Eq. 4.1. below shows how the interaction of W and I can
 1803 be used to generate P such that the size of S is the same as a natural card. This means that

1804 were someone to hold an actual credit card at a depth D from the viewer the retinal
 1805 impression on each eye would be identical to the displayed width W and parallax P at a
 1806 viewing distance V from the screen.

1807 Eq. 4.1.
$$Parallax (P) = \frac{W I}{8.56} - I$$

1808 What is immediately clear is that the viewing distance V has no effect on this judgement, as
 1809 it is not in the calculation. This means that regardless of the viewing distance (0.5m, 1m or
 1810 2m) the parallax that the different W values generated were the same. I used 6.3cm as an
 1811 estimate for interocular distance (Dodgson, 2004). Table 4.1 below gives us the widths W
 1812 and respective parallaxes P which were used.

Width parameter, W, cm	Parallax parameter, P, cm
5.56	2.208
6.56	1.472
7.56	0.736
8.56	0
9.56	- 0.736
10.56	- 1.472
11.56	- 2.208

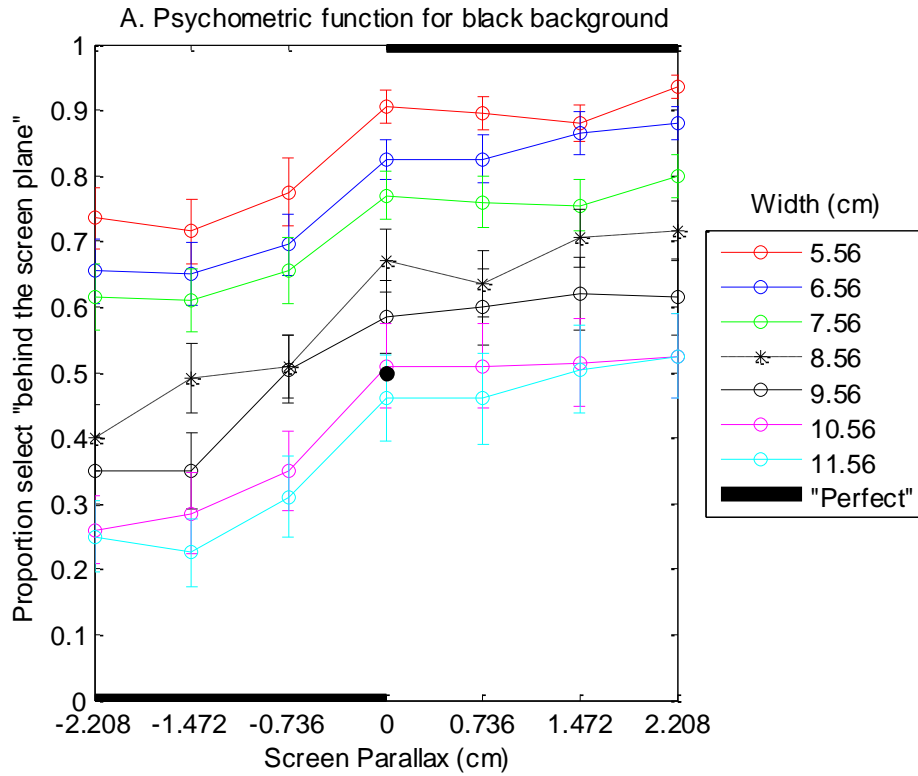
1813 **Table. 4.1.** Width W and corresponding parallax P parameters for the experiments. Parallax
 1814 is based on actual cm deviations away from one another on the monitor, and the signage is
 1815 based on typical convention (i.e. negative parallax values place the card in front of the screen
 1816 plane in depth compared to the participant). These 7 widths and 7 parallaxes were combined
 1817 for 49 different combinations in the absolute (black background) trials and the 7 widths were
 1818 combined with the zero and negative parallaxes for 28 different combinations in the relative
 1819 (noisy background) trials.

1820 4.2.5. Results

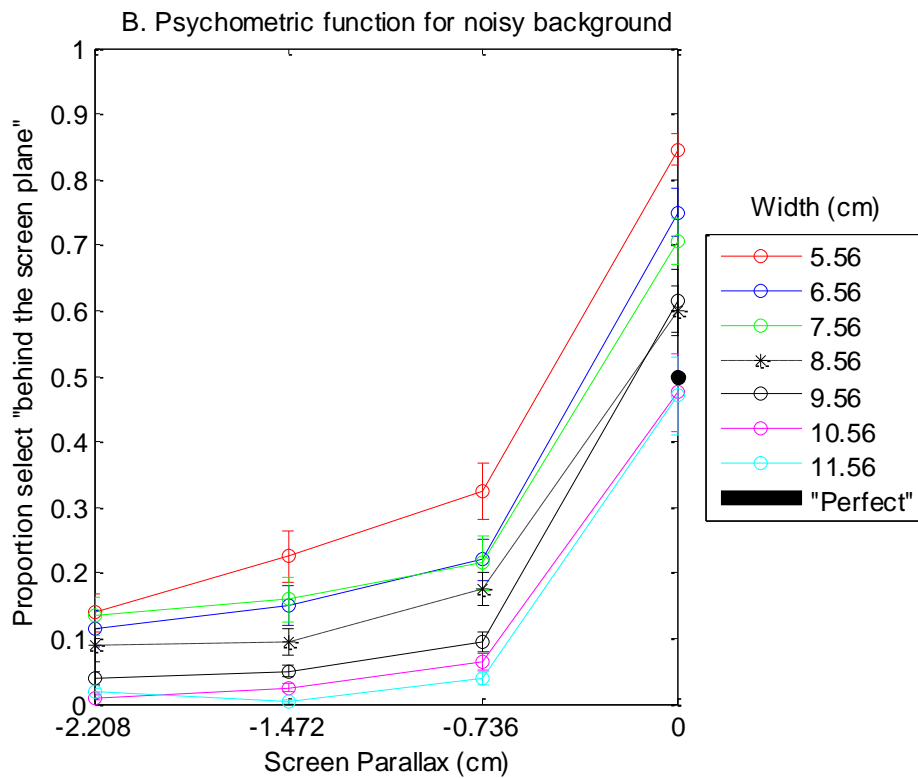
1821 Fig. 4.4. shows results for the absolute (A) and relative (B) disparity conditions. The
1822 proportion of “far” judgments is plotted against screen parallax in cm, with the different
1823 colored curves showing results for different card widths W . I assess the interaction between
1824 the two different depth cues (size and disparity) for both absolute and relative disparity. To
1825 do this I first consider what would happen in the ‘perfect’ scenario to help analyze the results.
1826 If the participants considered depth information from vergence as the only source of depth,
1827 ignoring familiar size cues of the image, then for negative parallax the participant should
1828 always choose ‘in front of’ the screen plane. At the zero disparity case (when the image
1829 displayed is on the screen plane) then participants should necessarily fall to chance in
1830 selecting either ‘in front of’ or ‘behind’ the screen plane, and finally for positive parallax the
1831 participants should always be perceiving the displayed stimulus as ‘behind’ the screen plane.
1832 This “perfect” performance is shown with heavy black lines in Fig 4.4.

1833 I evaluated statistical significance using a generalised estimating equation in SPSS, using
1834 inter-subject and global comparisons of the raw data with stimulus width, disparity, and type
1835 of disparity (relative or absolute, determined by the background) as variables. As expected,
1836 participants do use screen parallax to perform the task. Additionally, my results confirm the
1837 increased ability to distinguish depth with relative disparity compared to absolute disparity.
1838 In absolute disparity, the ability to distinguish the card being in front or behind the screen is
1839 significantly worse. On average across all non-zero parallax values, participants are correct
1840 on 60% of trials with a black background, but 89% with the noisy background, where
1841 “correct” is defined based on the sign of parallax. For example, consider the case where the
1842 card has the most negative parallax, -2.21cm, but has the smallest width, 5.56cm. Then,
1843 disparity suggests the card is ‘in front of’ the screen plane, but the familiar size cue indicates
1844 to the participant the card is ‘behind’ the screen, as it is smaller. With the noisy background
1845 Fig 4.4.B shows that participants were correct 84% of the time, while with the black

1846 background they were successful only 28% of the time. My statistical analyses confirmed this
1847 significant difference (paired T-test on n=9 different participant observations, $P = 0.001$).
1848 Participants also clearly base their depth judgment partly on the on-screen size of the credit
1849 card. That is, they have a tendency to report that smaller stimuli are further than the screen
1850 while larger stimuli are in front. This tendency is particularly strong for the black background
1851 condition, Fig. 4.4.A. In this condition, the stimulus was presented in isolation, so essentially
1852 only absolute disparity cues were available. The human visual system is particularly sensitive
1853 to relative-disparity cues between nearby objects, and is not very sensitive to absolute
1854 disparity. Accordingly, in this condition participants gave significantly more weighting to the
1855 depth cue implied by the size of the familiar object, Cards that were displayed smaller than
1856 the familiar size are deemed to be 'behind' the screen plane significantly more than cards
1857 displayed larger than the familiar size, regardless of the disparity that they were displayed
1858 at. However it is clear that disparity still makes some contribution to the decision about
1859 depth, or else the gradient of the different width lines would be flat, and each line would be
1860 a perfectly horizontal line, going down in proportion from red to cyan as the size increased
1861 from 5.56cm to 11.56cm.



1862



1863

1864 **Fig. 4.4.** Psychometric functions for black (A, absolute screen parallax) and noisy (B, relative

1865 screen parallax) backgrounds. The horizontal axis shows screen parallax, and each line type

1866 *is a different screen width of the displayed credit card stimulus. Also shown is the 'perfect'*
1867 *performance, if participants made judgments based solely on disparity depth information,*
1868 *shown with a thicker, black line.*

1869 However the story is almost the opposite for relative disparity (Fig 4.4.B). Here, it is apparent
1870 that disparity is the overriding depth cue, not the familiar size cue as for absolute disparity.
1871 Again it is clear that the familiar size cue is still factored into consideration, from the fact that
1872 the lines are separated. The separation is a lot less pronounced than it is in the absolute case
1873 (Fig. 4.4.A.), however. Furthermore, considering some of the more extreme experimental
1874 setups I can see that disparity is the driving depth cue in the relative disparity case. For
1875 example when the card is displayed in front of the screen (at any disparity) at 5.56 cm width,
1876 familiar size should be saying the card is displayed behind the screen plane, as it is much too
1877 small to be displayed in front. However the participants correctly judge the card to be
1878 displayed in front of the screen plane (based on the depth information from vergence; in this
1879 case the familiar size cue is considered incorrect) at least 68% of the time (for the smallest
1880 negative parallax of -0.736cm). This cannot just reflect the occlusion cue, i.e. the fact that
1881 the card occluded the textured background, because participants judged the card to be
1882 *behind* the screen plane when the parallax was zero.

1883 I calculated d_{prime} to consider the strength of the signals in the stimulus I was creating
1884 (assuming Gaussian statistics). I calculated a d_{prime} value based on "near" and "far"
1885 judgements for the black background case and calculated a value of d_{prime} of 0.5124.

1886 A limitation of the noisy-background condition is that no positive parallaxes were presented.
1887 This was because there was no unproblematic way of doing this. A stimulus with positive
1888 parallax must either be occluded by the zero-parallax background surface, or be seen
1889 through a transparent surface, or be seen through a hole cut in the surface. Fortunately,
1890 there does not appear to be any significant difference between the zero disparity cases for
1891 the black and the noisy backgrounds. In both cases, participants show a bias towards

1892 reporting that objects are further than the screen plane. This is a helpful indication that the
1893 noisy background only provided a reference surface for comparison and did not otherwise
1894 affect the results. As can be seen, once disparity information is removed (i.e. once the display
1895 reverts back to 2D, rather than S3D) the reliance on the familiar size cue increases, as
1896 displayed by the increased gradient of the slope in Fig.4.4.B. between the negative parallax
1897 value of -0.736cm and the 2D case of 0cm parallax.

1898 4.2.6. Discussion

1899 This experiment considers the interaction of the familiar size and vergence based
1900 depth cues when the cues either both suggest the stimuli is displayed correctly in front of or
1901 behind the screen, or the case where the different depth cues conflict with one another. I
1902 conducted the initial experiment to consider whether or not the structure of the main
1903 experiment would be a suitable setup, and also to determine how likely it would be that the
1904 main experiment would yield any informative results. I conclude that it would be worth
1905 conducting the full study based on the significant results gathered in this experiment.

1906 In this experiment, I asked participants to select whether the card is in front or behind the
1907 screen plane. This immediately privileges disparity information, since this relates specifically
1908 to depth as a matter of geometry, whereas the familiar size cue only implicitly relates to
1909 depth; participants know that images of familiar objects can be drawn larger or smaller than
1910 reality (Hudson, 1967). Even so, my results indicate that participants relied on the familiar
1911 size cue when the information available from disparity was limited. Based on this I believe it
1912 would be more suitable to change the question, to ask, for example, 'does the card appear
1913 bigger or smaller than it should', to determine also whether this would suggest a different
1914 interaction between the two cues to depth and size. It would also be useful if the weightings
1915 given to the different depth cues in each case could be quantified, via a modelling process,
1916 and hence in my full experiment I decided to do this.

1917 My experiment suggests that the structure I opted for is a sensible one for considering the
1918 cue conflict that can occur between familiar size cues and vergence based depth judgements.
1919 Based on the initial experiment I structure my main study as three separate experiments,
1920 described below.

1921 4.3 General methods

1922 4.3.1 Participants

1923 Work was approved by the Newcastle University Faculty of Medical Sciences Ethics
1924 Committee. Participants were recruited via an internal volunteer scheme at Newcastle
1925 University's Institute of Neuroscience, on the basis that they had no visual problems other
1926 than wearing glasses or contact lenses. There were three viewing distances that I conducted
1927 the main experiment at. Due to availability, no participants took part in all of the different
1928 viewing conditions. Seven subjects (M: 4, F: 3) participated in the experiments at 100cm.
1929 There were 10 (M: 2, F: 8) participants at 50cm, and 11 (M: 5, F: 6) participants at 200cm. All
1930 participants were initially naïve to the study but any that came back for different experiments
1931 and viewing distances will have known what the study was looking at. None of the
1932 participants from experiment 1 at 100cm repeated the experiment, due to availability. There
1933 were 8 participants (M: 2, F: 6) who took part in the other two viewing distances of the
1934 experiment, and these will no longer have been naïve to the task, although no explanation
1935 of what the study was aiming to discover was told to the participant until the full end of their
1936 participation (i.e. once the final experiment they were doing was completed). Participants
1937 were given a £10 gift voucher for every viewing distance they completed. Table 4.2 below
1938 shows the participants involvement in the experiment. A green square indicates that they
1939 completed the experiment at that distance, a red square indicates they did not complete the
1940 experiment at that distance

Subject ID	Gender (M/F)	Completed at 50cm?	Completed at 100cm?	Completed at 200cm?
1	M	Green	Red	Green
2	F	Green	Red	Red
3	M	Green	Red	Green
4	F	Green	Red	Green
5	F	Green	Red	Red
6	F	Green	Red	Green
7	F	Green	Red	Green
8	F	Green	Red	Green
9	F	Green	Red	Green
10	F	Green	Red	Green
11	M	Red	Green	Red
12	M	Red	Green	Red
13	F	Red	Green	Red
14	M	Red	Green	Red
15	F	Red	Green	Red
16	F	Red	Green	Red
17	M	Red	Green	Red
18	M	Red	Red	Green
19	M	Red	Red	Green
20	M	Red	Red	Green

1941 **Table 4.2.** Participation information on subjects. Green squares indicate the subject in
1942 question completed the experiment at that distance, a red square indicates they did now. The
1943 sex of the participants is also shown, and the subject ID was assigned to them chronologically.

1944 4.3.2 Equipment

1945 The same equipment was used in the main experiment as had been used in the initial
1946 experiment.

1947 4.3.3 Procedure

1948 In the main experiment the subjects' objective was to select whether the image
1949 presented appeared 'bigger' or 'smaller' than the reference card provided. Once a decision
1950 had been made, the subject clicked the up arrow key or down arrow key for bigger or smaller
1951 respectively. This was done at viewing distances of 50cm, 100cm and 200cm. I call these
1952 experiments the size/vergence interaction experiment(s). After the subject selected their
1953 answer using the keyboard, the next image would appear instantly. Between each image the
1954 subject was encouraged to look down at the reference card. Note that the question is
1955 different to that asked in the initial experiment (Hands, Khushu, & Read, 2014), and asks the
1956 question recommended by the experiment discussion.

1957 In addition to the main experimental setups, each cue was analysed separately with further
1958 experiments. For size, the subject was presented images that changed in width and height
1959 but had a constant disparity of zero (i.e. flat on the screen). The subject still wore the S3D
1960 passive glasses to maintain consistency and once again was asked to decide whether the
1961 image was bigger or smaller than the reference card. I call this experiment the size
1962 judgements in 2D experiment. For vergence, the size of the card remained the same as the
1963 reference (8.56cm width, 5.398cm height) whilst the parallax changed. The subject was
1964 required to decide whether the card was appearing in front of the screen (negative parallax)
1965 or behind the screen (positive parallax), using the keyboard. I refer to this as the vergence-
1966 based depth judgement experiment. The experiments were conducted in the order of: main
1967 experiment, size judgements in 2D experiment, vergence based depth judgement
1968 experiment. There is potential for practice effects as all participants completed the

1969 experiment in this manner, which I consider more in section 4.8 (discussion). This data,
1970 combined with my data for the 'chance' instances in the experiment, will help us in the
1971 mathematical modelling in assessing any bias that occurred on an individual basis.

1972 4.3.4 Parameters

1973 The parameters used in the main experiment consisted of the same parameters
1974 used in the initial experiment. A subset of the parameters were used in the size, and
1975 vergence based depth judgement experiments, as explained above in section 4.3.3
1976 (procedure)

1977 4.3.5 Mathematical modelling

1978 Fig. 4.1. above shows the different letters in my mathematical calculations that are
1979 needed to be considered. These are used below in the equations to explain my calculation
1980 of the parallaxes used and also for the modelling. I can use this to generate some conclusions
1981 on key relationships:

1982 The relationship between the ratios of perceived virtual depth and size and physical viewing
1983 distance and width is:

1984 Eq. 4.2.
$$\frac{W}{V} = \frac{S}{D}$$

1985 Using similar triangles I can also see that Parallax P can be equated to:

1986 Eq. 4.3.
$$P = I \left(1 - \frac{V}{D} \right) = I \left(1 - \frac{W}{S} \right)$$

1987 Combining these two I get the relationship that:

1988 Eq. 4.4.
$$S = \frac{IW}{I-P}$$

1989 To quantify my data I developed a model based on signal detection theory (Ernst & Banks,
1990 2002; Green & Swets, 1966) which assumes the perceived depth of familiar object relies on

1991 two competing mechanisms: That of the perceived size of the object (i.e. a size signal), and
1992 the binocular vergence to correctly assess the depth of the object (i.e. a parallax signal). I use
1993 modelling based on log values, as it allows for easier mathematical calculations in
1994 combinations, and is justified by Fechner's law. Hence:

1995 Eq. 4.4.
$$\ln(S) = \ln(I) + \ln(W) - \ln(I - P)$$

1996 So the perceived size of the card (which is the value I wish to measure, S) can be considered
1997 by looking at the interaction of I, W and P. In my experiment interocular distance (I) is a
1998 constant on average of 6.3cm (Dodgson, 2004) so I can consider the combination of two
1999 distinct signals, that from the familiar width information and the parallax information. I
2000 assume that these two signals are both centred on the true value of the displayed parameter
2001 subject to a bias, θ , and the correctness of the depth perception decreases subject to
2002 Gaussian noise (see Figs. 4.5. and 4.6.). Hence I consider the two signals:

2003 Eq. 4.5.
$$\omega \sim N(\ln(W) + \theta_w, \sigma_w)$$

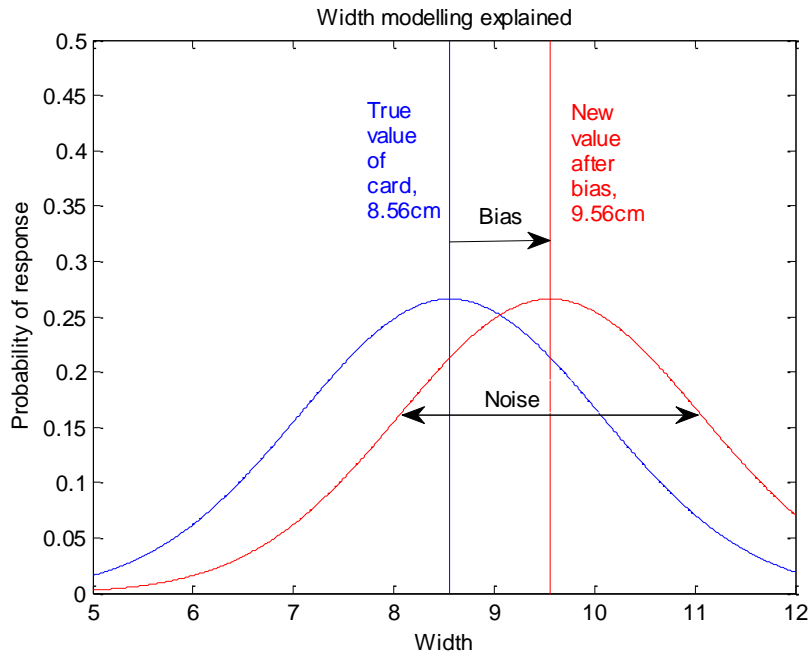
2004 Eq. 4.6.
$$\rho \sim N(\ln(I) - \ln(I - P) + \theta_p, \sigma_p)$$

2005 I then combine these two signals in a weighted sum to get:

2006 Eq. 4.7.
$$S = g_w \omega + g_p \rho$$

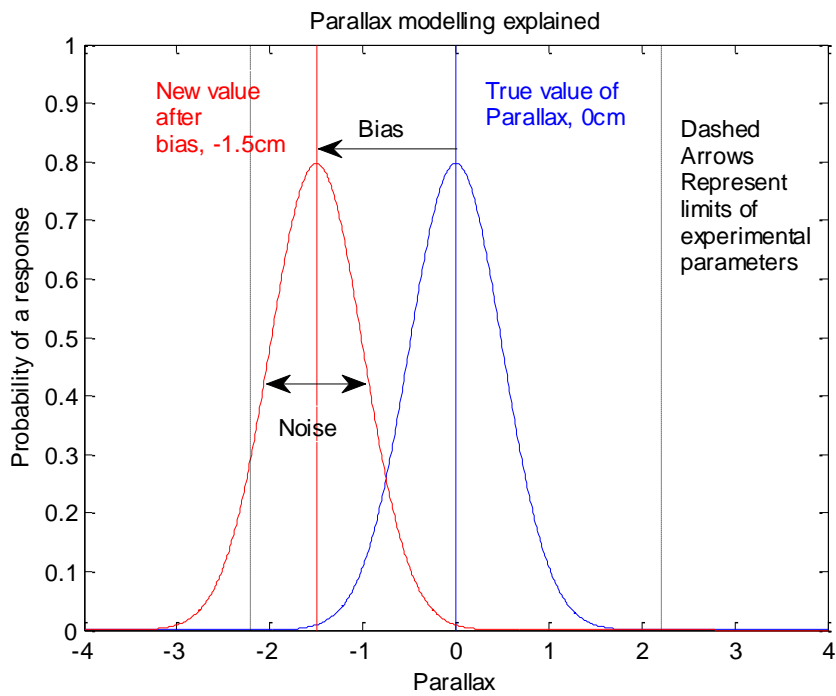
2007 And the value of S determines whether the model returns a value of 'too big', or 'too small',
2008 based on 6 free parameters: the respective bias values, θ_w , θ_p , noise values σ_w , σ_p , and weight
2009 values g_w and g_p . How I work with these two signals varies in each separate experiment and
2010 is explained further below. I opted to include a lapse rate of 5%, to better improve my fits.
2011 All three experiments were fitted using the same model to allow us to effectively compare
2012 across experiments. I use the Matlab code FMINSEARCH to determine the best fit of my
2013 model.

2014



2015

2016 **Fig. 4.5.** How the bias and noise contribute to the signal detection for the familiar size cue. In
 2017 this case the physically displayed width W is equal to the 'true' value of 8.56cm. The size bias
 2018 is a value of +1.0cm in this example and the noise is a value of 1.5cm. This means in this case
 2019 my size signal would be modelled as: $W \sim N(\log(9.56), 1.5)$.



2020

2021 **Fig. 4.6.** *The parallax signal is slightly more complicated than the size signal model but the*
2022 *premise is very similar. In this example the parallax shown is 0cm, but the bias moves the*
2023 *signal in front of the screen plane, and in this example the noise is considerably less so the*
2024 *peak is higher. Here the parallax signal is modelled (due to true value $P = 0\text{cm}$) as:*

2025 $P \sim N(-1.5, 0.5)$

2026 It is important to note that in experiments conducted with similar modelling strategies as
2027 that above - such as Ernst & Banks (2002) - the different cues could be separated completely
2028 (i.e. only haptic cues, no visual; only visual, no haptic and a combination of both haptic and
2029 visual cues). In my experiment it isn't possible to measure the cues independently of one
2030 another, as both cues are visual, and changing one cue has an indirect impact on the other.
2031 Due to this, some assumptions have been made that the standard weighted sum model
2032 applies to our experimental setup, and that by constraining one of the cues to depth and size
2033 to be a constant, rather than allowing it to vary, I isolate the cues to depth as well as possible,
2034 as any changes in perception should be driven solely by the change in the other perceptual
2035 cue.

2036 4.4 Experiment 1: Vergence-based depth judgments

2037 The initial aim of experiment 1 was to attempt to measure participants' sensitivity
2038 for depth – essentially take a measurement of their stereoacuity – and consider the value of
2039 the bias associated with participants' depth judgements. The two parameters associated
2040 with the parallax signal (θ_p and σ_p) are the important ones here and are considered in the
2041 fitting to the data.

2042 4.4.1 Methods

2043 In experiment 1 the methods are exactly as in the general methods, as the two
2044 backgrounds were used, and the parallaxes associated with them. In this experiment
2045 however the displayed width was set to be the true displayed value of 8.56cm, and did not

2046 change. Thus the experiment was a shorter one than experiment 3, with a total of 220 trials
2047 in the experiment (1 width x (7 absolute parallaxes + 4 relative parallaxes) x 20 repeats). In
2048 this experiment the question asked to participants was slightly different also; participants
2049 were asked to answer if the displayed card was 'in front of or behind the screen plane', as
2050 they had done for the initial experiment. This was because I was in effect attempting to
2051 establish a value for the participants' stereoacuity.

2052 4.4.2 Model

2053 In the model for experiment 1, I consider the condition where the weight assigned
2054 to the width signal, g_w , is zero. Thus the equation collapses to a single Gaussian model. Hence
2055 I can solve this analytically to get the probability of saying 'in front' of the screen plane is:

2056 Eq. 4.9.

$$\begin{aligned} 2057 \quad P(\text{infront}) &= \frac{1}{\sigma_p \sqrt{2\pi}} \int_{-\infty}^0 dp \exp\left(-\frac{(p - \ln I + \ln(I - P) - \theta_p)^2}{2\sigma_p^2}\right) \\ 2058 \quad &= \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{-\ln I + \ln(I - P) - \theta_p}{\sigma_p \sqrt{2}}\right)\right) \end{aligned}$$

2059 I then fit the model to my data using FMINSEARCH to get the best results.

2060 4.5 Experiment 2: Size judgments in 2D

2061 The initial aim of experiment 2 was to attempt to measure participants' sensitivity
2062 for correct familiar size and to consider the bias and noise of the size signal. The two
2063 parameters associated with the width signal (θ_w and σ_w) are the important ones here and
2064 are considered in the fitting to the data.

2065 4.5.1 Methods

2066 In experiment 2, the methods are exactly as in the general methods, as the two
2067 backgrounds were used, and the full range of the widths of the displayed cards. In this

2068 experiment however the displayed parallax was set to always be a value of 0cm, hence in the
2069 screen plane, and did not change. Thus the experiment was a shorter one than experiment
2070 3, with a total of 280 trials in the experiment (1 parallax x 7 widths each for both black and
2071 noisy backgrounds x 20 repeats). In this experiment the question was as it was in the general
2072 methods.

2073 4.5.2 Model

2074 In the model for experiment 2 I consider the condition where the weight assigned to
2075 the parallax signal, g_p , is zero. Thus the equation collapses to a single Gaussian model. Hence
2076 I can solve this analytically to get the probability of saying 'too big' when compared to a
2077 standard credit card:

2078 Eq.4.10.

$$2079 \quad P(\text{too big}) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{\ln W + \theta_w - \ln S_{cc}}{\sigma_w \sqrt{2}} \right) \right)$$

2080 where S_{cc} is the correct size of the credit card (8.56cm). With this formula it can be seen that
2081 if the width of the displayed card is the correct size of 8.56 and the bias, θ_w is zero, the entire
2082 equation falls to chance (0.5), which is the result I would want.

2083 I then fit the model to my data using FMINSEARCH to get the best results.

2084 4.6 Experiment 3: Size/vergence interaction

2085 The original aim of this study was to consider the interaction between the two pieces
2086 of depth information when they were conflicting with one another. Experiment 3 is the
2087 experiment that really considers the depth cue combination problem. Experiments 1 and 2
2088 could be considered supplementary to this one, and the models for experiments 1 and 2 are
2089 smaller cases of the experiment 3 model (where one of the weights was considered to be
2090 zero). I believe that the model fits for experiments 1 and 2 will help us to justify my

2091 experiment 3 model. All 6 parameters are considered in the fitting here (θ_p , θ_w , σ_p , σ_w , g_p ,
 2092 and g_w).

2093

2094 4.6.1 Methods

2095 In experiment 2 the methods are exactly as in the general methods, as the two
 2096 backgrounds were used, and the full range of the widths and parallaxes of the displayed cards
 2097 for each background. Thus there were 1540 trials in each experiment 3 that the participants
 2098 took part in (49 absolute combinations of size and parallax + 28 relative combinations of size
 2099 and parallax x 20 repetitions).

2100 4.6.2 Model

2101 The model that I consider for the experiment 3 combines both of the original 1D
 2102 models into a 2D model. Here I integrate Eq. 4.7. to get:

2103 Eq. 4.11.

2104 $P(\text{too big})$

$$2105 = \frac{1}{\sigma_p \sigma_w 2\pi} \int_0^\infty dw \exp\left(-\frac{(w - \ln W - \theta_w)^2}{2\sigma_w^2}\right) \int_{(\ln S_{cc} - g_w w)/g_p}^\infty dp \exp\left(-\frac{(p - \ln I + \ln(I - P) - \theta_p)^2}{2\sigma_p^2}\right)$$

$$2106 \quad P(\text{too big}) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{g_w(\ln W + \theta_w) + g_p(\ln I - \ln(I - P) + \theta_p) - \ln S_{cc}}{\sqrt{2(g_w^2 \sigma_w^2 + g_p^2 \sigma_p^2)}} \right) \right)$$

2107 Here S_{cc} is again the size of a standard credit card (8.56cm), and the other parameters are as
 2108 defined in the general methods.

2109 4.7 Results

2110 I used an omnibus ANOVA in matlab to assess significance of factors and interactions
 2111 between them for my experimental results at all viewing distances, pooled across

2112 participants. These results are shown in table 4.3 below, where I report the viewing condition
 2113 (black or noisy) and then consider different factors (experiment number, virtual size of credit
 2114 card, parallax the card was displayed at and the viewing distance of the participant) as well
 2115 as the respective 2-way interactions. I report the F-statistics and also the respective P-values,
 2116 as well as the DFN and DFD (degrees of freedom in the numerator and denominator
 2117 respectively).

Background	Factor/interaction	F	DFN	DFD	P-value
Black	Experiment	0.01	2	1	0.9899
	Width	170.9	6	1	0
	Parallax	5.57	6	1	0
	Viewing distance	80.23	2	1	0
	Experiment*Width	0.42	12	2	0.8672
	Experiment *Parallax	102.67	12	2	0
	Experiment *Viewing distance	6.85	4	1	0.0001
	Width*Parallax	7.92	36	1	0
	Width*Viewing distance	7.12	12	1	0
	Parallax*Viewing distance	3.31	12	1	0.0005
	Noisy	Experiment	44.99	2	1
Width		89.65	6	1	0
Parallax		4.86	3	1	0.0033
Viewing distance		35.27	2	1	0
Experiment*Width		0.62	12	2	0.7098

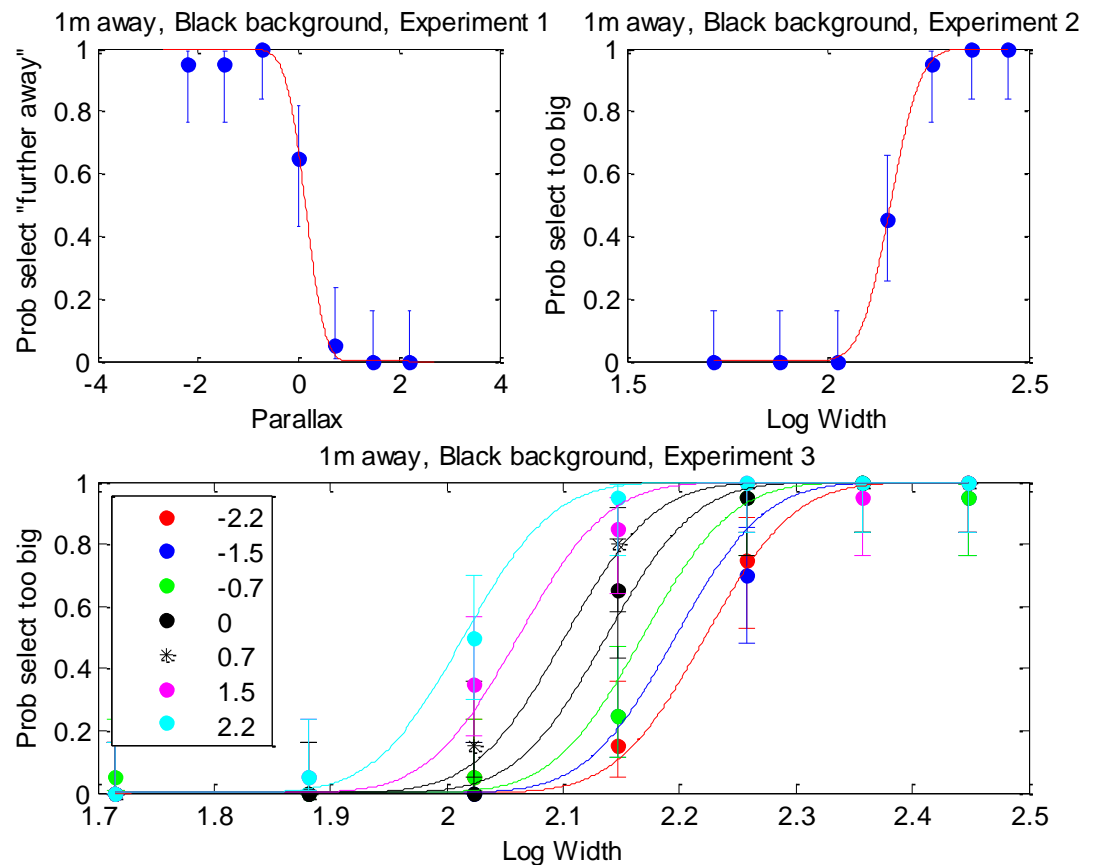
	Experiment *Parallax	184.27	6	2	0
	Experiment *Viewing distance	7.55	4	1	0.0001
	Width*Parallax	16.33	18	1	0
	Width*Viewing distance	9.06	12	1	0
	Parallax*Viewing distance	0.68	6	1	0.6646

2118 **Table 4.3.** Summary of factors and interactions in an ANOVA analysis of my results for black
2119 and noisy background respectively. Non-significant results are highlighted in red. Significant
2120 factors or interactions are highlighted in green. Table reports back F values, the degrees of
2121 freedom numerator and denominator, and the respective p-values.

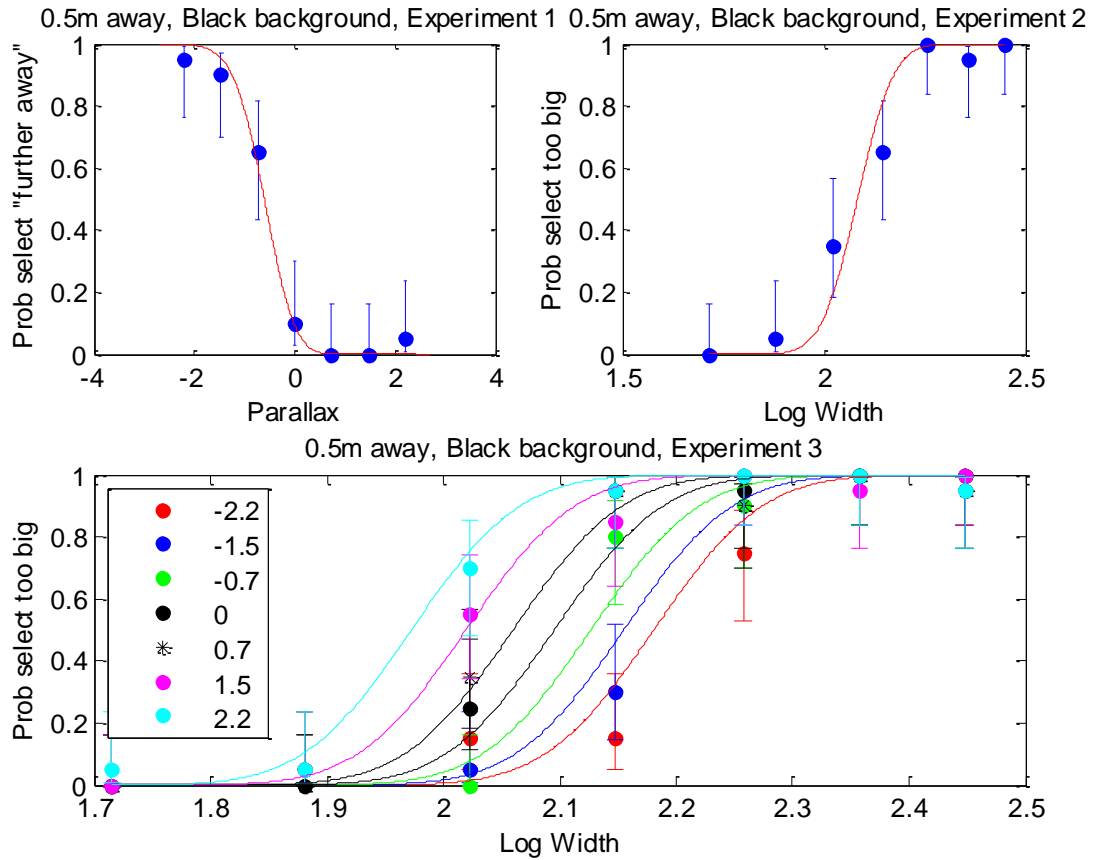
2122 As can be seen the displayed width, the displayed parallax and the viewing distance the
2123 participant was sat at are significant factors in results for both noisy and black backgrounds.
2124 Many of the interactions were also significant. This implies that changing the size of the
2125 displayed card, changing the parallax the card is displayed at and changing the distance the
2126 participant sees the card from all have a significant effect on how the card is perceived.

2127 I consider the results and fitted models in Figs. 4.7 to 4.11. These are fits from different
2128 participants in either the black or noisy background condition. These are typical of the
2129 population of fits and patterns. In doing this I try to establish whether my model with 6 free
2130 parameters is effective, and what the underlying trends might be for the experimental
2131 results. As can be seen from Figs. 4.7. to 4.11. my model fits the data well, and captures the
2132 pattern of the 63 different datapoints (absolute disparity) or 42 datapoints (relative
2133 disparity) with only 6 free parameters. Figures show individual participants data and is not
2134 averaged or pooled across participants here, as the results were all very similar. Participants

2135 are more likely to say the virtual card displayed is 'too big' more as the vergence depth
 2136 information moves the card further away from the participant (i.e. from negative to positive
 2137 parallax) for any given size of the virtual card. Participants also consider the card 'too big'
 2138 more often as the virtual size increases, regardless of disparity. This trend is exhibited in all
 2139 three experimental setups (where applicable).



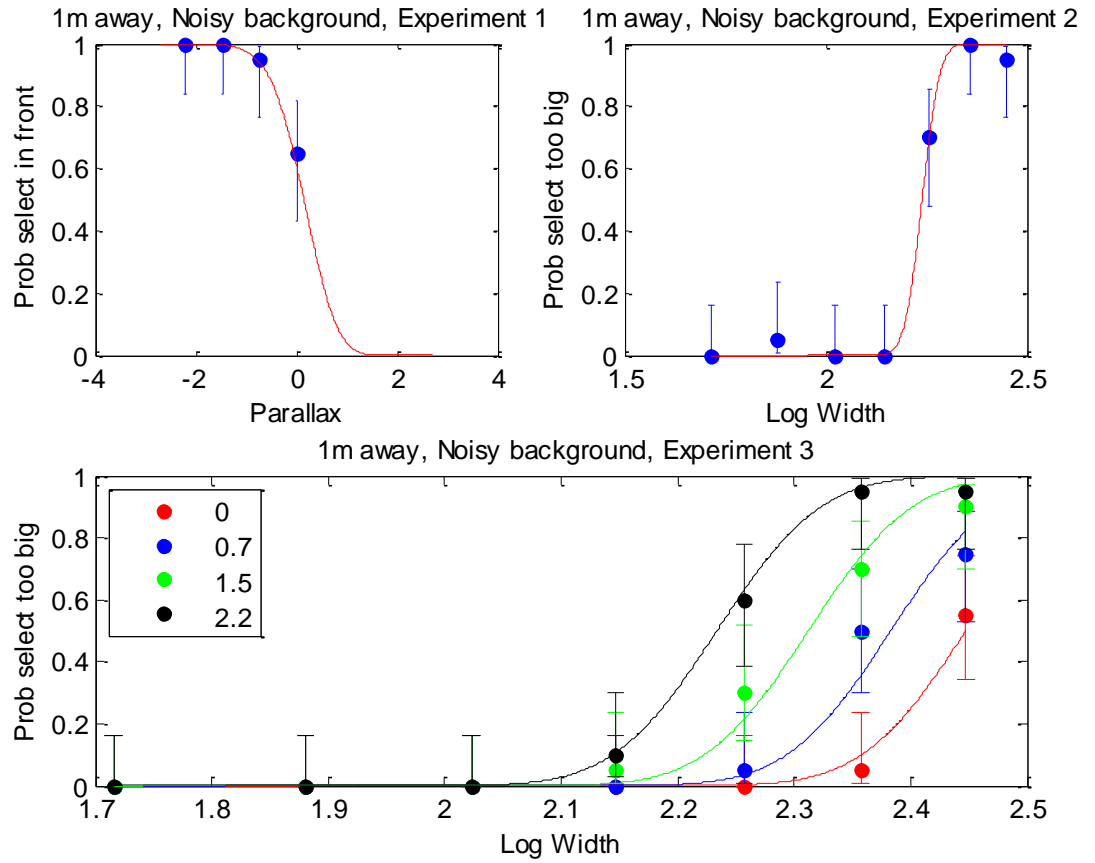
2140
 2141 **Fig 4.7.** Data and model fits for all three experiments for one participant at a viewing distance
 2142 of 1m with absolute disparity (black background). Each different colour represents a different
 2143 parallax value. The x axis represents the parallax values (experiment 1) or the log of the width
 2144 of the virtual credit card and the y axis is the probability of answering the question as either
 2145 'in front' (experiment 1) or 'too big' (experiments 2 and 3). Each data point with confidence
 2146 intervals represents $n = 20$ trials.



2147

2148 **Fig 4.8.** Data and model fits for all three experiments for one participant at a viewing distance

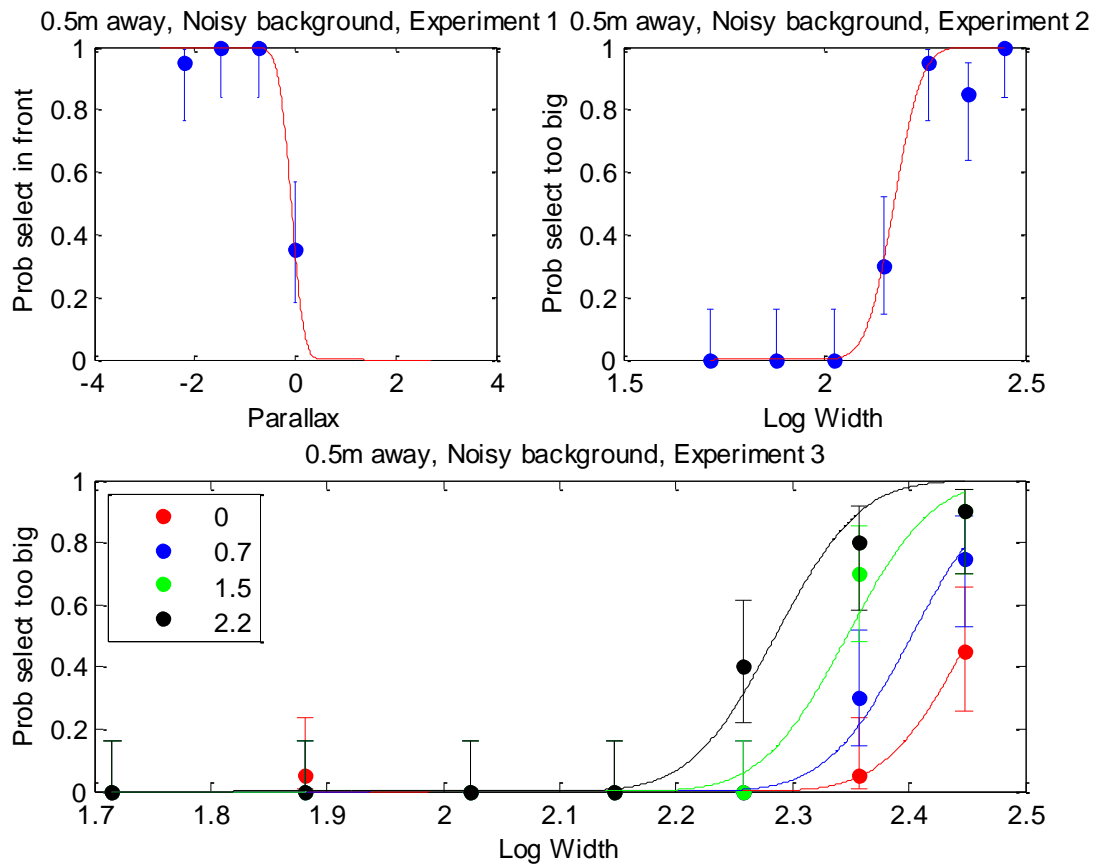
2149 of 0.5m with absolute disparity (black background). All other information as in Fig. 4.7.



2150

2151 **Fig 4.9.** Data and model fits for all three experiments for one participant at a viewing distance

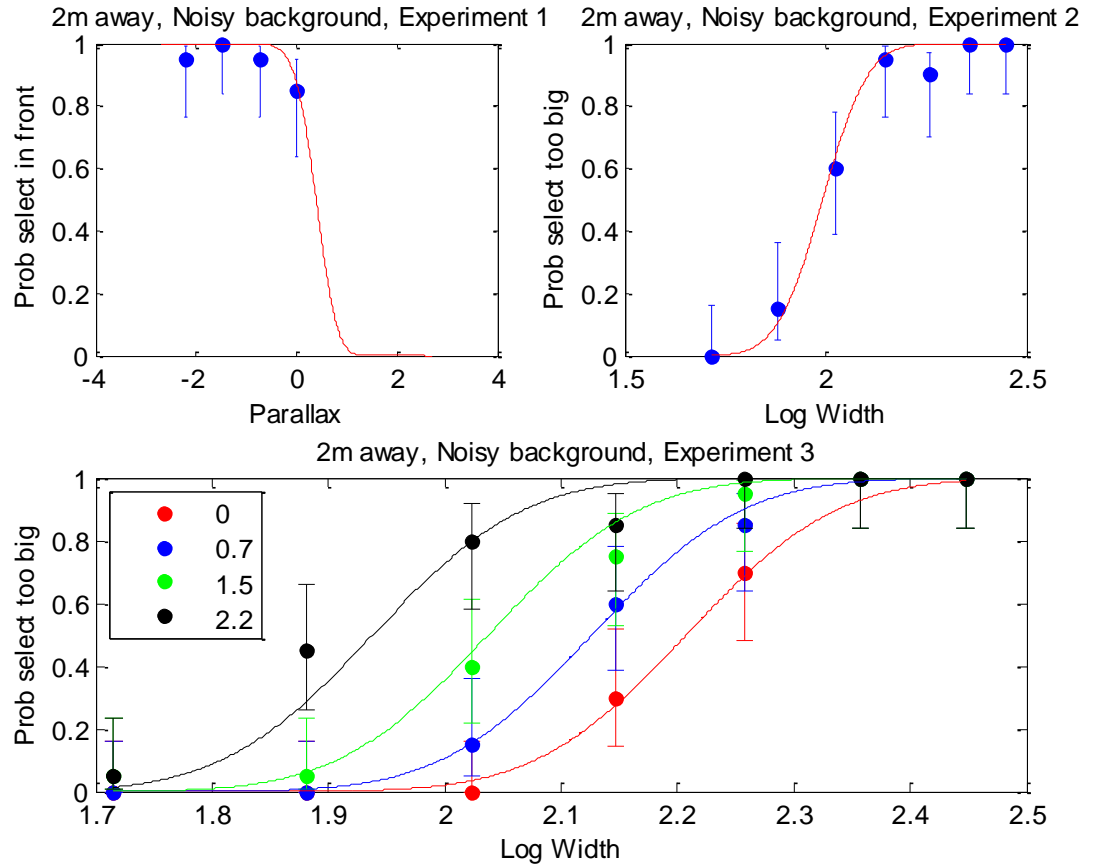
2152 of 1m for relative disparity (noisy background). All other information as in Fig. 4.7.



2153

2154 **Fig 4.10.** Data and model fits for all three experiments for one participant at a viewing

2155 distance of 0.5m for relative disparity (noisy background). All other information as in Fig. 4.7.



2156

2157 **Fig 4.11.** Data and model fits for all three experiments for one participant at a viewing
 2158 distance of 2m for relative disparity (noisy background). All other information as in Fig. 4.7.

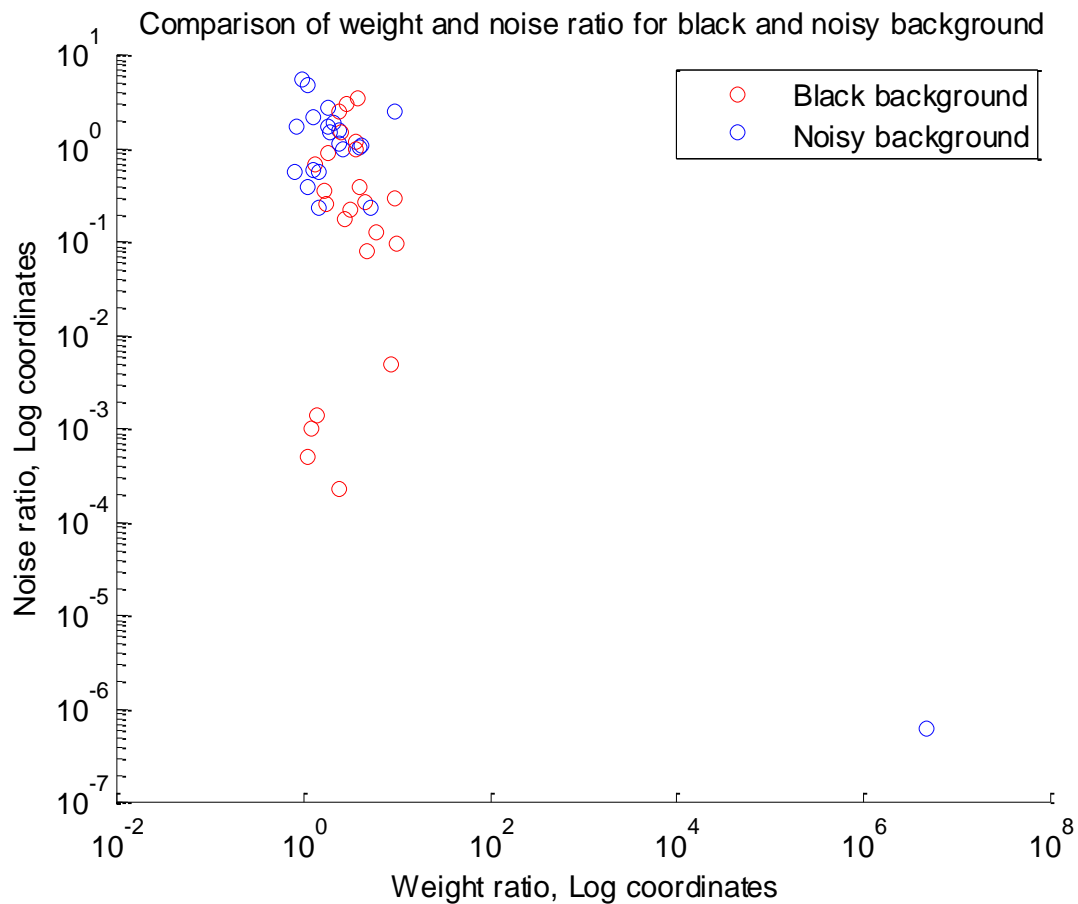
2159 I consider whether I could constrain the weights of the respective signal cues to depth (size
 2160 and disparity) by the different noise values that each signal had in my model. This would
 2161 reduce my number of free parameters to 4 instead of 6, and would therefore be a more
 2162 robust model. The constraint I wanted to apply to the weights was the following:

2163 Eq. 4.12.

2164
$$G_W^2 + G_P^2 = 1 \text{ and } \frac{G_W}{G_P} = \frac{\sigma_P}{\sigma_W}$$

2165 This suggests that the weights and the noises are inversely proportional. This would be the
 2166 optimal way to integrate the two different depth cues, as seen in previous studies (Ernst &
 2167 Banks, 2002). To consider if this was an applicable constraint I first considered the

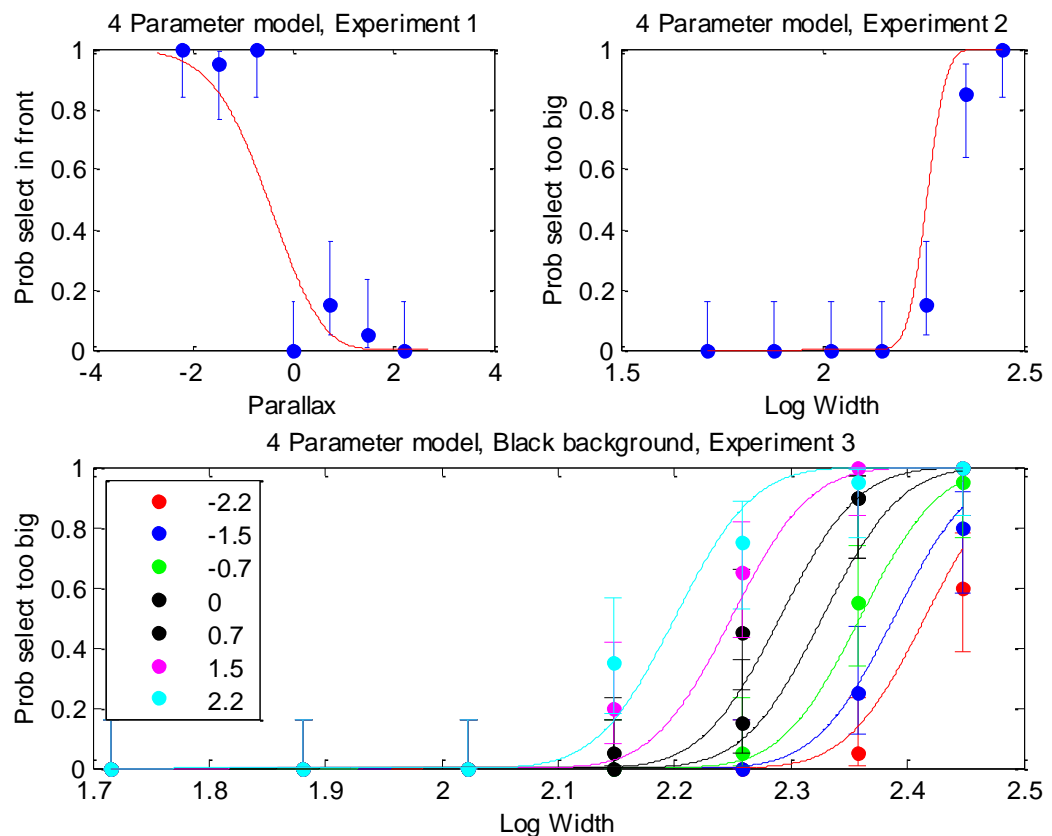
2168 relationship between the noise ratio and the weight ratio for black and noisy backgrounds
 2169 respectively for the size/vergence interaction experiment, in log axes (Fig. 4.12. below). The
 2170 ratio considers the weight (or noise) of the size signal in comparison to the weight (or noise)
 2171 of the parallax signal.



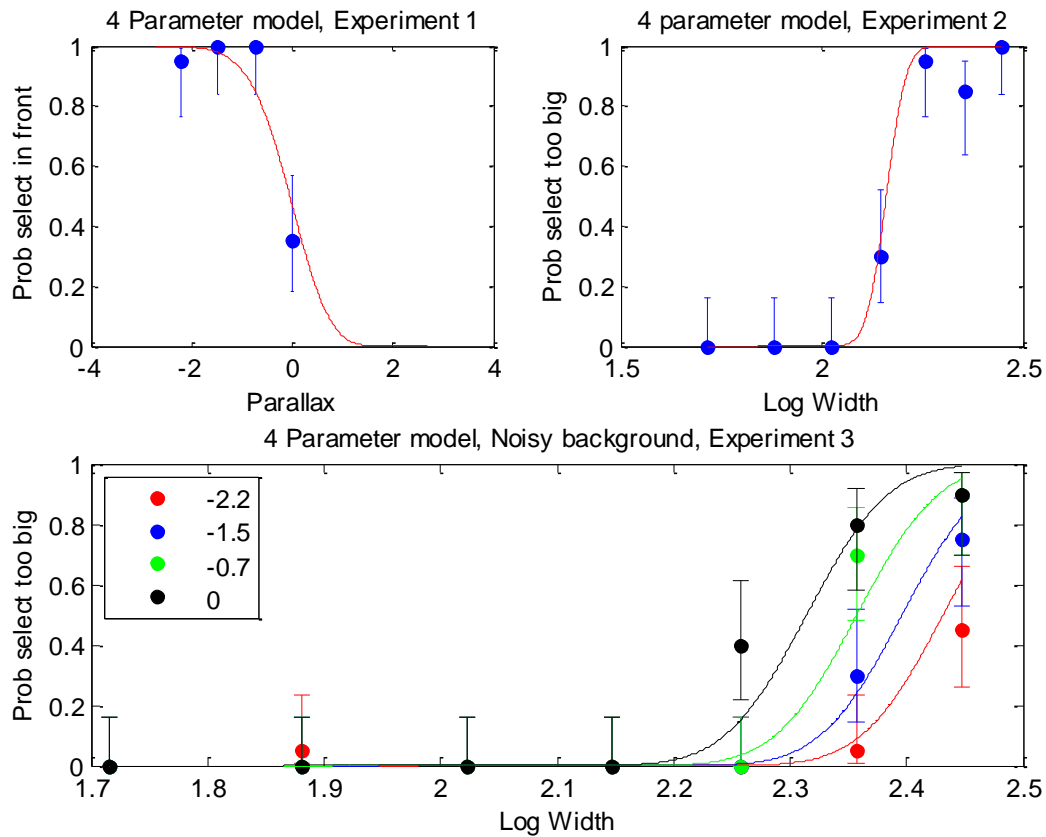
2172
 2173 **Fig. 4.12.** Comparison of the weight and noise ratios for black background (red circles) and
 2174 noisy background (blue circles) respectively. As can be seen there are some extreme outliers,
 2175 but the relationship appears to be negatively correlated, as would fit with previous studies.

2176 As can be seen, the relationship between the two weight and noise ratios appears to be
 2177 negative. The correlation values of all the data return -0.1314 and -0.2395 for absolute (black
 2178 background) data and relative (noisy background) data respectively. This gives us an initial
 2179 indication that the model works as well with 4 parameters as 6, as the 4 parameter model
 2180 has a correlation of -1.

2181 I also consider the log likelihood ratio test on my 4 and 6 parameter models (Wilks, 1938).
 2182 This gives us an indication further of whether the 4 parameter model is a viable robust
 2183 alternative. I find that in all but one case for each background (a different participant in each
 2184 case, both at viewing distance of 2m) there is no significant benefit to using the 6 parameter
 2185 model over the 4 parameter model, and hence conclude that the relationship between the
 2186 noise of the signals in my model is inversely proportional to the weight assigned to the signal
 2187 for depth. This agrees with previous literature and also represents my data well, suggesting
 2188 I have found an optimal cue combination technique when considering two conflicting cues
 2189 to depth (familiar size and vergence depth information).



2190
 2191 **Fig. 4.13.** Data and 4 parameter model fits for all three experiments for one participant at a
 2192 viewing distance of 2m for absolute disparity (black background). In this model the weights
 2193 assigned to the two signals are calculated using Eq. 4.12 above. All other information as in
 2194 Fig.4.7.



2195

2196 **Fig. 4.14** Data and 4 parameter model fits for all three experiments for one participant at a
 2197 viewing distance of 0.5m for relative disparity (noisy background). In this model the weights
 2198 assigned to the two signals are calculated using Eq. 4.12 above. All other information as in
 2199 Fig. 4.13.

2200 Considering the model fits for the pooled data, the weights assigned to the size cue, G_w is
 2201 larger than the weight assigned to the vergence cue, G_p (0.9840 for G_w and 0.1782 for G_p in
 2202 the 1m case, 0.9453 and 0.3262 in the 0.5m case and 0.9565 and 0.2917 in the 2m case) for
 2203 absolute disparity (black background). The result of this is that the familiar size cue is the
 2204 more relied upon cue when depth information from familiar size and vergence depth
 2205 information conflict with one another. Participants also reported back that they preferred to
 2206 consider the size before the stereoscopic information, which further supports this.

2207 Participants reported finding relative depth judgements (i.e. the noisy background case) easy
 2208 to consider the depth. This is not new information but backs up what I already know about

2209 absolute and relative disparity information in that humans are better at relative depth
2210 judgements than absolute. This is also backed up by my model results, where the weights
2211 were more equally considered. This suggests the noise from the vergence signal was more
2212 reliable.

2213 I found no significant difference between the three viewing distances in this experiment,
2214 with all pooled models similar to each other.

2215 [4.8 Discussion](#)

2216 The inverse relationship between the noise associated with a signal and the weight
2217 assigned to that signal is concurrent with other studies that consider optimal cue
2218 combinations. It makes sense that if the human visual system considers one cue to depth
2219 (such as a familiar size cue) to be a lot more reliable than another (such as the vergence
2220 depth information) then the more reliable cue is the one the visual system applies more
2221 consideration to, and hence reflects the higher weight assigned to it in the combination of
2222 the information from the cues.

2223 This study found that the familiar size cue to depth was more heavily weighted (and hence
2224 had less noise associated with its signal). In the case of 2D content, it could be that exposure
2225 to images with things the wrong size (i.e. movie screens, television shows, pictures in books
2226 and on walls) forces the human visual system to become very adept at discerning depth
2227 based on the relative size of the familiar object. Indeed from a young age humans are
2228 exposed to many pictures in books, in frames and on devices with screens such as televisions,
2229 tablets and more recently smart phones which take up a disproportionate amount of space
2230 on the retina for the known familiar size of that object. Humans conversely may develop this
2231 ability to understand that the retinal space taken up is directly linked to the depth of the
2232 object, and not the size of it, due to many exposures of this type. However it is much rarer
2233 for humans (particularly in their developmental years) to experience a similar level of

2234 exposure to S3D technology. It could be this imbalance that causes humans to depend more
2235 heavily (and make more reliable judgements) on the familiar size cue to depth than the
2236 stereoscopic cue that comes from vergence information.

2237 This study found no significant difference between viewing distance and the cue to depth
2238 that was relied upon. This suggests that despite the accommodation vergence conflict effects
2239 of S3D (which I would expect to be somewhat lessened at the largest viewing distance of
2240 200cm) familiar size is still the stronger signal to depth. It would be interesting to consider
2241 distances of greater than 200cm to see if the pattern continues.

2242 This study used a credit card as its stimulus to remove internal depth cues such as roundness.
2243 However a credit card is a small familiar object and even at a distance of 50cm participants
2244 reported that they believed the card looked considerably smaller than standard, even when
2245 it was the correct size on the monitor, particularly in the absolute case, where the rest of the
2246 50inch monitor was completely black. It would be interesting to examine whether the
2247 relationship between the size of the object and the size of the monitor had any impact on
2248 the perception of the object at different viewing distances.

2249 This study confirmed what was already understood in that relative depth judgements are
2250 easier to make than absolute depth judgements (based on participant feedback and the
2251 data). While this is not surprising it does suggest that the experiment was set up correctly in
2252 terms of judging changes in vergence and disparity.

2253 The conflicts between two depth cues can become large enough to become robust, which
2254 can lead to one cue to depth and size dominating over the other one, although not to a point
2255 of complete ignorance (Hillis et al., 2004). This can be seen in the case of football players and
2256 the Subbuteo effect, where players can be seen as being very small on an enormous pitch in
2257 S3D displays. In this case, the discrepancy between the size cue to depth and the disparity

2258 cue to depth are in such conflict that the image warps. When the cues don't conflict quite as
2259 robustly the effect is not observed.

2260 A potential issue that this study faces is that of a question bias. If the study asks to make a
2261 judgement based on size ('is the card bigger or smaller than a standard credit card'), do the
2262 participants rely more heavily on the familiar size cue because the question considers size
2263 and not depth? This is not the case, as the initial experiment considered the same
2264 experimental setup but with the question changed to one associated with depth ('is the card
2265 in front or behind the screen plane?'). In this case the familiar size cue is still more heavily
2266 relied on to make judgements (Hands et al., 2014).

2267 The perception of depth in the image that the participant has is influenced by many different
2268 factors. Quite a few of these, while watching S3D content, suggest to the viewer that the
2269 true depth in the image is in fact at the screen plane. This causes a problem because the
2270 entire premise of S3D is based on using disparity in the image to generate depth and move
2271 an object away from the screen plane. Some cues to depth could be as simple as factors that
2272 allow the viewer to more correctly assess the depth position of the screen, and disrupt the
2273 disparity information to depth, such as light reflectance from the flat screen image, the
2274 lighting levels in the room or even small details such as dust on the frame and glass of the
2275 screen. Another important cue related to the flatness of the screen, and lack of literal depth
2276 in the image, is that of accommodation. The accommodation-vergence conflict is a well-
2277 documented issue with S3D displays. The focus cues to depth generated from viewing the
2278 image are necessarily all from the same depth on a conventional display (Watt et al., 2005).
2279 In my study, particularly in the vergence-based depth judgements experiment, this cue to
2280 depth from accommodation could pull preference from the participant to estimate depth
2281 closer to the screen than the disparity information would suggest. In the vergence-based
2282 depth judgements, the question was asked to judge absolutely: 'in front' or 'behind' the

2283 screen plane. Hence even if the accommodation cue to depth suggested a depth closer to
2284 the screen plane, the result would still be the same answer given.

2285 The order of the experiments was always conducted in the same order for all participants:
2286 main experiment, size judgements in 2D and then finishing with vergence based depth
2287 judgements. There is potential, because of this, that a practice effect may be present, as all
2288 of the vergence based depth judgement experiments were completed last. However this
2289 order was deemed to be the most suitable, as the main experiment considered the
2290 interaction between the two depth factors, which was the main focus of this study, and the
2291 vergence based depth judgements asked a different question. I could have potentially
2292 randomised the order in which the experiments were completed, in an attempt to control
2293 for any practice effects. It would be interesting to see if the results would be the same with
2294 this randomisation. I predict that the results would be analogous with what I have concluded
2295 here.

2296 One aspect which would be interesting to consider would be whether or not the participant
2297 had similar responses to the virtual card being displayed in positive parallax with a noisy
2298 background behind it. Hence instead of changing the disparity of the card only I would have
2299 also changed the background noise disparity. In this instance the card would be in front of a
2300 virtual background, but both card and background would be shifted (via disparity) behind
2301 the screen plane. This was not considered here, as the wording of the question would need
2302 to be different to that which had already been asked to participants in the absolute (black
2303 background) case. I also believed there was potential that the viewer might believe the
2304 background was always in the screen plane, and didn't want this to confound results.

2305 From my initial experiment, which asked a question based on depth in the image, I find that
2306 observers reported depth based on disparity where reliable information was available. These
2307 experimental results reinforce the conclusion that the human visual system is much more

2308 adept at judging depth from relative rather than absolute disparity. However, my results also
2309 indicate that other cues may be considered when the disparity information is unreliable,
2310 even when these cues – such as expected size – do not strictly relate to depth.
2311 The implications from this suggest that miniaturization, as an industry-known problem,
2312 comes from the fact that the disparity information is given in a relative setting, so the viewer
2313 can easily distinguish where in depth the disparity of the image suggests objects are, and
2314 only then considers whether they appear the correct size or not. In the case of absolute
2315 disparity the viewer appears to first use familiar size to determine where the object is, and
2316 then either confirm with disparity information, or else discount the disparity information and
2317 use the familiar size depth cue alone. This appears to be a similar principle for viewing 2D
2318 footage, in that the perceived size is the familiar size, and the perceived depth is considered
2319 second. The issue of miniaturisation is still a small one, as people still report enjoying
2320 watching S3D content with larger interaxial values and toed in configurations (such as those
2321 used in football match filming).

2322 4.9 Conclusion

2323 The various conflicts between different depth cues generated from showing S3D
2324 content on a 2D screen has been studied extensively. This study contributes further with a
2325 consideration of the interaction between two specific depth cues, namely that of the familiar
2326 size and vergence depth information, where problems that arise are usually described in
2327 industry as miniaturization. This study has found that the type of disparity is very important,
2328 and that more weight is given to disparity over familiar size in relative disparity cases and
2329 more to familiar size over disparity in absolute disparity cases.

2330 This study considered whether a familiar size cue to depth and depth information from
2331 vergence movements were combined in an optimal fashion, and if so which cue to depth was
2332 considered the more reliable. Using signal detection theory and mathematical modelling I

2333 found that despite individual variance there was a clear trend for participants to rely more
2334 heavily on the familiar size cue as an indicator of the displayed depth, and hence this had a
2335 more significant impact on judgements when depth information was conflicting between the
2336 familiar size and the vergence based depth information. A model was constructed to quantify
2337 this relationship and it reflects the findings that cues are combined in an optimal fashion, as
2338 has been found in the past with other conflicting cue information (Brumm et al., 2010). In
2339 this study more reliability is associated with relative depth judgements than absolute.

2340 5. An eyetracker study on shifting the screen plane

2341 5.1 Introduction

2342 Leonardo da Vinci famously complained that flat paintings could never give a true
2343 impression of depth, because in real scenes the two eyes see different aspects of an object
2344 (da Vinci, 2013). Since Wheatstone's (Wheatstone, 1838) invention of the stereoscope, this
2345 limitation has been overcome, and today many forms of technology exist which are able to
2346 show the left and right eye a slightly different image of the same scene, including polarised
2347 light filters, active shutter glasses and parallax barriers (Banks et al., 2012; Burks, Harper, &
2348 Bartha, 2014; Devernay & Beardsley, 2010; Pastoor & Wöpking, 1997; Snowden et al., 2012).
2349 Advances in digital technology mean that S3D displays are more accessible than ever before
2350 (Karajeh et al., 2014): Consumers are now able to possess 3D-capable television sets in their
2351 own home (Darlin, 2005; Read, 2014); several videogame manufacturers have produced 3D
2352 versions (Schild & Masuch, 2011) and a number of companies are developing virtual reality
2353 headsets which incorporate S3D (Earnshaw, 2014).

2354 However, S3D content, especially live-action, remains complex and expensive to produce. A
2355 production standard mirror rig setup (including cameras) for S3D filming can easily cost more
2356 than \$1,000,000. Given that filming an event usually requires many different camera angles
2357 and hence many different rigs, filming a football game in S3D could require as much as
2358 \$10,000,000 of equipment (based on a minimum of 9 cameras needed, although typically
2359 the average is 12-15). These rigs have to be very precisely aligned to avoid distortions, and
2360 usually require extra personnel to operate, e.g. specialist 3D focus / convergence /
2361 interocular pullers as well as stereographers. Extra consideration also needs to be given to
2362 editing, since when changing aspects such as colour saturation and brightness, both eyes
2363 need to be changed equally or distortions quickly appear (Cavanagh, 1987). Even for
2364 computer-generated S3D content, more rendering hours and more calculations are needed.

2365 Sometimes more than two renderings of the same scene are required, since the
2366 stereographer may decide that different regions of the scene need to be rendered with
2367 different camera parameters.

2368 Given these issues, producers of S3D content occasionally use a shortcut rather than
2369 capturing every scene in S3D or converting it to S3D in post-production. They take 2D content
2370 and simply replicate the single camera lens image in both the left and the right eye, after
2371 offsetting them horizontally in opposite directions. The effect of this, geometrically, is to shift
2372 the planar 2D image back behind the screen plane; accordingly, I will call this “shifted 2D”.
2373 The shift has to be behind rather than in front of the screen plane to avoid window violations.
2374 This shift is believed in the industry to create an illusion of depth which, while not as
2375 compelling as true S3D, is nevertheless more impressive than conventional 2D.

2376 This belief is not unreasonable. 2D images contain many pictorial cues to three-dimensional
2377 structure, including perspective, shading, texture cues and apparent size. These can even
2378 trigger reflex vergence eye movements, implying that the brain accepts these depth cues at
2379 a basic perceptual, rather than simply cognitive, level. 2D video content includes still more
2380 powerful depth cues, such as structure from motion (Ringach, Hawken, & Shapley, 1996).
2381 However, there is evidence that the visual system detects the flat picture plane, and that
2382 perception is powerfully influenced by this. Indeed, this seems to be a key reason why
2383 pictures and photographs look ‘correct’ across a wide range of viewing angles, even though
2384 the image on the retina is changing profoundly (Hands, Smulders, & Read, 2015). The visual
2385 system appears to detect the screen plane and correct for the oblique viewing angle. An
2386 undesirable side-effect of this is that one remains aware at some level that the image is ‘only
2387 a picture’, projected onto a flat screen plane, rather than genuinely existing in three-
2388 dimensional space.

2389 There is a wealth of evidence that weakening the cues to the existence of the screen plane
2390 results in a stronger impression of depth. This goes back at least to Tscherning, cited in Ames
2391 (Ames Jr, 1925). Binocular disparity is a powerful cue to the flatness of the screen plane, so
2392 weakening disparity cues is an immediate way of reducing the salience of the screen plane.
2393 Tscherning discusses the depth illusion produced when 2D pictures are viewed through
2394 Javal's iconoscope, an optical device which presents the same image to both eyes. Simply
2395 viewing a picture from a greater distance produces a similar effect, but the iconoscope also
2396 disrupts the relationship between convergence and viewing distance, a manipulation which
2397 itself increases the depth illusion (cited in Ames Jr, 1925). The zograscope (Koenderink,
2398 Wijntjes, & van Doorn, 2013) worked in a similar way. Claparède (Claparède, 1904) discussed
2399 the "paradox of monocular stereopsis": the stronger depth illusion created when 2D pictures
2400 are viewed monocularly, again because this removes the binocular cues to flatness. Ames
2401 reports blurring the image in one eye also strengthens the depth illusion, especially if a
2402 cylindrical lens is used to blur vertical lines while leaving horizontal ones sharp. Again, this
2403 presumably disrupts disparity cues to flatness, leaving pictorial cues free to dominate.
2404 Binocular cues are not the only ones indicating the screen plane. Accommodation is a
2405 monocular cue to flatness, at least at near viewing distances, so removing this cue (by
2406 viewing through a small hole) or disrupting it (by viewing through positive or negative lenses)
2407 also strengthens the illusion of depth. Ames reports that viewing a flat image through a
2408 mirror produces the same effect. This is presumably by introducing uncertainty as to the
2409 position of the picture in space: the frame removes the continuity between the observer and
2410 the picture via the surrounding objects and surfaces, while the mirror's surface presents a
2411 competing candidate for picture plane. Perhaps most interestingly for the present paper,
2412 Ames also discusses "Changing the convergence of the eyes from that normally required by
2413 the distance from which the picture is viewed", by placing prisms in front of the eyes (Duane,

2414 1900; Verhoeff, 1935). This is directly equivalent to the “shifted 2D” exploited by current S3D
2415 producers.

2416 Thus, there are good grounds for expecting “shifted 2D” to produce a stronger illusion of
2417 depth than “native 2D”, presented on the screen plane. The shift introduces uncertainty
2418 about the location of the picture plane, with the physical screen presenting an alternative
2419 candidate, much as in Ames’ mirror experiment, while the vergence is now further than the
2420 physical distance of the plane. The manipulation should therefore reduce the salience of the
2421 flat screen plane, reducing conflict with monocular depth cues within the content and thus
2422 producing a stronger impression of depth. The early literature was purely qualitative, while
2423 more recently many of these effects have been examined quantitatively (Koenderink et al.,
2424 2013; Koenderink, van Doorn, & Kappers, 1994; Vishwanath & Hibbard, 2013). However, to
2425 my knowledge the present study represents the first quantitative examination of the effect
2426 of the “shifted 2D” manipulation on the experience of depth.

2427 Humans use extraocular muscles around the eye to align the foveas of the retinas onto points
2428 of interest in any particular scene in front of them, be it in the natural environment or on a
2429 viewing medium such as a screen. The introduction of S3D technology may have an
2430 interesting effect on where the attention of viewers is drawn to when content is viewed in
2431 S3D over 2D. It is difficult to ask this question subjectively, but, with the advancement of
2432 eyetracking technology, it is possible to measure gaze position on the screen, and consider
2433 whether there is a difference in where the visual system attends to while watching S3D
2434 content compared to 2D content.

2435 An important survival mechanism utilised by the visual system is to very quickly fixate on
2436 objects that are approaching the viewer (Franconeri & Simons, 2003). If the object in
2437 question is not pertinent to the task at hand being completed (whatever it may be), it should
2438 still be considered and searched with a matter of urgency. It is also vital for the visual system

2439 to fixate on objects closer than further away, and hence naturally the visual system focuses
2440 on objects closer to the viewer.

2441 Among the factors which are important in determining the depth of an approaching object,
2442 and hence determining both whether the object is indeed approaching , and how far away
2443 in depth it is, are stereoscopic depth (including both changes in the disparity between the
2444 left and right retinas images (Brenner & van Damme, 1999) and interocular velocity
2445 differences), vergence, and changes in the size of the object, particularly when considered
2446 relative to the background (looming) (Gregory, 2015; Hands et al., 2014). For example if the
2447 object is increasing in size, uniformly with the background, and other warping effects are
2448 occurring to the background objects, the focus object itself isn't approaching, the viewer is
2449 moving closer to the objects in question. Another example of how the size of an object could
2450 affect attention is to consider the familiar size of the object and compare it to the
2451 background. If it is larger than it typically should be, it may be closer than normal and should
2452 be attended to (Hands et al., 2014; Yamanoue et al., 2006).

2453 Caballero et al. considered stereoscopic depth in an attentional study and determined that
2454 stereoscopic depth is an important factor in where attention is focussed (Caballero, López,
2455 & Saiz-Valverde, 2008).

2456 The attention that is paid to approaching and moving objects is something creators of
2457 television have tried to utilise to draw the viewer's attention to certain objects and key
2458 moments in scenes. It is regularly used in advertising and also in production of movies and
2459 television, as well as in games consoles (Schild & Masuch, 2011). In 2D the objects that
2460 producers wish to have the attention of the viewer at a certain time can be altered using
2461 movement, colour, focus, and a change in size and shape; however, due to technological
2462 limitations, binocular viewing cannot be used to generate a sense of looming. With the
2463 growing popularity of S3D technology in the home, and it becoming more accessible because

2464 of falling prices (Darlin, 2005), there is now a potential to utilise the S3D technology to
2465 attempt to draw viewers' attention to a non-focus part of the scene, using binocular vision
2466 and disparity to increase attention to an object, while also utilising other cues such as size,
2467 shape, colour and focus (Welchman et al., 2005). Typically people pay attention to things
2468 closer to them (Franconeri & Simons, 2003), and so S3D could be utilised to draw the
2469 attention of the viewer. Approaching objects could now be depicted with both looming and
2470 stereoscopic depth cues to draw the attention of the viewer to the desired object with even
2471 greater strength. This would suggest that there is potential for S3D content to attract
2472 attention in different locations to 2D content.

2473 Eyetracking technology has been utilised in many different studies in measuring vergence
2474 and then from the vergence measurements calculating where in depth the participant is
2475 looking. Studies such as that conducted by Duchowski et al. attempted to use an eyetracker
2476 with a customised Wheatstone stereoscope to measure and model vergence movements
2477 from participants while viewing images with virtual depth (Duchowski, Pelfrey, House, &
2478 Wang, 2011; Wheatstone, 1838). They reported that not only did the vergence movements
2479 correspond to the depth displayed on the scene, but when considering the jitter in the
2480 function of depth perception (caused primarily by saccades) they could smoothen the
2481 function of depth perception with a quadratic filtering method. To aid in surgical procedures,
2482 Mylonas et al. considered both binocular eyetracking in robots and using the link between
2483 horizontal disparities and viewing distance (Mylonas, Darzi, & Yang, 2004). Developing an
2484 integrated stereo viewer and eyetracker they successfully considered vergence changes in
2485 their study. Hillaire et al. used eyetracking to consider where the participant was attending
2486 to (in the paper they refer to this position as the focus point) and suggest algorithms to
2487 'decimate' some of the periphery details to allow for faster rendering (Hillaire, Lécuyer,
2488 Cozot, & Casiez, 2008). This is an interesting use of eyetracking technology in speeding up

2489 processing time, which could become even more important if a study utilises S3D technology,
2490 and hence needs to process two separate images for each retina.

2491 Viewers have watched 2D scenes from a very young age, in books, television and more
2492 recently handheld appliances such as smartphones and tablets (Read, 2014). In doing so,
2493 have they learned a certain way to view content that applies to 2D, and hence fixate on
2494 specific points on a scene because they are the important points that demand attention?
2495 Does the introduction of S3D technology mean that participants fixate somewhere else on
2496 the screen compared to 2D content? Or does S3D technology simply augment the attention
2497 viewers already assign to objects in the scene? Do participants tend to fixate on objects of
2498 interest (typically the focus of the scene at that moment) regardless of whether content is
2499 displayed in 2D or S3D? Due to how important vergence and disparity cues are to depth
2500 perception, it may be that the introduction of artificial depth from S3D technologies affects
2501 the saliency of the content in question, as adding disparity could cause a subsection of the
2502 content to become more salient, which would affect the fixation and gaze position in
2503 theoretical models such as dynamic routing (Tsotsos, Culhane, Wai, Lai, Davis & Nuflo, 1995)
2504 and feature integration theory (Itti, Koch & Niebur, 1998). I predict that the gaze position
2505 and fixation of participants is very much content driven, and that the introduction of
2506 stereoscopic depth from disparity in the content will have little to no effect on fixation.

2507 In this chapter I aim to consider whether the technique of shifting the plane that 2D content
2508 is shown at away from the screen plane to a virtual plane at a depth behind the screen, as
2509 used in industry, is a viable alternative to native S3D, considering whether the perceived
2510 depth impression is comparable and analogous between the two. I also consider eyetracking
2511 data taken from the different clip types to consider if, for the same content, viewers fixate
2512 to different locations during S3D viewing compared to 2D viewing.

2513 5.2 Material and methods

2514 5.2.1 Equipment

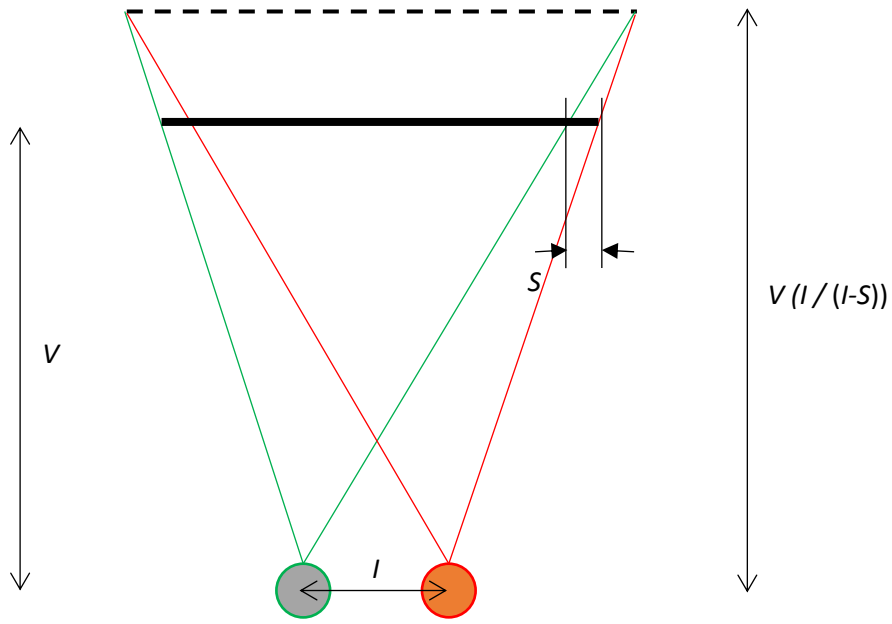
2515 The stimuli were displayed on a 21.5" passive stereoscopic 3D display monitor (AOC
2516 D2367ph, http://www.aocmonitorap.com/v2015/nz/product_display.php?id=409) in a
2517 room which had regular, constant background luminance of 161.2cd/m² (average of ten
2518 measurements made using a Minolta LS – 100 photometer). The monitor resolution was
2519 1920 x 1080 pixels, 47.6 cm wide x 26.8cm high. The monitor was of the patterned-retarder
2520 type where left and right images are separated by circular polarisation and displayed on
2521 alternate pixel rows, halving the vertical resolution. The stimuli were created in side-by-side
2522 S3D format, thus halving the horizontal resolution, which the monitor converted to row-
2523 interleaved format. A chinrest was used to ensure that each subject viewed the content from
2524 the same position both horizontally and vertically with each trial, to ensure other effects,
2525 such as viewing distance and viewing angle (Brenner & van Damme, 1999; Hands et al.,
2526 2015), were not factors in determining immersion. A height adjustable chair was used to
2527 ensure the participants were comfortable during the experiment. Throughout the
2528 experiment participants wore passive S3D glasses, so that they could not ascertain whether
2529 content was being shown in 2D or S3D by the presence or absence of glasses. The viewing
2530 distance was 100cm. Participants gaze was tracked using an Eyelink Eyetracker 1000, on an
2531 angled binocular configuration with a 25mm wheel lens. The eyetracker was positioned 55cm
2532 away from the participant, underneath the monitor.

2533 5.2.2 Stimuli

2534 The stimuli were 13 separate 30 second clips from the BSkyB production 'Micro
2535 Monsters with Sir David Attenborough', which was filmed in S3D. Clips were chosen from 2
2536 episodes that were made available by BSkyB for the study, and were chosen so that the 30
2537 second timespan started and ended at a sensible place, avoiding starting or stopping the clip

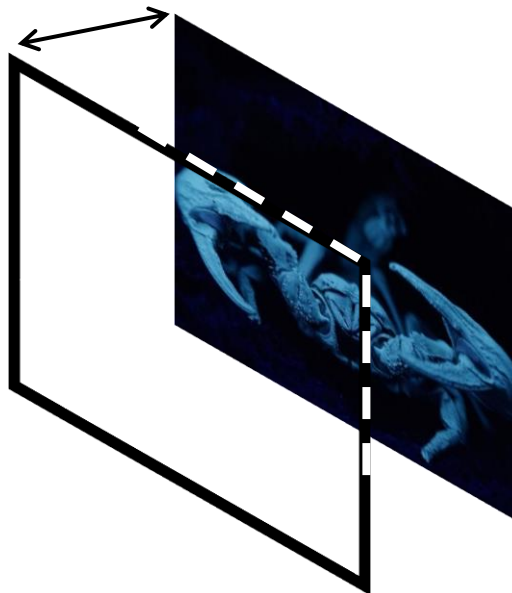
2538 midsentence, and also to be sure the clips were engaging. Both the left and right eye of each
2539 clip was made available in an AVI file for the study. The software program 'Stereo Movie
2540 Maker' (available at <http://stereo.jpn.org/eng/stvmkr/index.html>) was used to modify the
2541 clips. The subsequent modified clips were each displayed in four different ways:

- 2542 • Native S3D - showing the left clip to the left eye and the right clip to the right
2543 eye, as typically done in S3D content displays.
- 2544 • Native 2D - showing the left clip to both the left and right eye. (Note that this
2545 will have been different to the 2D production that BSkyB showed on its channels,
2546 as a different editing procedure will have been used for the 2D footage.)
- 2547 • Shifted S3D - as for Native S3D but in this case the left image was shifted left by
2548 56 pixels and the right image was shifted right 56 pixels.
- 2549 • Shifted 2D - as for Native 2D but this time the original left clip was shifted left by
2550 56 pixels and displayed to the left eye, while the *same* clip was shifted right by
2551 56 pixels and displayed to the right eye. Geometrically, this is equivalent to
2552 displaying flat 2D content on a plane behind the monitor screen (Fig. 5.1.), as if
2553 viewed through a glass window (Fig. 5.2.). Shifting each eye's image by a distance
2554 S in this way increases the geometrically-defined distance by a factor $I / (I-S)$,
2555 where I is the observer's interocular distance, in the same units as S . This factor
2556 is independent of the screen width and of the viewing distance. For my
2557 experiment, after resizing for display purposes, the displayed shift was 60 pixels
2558 = 1.44cm, meaning that for an observer with eyes 6.3cm apart (Dodgson, 2004),
2559 the geometry specified an increase in distance of 1.28. Thus at my 100cm-
2560 viewing distance, the shift places the virtual content 128cm from the observer
2561 according to the binocular geometry (Fig. 5.1.).



2562

2563 **Fig. 5.1.** Geometry of my experiment. The interocular distance I is for calculations assumed
 2564 to be 6.3cm. By shifting the resized images 30 pixels in each eye, the virtual image is moved
 2565 behind the screen by a factor of 1.28. So when the viewing distance V is 100cm, the plane of
 2566 the image should appear to be 128cm away.



2567

2568 **Fig. 5.2.** Diagram to explain the concept of shifted 2D clips. The frame of the television is
 2569 shown, and the clip is moved in depth behind it, using disparity. My aim is to determine if this

2570 *gives as good a sense of immersive depth as native S3D, as it is a lot easier and less expensive*
2571 *to create content like this.*

2572 This resulted in 52 different stimuli for the participant to look at, 4 versions of each of 13
2573 clips. The 52 stimuli were shown in the same order to each of 9 participants. The four versions
2574 of each clip were presented consecutively, with the four conditions coming in a different
2575 order for each clip, to ensure that, e.g. the shifted 2D clip, wasn't the last of the four
2576 repetitions each time, which could risk affecting the results if the participant had lost interest
2577 in the clip by the fourth repeat.

2578 Due to a formatting issue with creating the shifted clips, the frame rates were 25 fps for the
2579 native and 12 fps for the shifted clips. Participants did not report noticing any differences
2580 between the frame rate of the clips in terms of quality or flicker, and indeed the authors
2581 could not reliably detect which clips were displayed at which frame rate. As shown below,
2582 the results did not correlate with frame rate.

2583 5.2.3 Procedure

2584 Participants were asked to sit in the chair comfortably, wearing the glasses and
2585 resting their chins on the chinrest. They were then given a brief explanation of the eyetracker
2586 technology and a very basic explanation of why it was being used in the experiment (i.e. to
2587 record where participants are looking during the clip presentation). The participant, before
2588 beginning the experimental trials, went through a calibration and validation process using
2589 the eyetracker, to be sure the recording was correctly measuring where the participant was
2590 looking on the screen. The validation process gave an average deviation from target and a
2591 maximum deviation value in degrees. If the average deviation was more than 0.5° away or
2592 the maximum was more than 1° away from the validation target on the screen a recalibration
2593 took place. This was the limit for 'good' validations according to the eyetracking software.
2594 Each individual trial then showed an initial timing clip, to ensure that recording was

2595 accurately started when the clip began, and then the 30 second clip of the content was
2596 shown. After each clip the participants were asked to assess the perceived depth in the
2597 image, stressing that the actual content (i.e. how interesting it was) wasn't important to the
2598 study, and to give a score on a 7 point Likert scale, with 1 being "terrible" or "not noticeable
2599 3D" to 7 being "fantastic, immersive 3D". Once the participant reported their score on the
2600 impression of the depth in the clip the trial was finished and the next trial would begin.
2601 Between each trial another validation test was completed to ensure the participants head
2602 hadn't moved away from a position acceptable for the eyetracker (0.5° average, 1°
2603 maximum) and, if the validation failed, a recalibration took place immediately before the
2604 next experimental trial. The consequence of the procedure being structured as such resulted
2605 in different experiments taking different lengths of time. Some participants only needed very
2606 few calibrations, while others needed a recalibration on roughly a 3 trial basis. Participants
2607 were allowed to move away from the headrest whenever they wanted to, but, if they did, a
2608 recalibration took place before the experiment was allowed to continue. Both left and right
2609 eye measurements were made, with the intention of being able to calculate the position in
2610 depth that the eye was verging to, by considering the parallax between the two eyes.
2611 Participants were given a £10 shopping voucher for their participation.

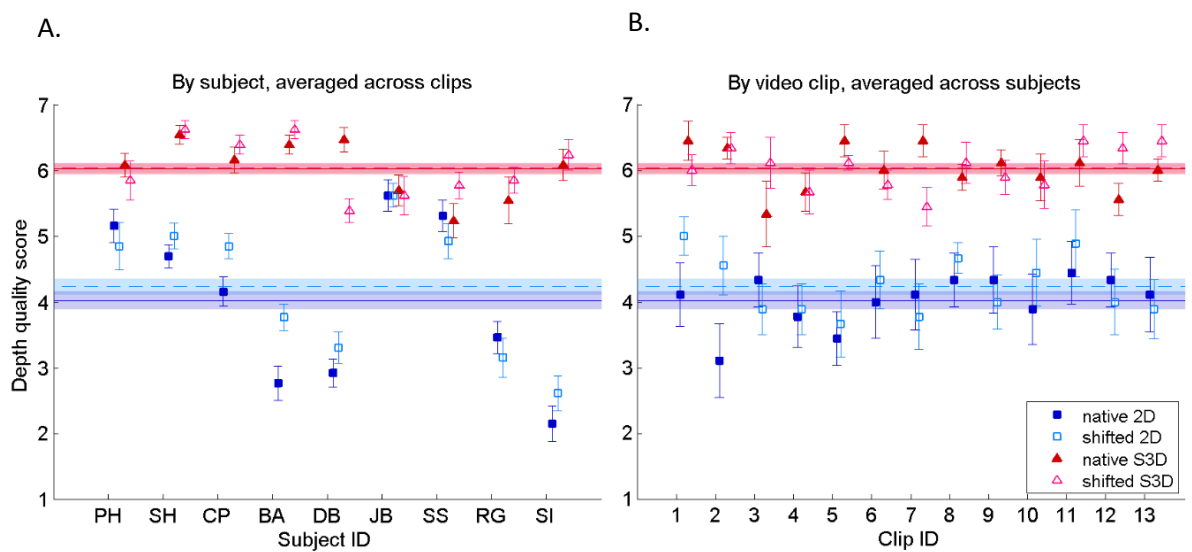
2612 5.2.4 Participants

2613 Participants were recruited via an internal volunteer scheme at Newcastle University
2614 Institute of Neuroscience, on the basis that they had no visual problems. The work was
2615 approved by Newcastle University Faculty of Medical Sciences Ethics Committee. 9
2616 participants (6 female, all naïve; 3 male, author PH and 2 naïve) were used in the study. Naïve
2617 participants were not informed of the experimental aims or hypotheses, until the experiment
2618 was completed, at which point they were debriefed on the aims of the study, however they
2619 will have been able to ascertain that there was something different about the repeated
2620 versions of the clips.

2621 5.3 Results

2622 5.3.1 Subjective ratings

2623 Fig. 5.3 shows the average score for each different viewing condition, (A) for the
2624 different subjects and (B) for the different clips. It is immediately clear that the depth ratings
2625 are substantially higher for S3D (red, triangles) than for 2D (blue, squares). However, there
2626 is very little difference between the native (filled shapes, solid lines) and shifted (empty
2627 shapes, dashed lines) formats. For S3D, shifting has no effect (mean rating 6.02 for native vs
2628 6.03 for shifted). For 2D, depth ratings are marginally higher for shifted (mean rating 4.03 for
2629 native vs 4.23 for shifted), but this difference is not significant. A two-way repeated-
2630 measures ANOVA on each subject's average ratings across the 13 clips, with stereo (2D vs
2631 S3D) and plane (native vs shifted) as factors, found a highly significant main effect of stereo
2632 condition ($F=19.9$, $P=0.002$), but no main effect of plane ($F=1.10$, $P=0.33$) and no interaction
2633 between plane and stereo condition ($F=0.634$, $P=0.45$). There was no evidence that the clips
2634 themselves differed in the depth impression they produced. For example, a Kruskal-Wallis
2635 test finds no difference between the ratings given to the 13 different native 2D clips ($P=0.44$).



2636

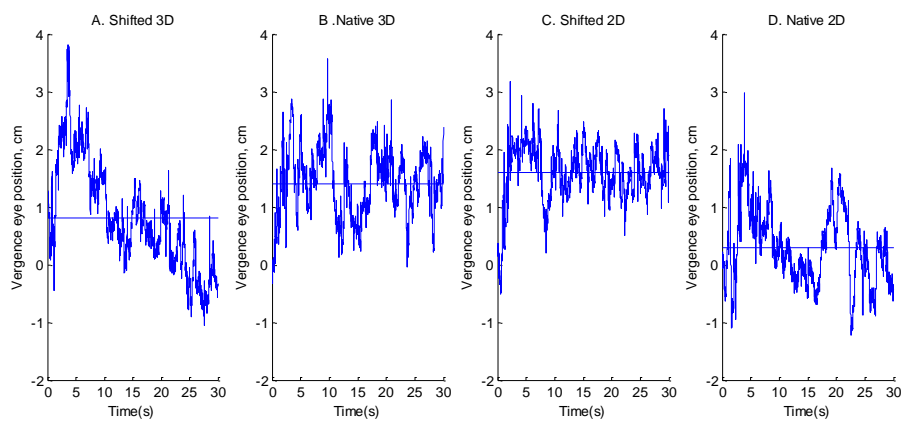
2637 **Fig. 5.3.** Depth quality scores for each of the four different viewing conditions, (A) for the 9
2638 different subjects, averaged across the 13 video clips, and (B) for the 13 different clips,
2639 averaged across all 9 subjects. Blue squares show results for 2D, red triangles for S3D; filled
2640 symbols / solid lines are for native content, empty symbols / dashed lines are for content
2641 shifted behind the screen plane. Errorbars show $\pm 1SEM$ of the 9 subjects' judgments for
2642 each data-point; points are offset horizontally so that errorbars do not overlap. Horizontal
2643 lines show means for each condition, averaged across content & subjects.

2644 5.3.2 Vergence eye measurements

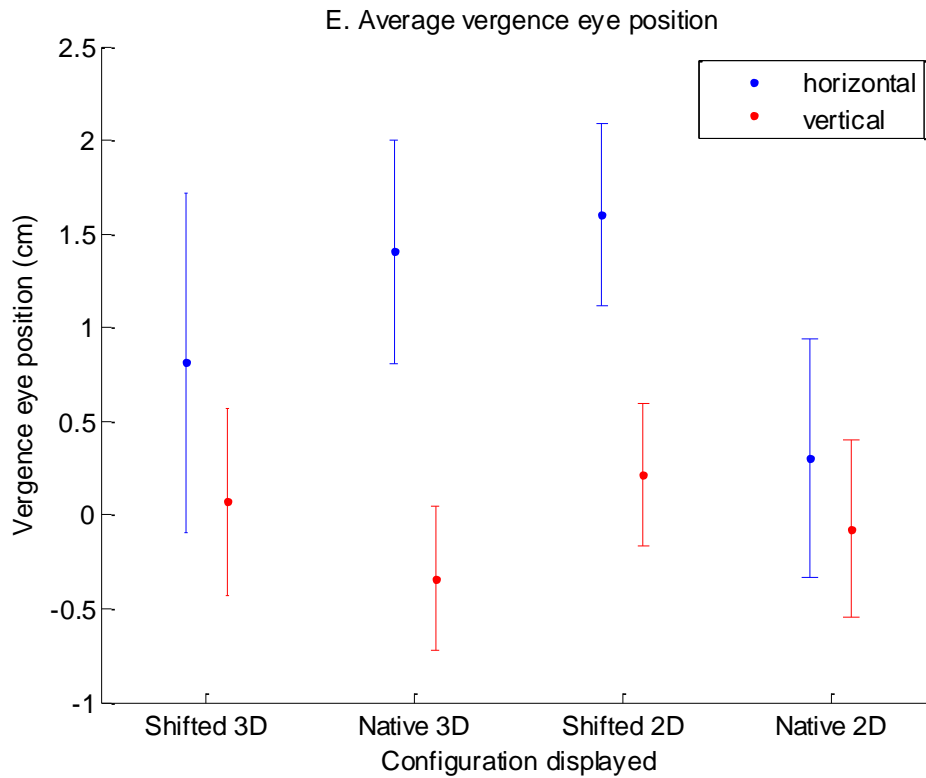
2645 I converted the .edf files from the eyetracker to .mat files and discarded most of the
2646 information the Eyelink eyetracker collects that was not pertinent to this study. I consider
2647 the left and right eyes' gaze position during the 30 second clip. Because of how regularly the
2648 eyetracker took measurements (once every 0.002 seconds) each trial had 15,000 left and
2649 right eye positions in x and y coordinates on the screen, with the top left corner being (0,0).
2650 As we have seen in section 5.3.1, the industry trick used by producers of S3D television and
2651 cinema content (to occasionally shift the 2D scene behind the screen plane using disparity
2652 between two identical images) gives a different impression of depth than 'true' S3D content.
2653 It is clear that the 2D content is being shown on the flat surface, at a set depth. In the case
2654 of S3D the disparity and vergence depth information for the content is different to that of
2655 the surface it is being displayed on, and hence the cues to depth are in conflict. Because of
2656 this in the following experiment we now consider the 'shifted' and 'native' cases to be two
2657 different samples taken of S3D and 2D content.

2658 The first analysis conducted on the eyetracking data was to consider the measured parallax
2659 between the right and left eye gaze positions in cm, as recorded by the eyetracker when both
2660 eyes were being detected successfully at the same time. This was to consider if the visual
2661 system reacts differently to shifted 2D and native S3D content, in terms of vergence. I

2662 estimate the vergence at each time point by considering the difference in location
2663 horizontally between the left and right eyes gaze position. This is given as a measurement in
2664 pixels. Hence by using a simple pixel to cm conversion for the monitor in question, the
2665 parallax between the eyes position (and hence an estimation of the vergence) for each time
2666 point was calculated. An example for a participant for one trial is shown below to illustrate
2667 in Fig. 5.4. Sub figures A, B, C and D show the different format the trials were shown in and
2668 the recorded parallax between the gaze positions of right and left eyes. The blue solid line
2669 indicates the average depth measurement for that trial. Sub figure E shows the average value
2670 for both the horizontal and vertical parallax values in cm for each configuration, in blue and
2671 red respectively, with 95% confidence intervals displayed. Even with rigorous calibration,
2672 eyetrackers can be subject to a considerable amount of error in calculating where the eyes
2673 are fixating while recording data, and this is reflected in my vertical parallax measurements
2674 which my results suggest were different from zero. However the horizontal parallax
2675 measurements are larger than the vertical parallax. These are subject to a high level of noise
2676 with large standard deviation values, which can be attributed to the noise of recording.



2677

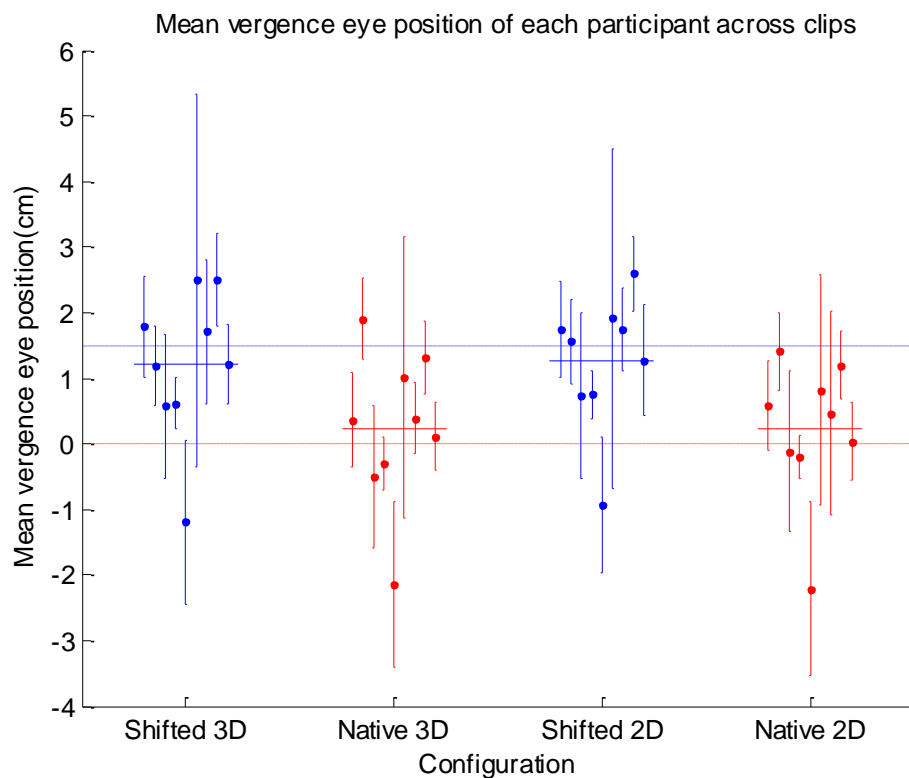


2678

2679 **Fig. 5.4.** Participant CPs vergence eye measurements for clip number 6 using left and right
 2680 eye gaze positions for different configurations (A – D). X axis is the time of the 30 second clip,
 2681 y axis is the measured depth calculated from the difference between the gaze positions of the
 2682 left and right eyes. The average value is shown as a solid blue line in each figure. Average
 2683 vergence eye measurement for each configuration is shown also (E) in blue. Errorbars show
 2684 ± 1 standard deviation. Vertical vergence eye measurements between left and right eyes gaze
 2685 is shown in red.

2686 This was very typical of all measurements made on all participants, shown in Fig 5.5: mean
 2687 values that suggest a change in vergence depending on the configuration. Across
 2688 participants, the average depth measured using vergence movements for the shifted S3D
 2689 content was 125.6cm, for the shifted 2D content 127.3cm, for the native S3D 102.0cm and
 2690 for the native 2D 101.5cm, calculated from the parallax in a similar procedure as used above
 2691 to calculate the depth of the screen plane. Thus the shifted and native depth measurements
 2692 were, in each case, near to the depth plane that the content was displayed on (100cm,

2693 physical screen plane for native, and 128cm, virtual plane through a window for shifted;
 2694 dashed lines in Fig 5.5). Of course, the S3D content varies in depth, but the depth in S3D clips
 2695 is typically centred on the screen plane, with some content in front of it and some behind it.
 2696 Thus it is not surprising that in Fig 5.5. the native and shifted S3D content is shown to have a
 2697 similar mean parallax value as the 2D content. For both native and shifted content, people
 2698 on average verge around the depth plane of the content. This was despite the fact that the
 2699 content plane was purely virtual in the shifted conditions, whereas in the native conditions,
 2700 the mean depth of the content plane was on the physical screen, so the binocular depth cues
 2701 were also supported by the accommodation cues and other cues to the physical screen
 2702 plane.



2703

2704 **Fig. 5.5.** Mean vergence eye position of each participant across clips. The vergence eye
 2705 position was averaged across each clip initially and then across all clips for each participant,
 2706 shown by configuration. Errorbars show ± 1 standard deviation. As can be seen, the value of
 2707 the native vergence eye measurements were, on average, lower than that of the shifted

2708 configurations. Mean across participants is shown by solid bars. A blue dotted line indicates
2709 zero movement, and a red dotted line indicates 1.5cm (the amount I introduced when
2710 shifting the clips).

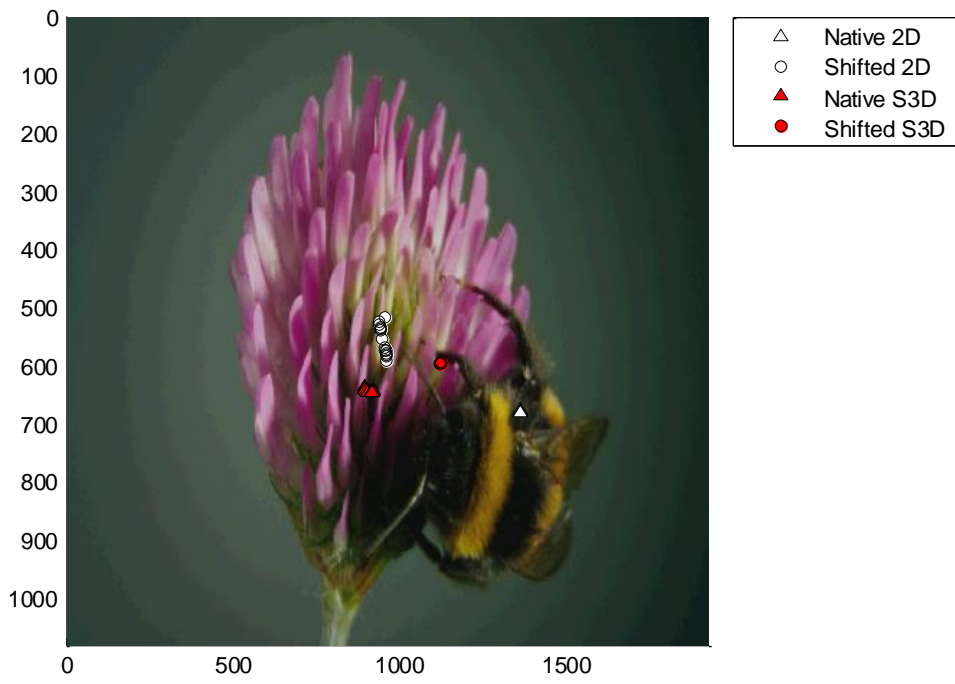
2711 To analyse the statistical significance, I calculated the mean parallax measurements per clip
2712 for the 2D and S3D configurations for each participant. I then calculated the mean parallax
2713 measurement per subject for the shifted 2D and native S3D configurations by averaging over
2714 the 13 clips. I then conducted paired t-test analysis on the n=9 data points and found a
2715 significant difference between depth measurements for shifted 2D content and depth
2716 measurements for native S3D content based on eye position ($P < 0.0001$, paired t-test on
2717 n=9 subjects). This confirms that despite a considerable amount of noise in the eyetracking
2718 data my measurements were successfully able to discriminate parallax shifts in the content.

2719 I consider whether there was a difference in the spread of the parallax measurements
2720 between configurations. There are two reasons we might expect such differences. First, we
2721 might expect measurements for shifted (or native) 2D content, where the screen parallax is
2722 constant, to have less variation in parallax than measurements for native (or shifted) S3D
2723 configurations, where the screen parallax is constantly varying. Second, vergence
2724 measurements for the shifted conditions might be more variable than those for the native
2725 conditions, since the native conditions provide additional cues to the location of the screen
2726 plane. However, paired t-test analysis on the standard deviation of the shifted and native
2727 vergence measurement (conducted as above with the mean) showed there was no
2728 significant difference between conditions in the average standard deviation values for
2729 participants across clips. The closest p-value to significance was that when comparing shifted
2730 and native S3D content ($P = 0.08$, paired t-test on n=9 SD values, averaged over times and
2731 clips). I have not analysed why this is the case, but suggest it could be attributed to the depth
2732 from parallax inside the individual S3D clips, which I did not analyse. This analysis is

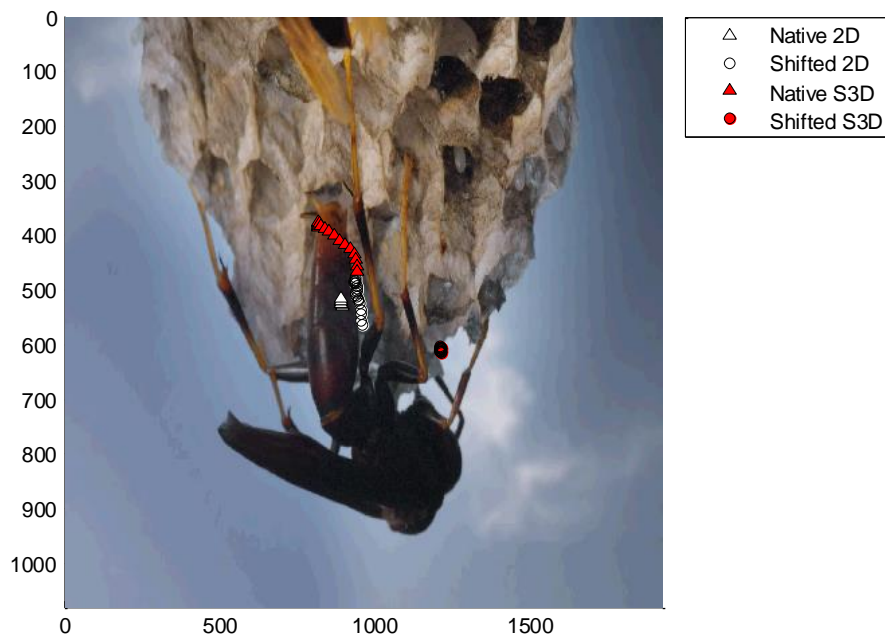
2733 necessarily limited to the content in question, so it could explain why the standard deviation
2734 measurements were not significantly different.

2735 5.3.3 Gaze position during the clips

2736 Our second analysis was centred on assessing where participants actually focussed
2737 attention during the showing of content. To consider this I took the average of the left and
2738 right eye's position both horizontally and vertically for each time measurement, and plotted
2739 the position over the corresponding frame in the image. Because the number of frames was
2740 considerably lower than the number of measurements each frame showed 20 different eye
2741 measurements for native content and 40 for shifted content. A selection of three examples
2742 is shown below in Fig. 5.6. Qualitatively, participants appeared to follow the focus object of
2743 the scene in question, and after the experiment in feedback they reported that while they
2744 noticed a difference between the clips (shifted 2D, native S3D etc.) they felt as though they
2745 were still watching in the same way.



2746



2747

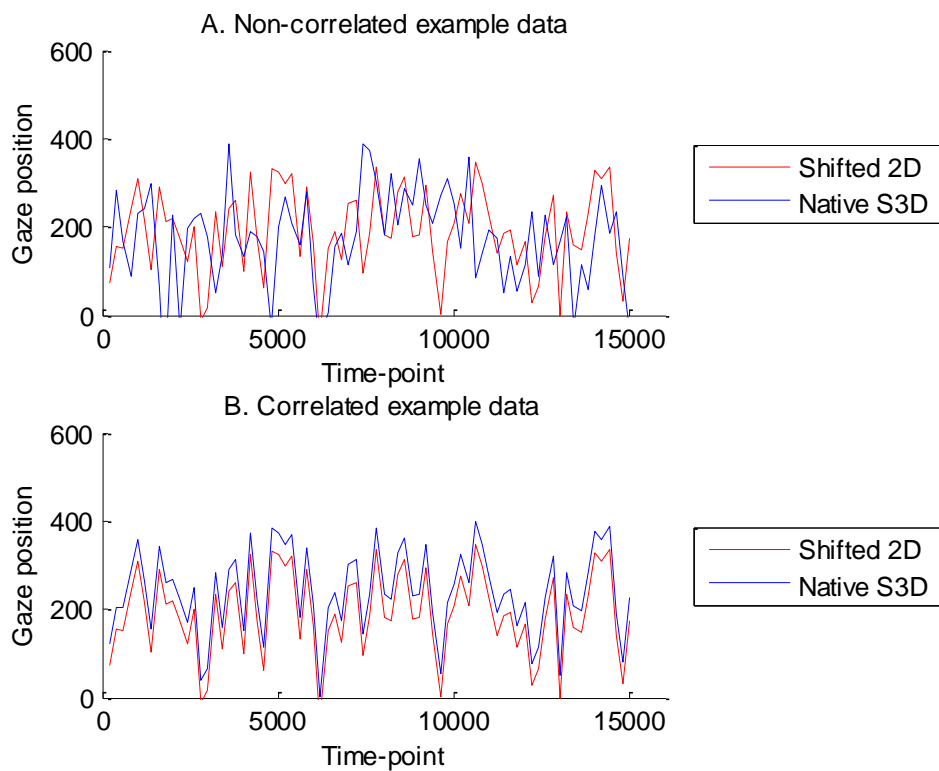
2748 **Fig. 5.6.** 3 Eye gaze positions from example frames of three different clips. As can be seen,
 2749 the attention in that frame appears to be drawn mostly to the focus object in the scene. The
 2750 discrepancies between the different gaze positions for different configurations are non-
 2751 significant when the entire clips are considered.

2752 I conducted analysis to consider where participants actually looked during the clips by
 2753 considering gaze position. For each time-point of recording (15,000 recordings for each 30
 2754 second clip) I calculated the average gaze position by taking the mean of the left and right
 2755 eyes' horizontal and vertical eye position. E.g. if the left eye was at position (2, 2) and the
 2756 right eye was at (4, 4), then for that time-point the gaze position was calculated as (3, 3).

2757 Because each clip is inherently different, and participants may not necessarily look in the
 2758 same place for the same content, for this analysis I didn't average across clips or participants,
 2759 but instead compared data per participant and clip. T-test analysis on configuration revealed
 2760 there was no significant difference between gaze positions for any configurations across clips
 2761 and participants (paired t-test on n=15,000 datapoints). This suggests that the participants

2762 attended to the same features in the clips throughout, regardless of the configuration that
2763 the clip was shown in.

2764 I should also consider potential correlations in the data, and demonstrate this with simulated
2765 data in Fig. 5.7. below. T-test analysis on 5.7.A would show no significant difference in gaze
2766 position when clearly there is a bigger difference between them (both datasets random
2767 numbers generated from a normal distribution, Pearson's rank correlation coefficient =
2768 0.0023, paired t-test analysis on n=15,000 data points, P=0.776) and 5.7.B (where datasets
2769 have a correlation of 1.0). It stands to reason that the different clips and participants might
2770 have significantly correlated data.

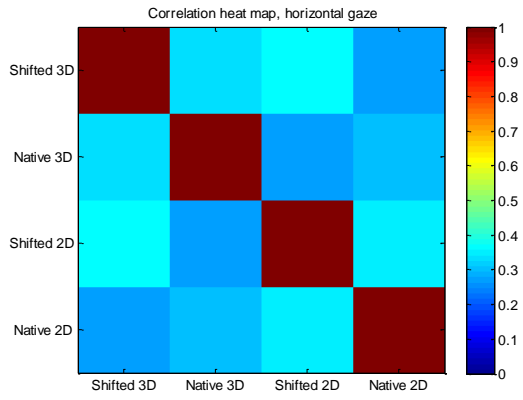


2771

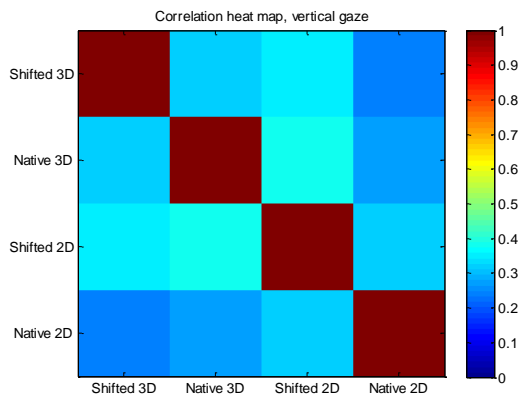
2772 **Fig. 5.7.** Example data to show that correlations are an important consideration. A) Two
2773 random datasets generated from the same normal distribution, with a very low correlation
2774 (0.0023). B) The random dataset for shifted 2D simulated data has been kept, but the native
2775 S3D data has now been directly correlated (by adding 50 to the gaze position).

2776 To consider the correlations, I calculated the average gaze position (averaged as explained
2777 above across left and right eyes for each participant for each clip), both horizontally and
2778 vertically. Some of the clips have missing data, which can occur when the participant blinks
2779 (pupil data lost) or when the participant looks away from the screen (pupil information too
2780 warped to calculate a gaze position). To deal with this issue I removed all data across each
2781 clip and participant separately if any configuration had missing horizontal or vertical data for
2782 that time point. I then calculated Pearson's rank correlation coefficients between pairs of
2783 configurations (e.g. native vs shifted 2D, native 2D vs shifted S3D) on the pooled data. All
2784 clips were similar in the pattern of their correlations when considered separately so I believe
2785 this is justified.

2786 I use heat map plots to consider the correlation data. In these, high correlations between
2787 respective configurations would be indicated with a red square, and lower correlations
2788 would be blue. Understandably the correlation values on the diagonal are 1 so these squares
2789 are completely red. If the heat map had a structure of red in the top left and bottom right
2790 quadrants and blue in the top right and bottom left, this would indicate people look in
2791 different places when they consider S3D content and 2D content. The correlation data for
2792 participant one is shown in fig. 5.8 as an example heat plot of the correlations. All the heat
2793 maps were very similar in structure. Fig. 5.9. shows the heat map for the correlations
2794 averaged across participants and gaze position.



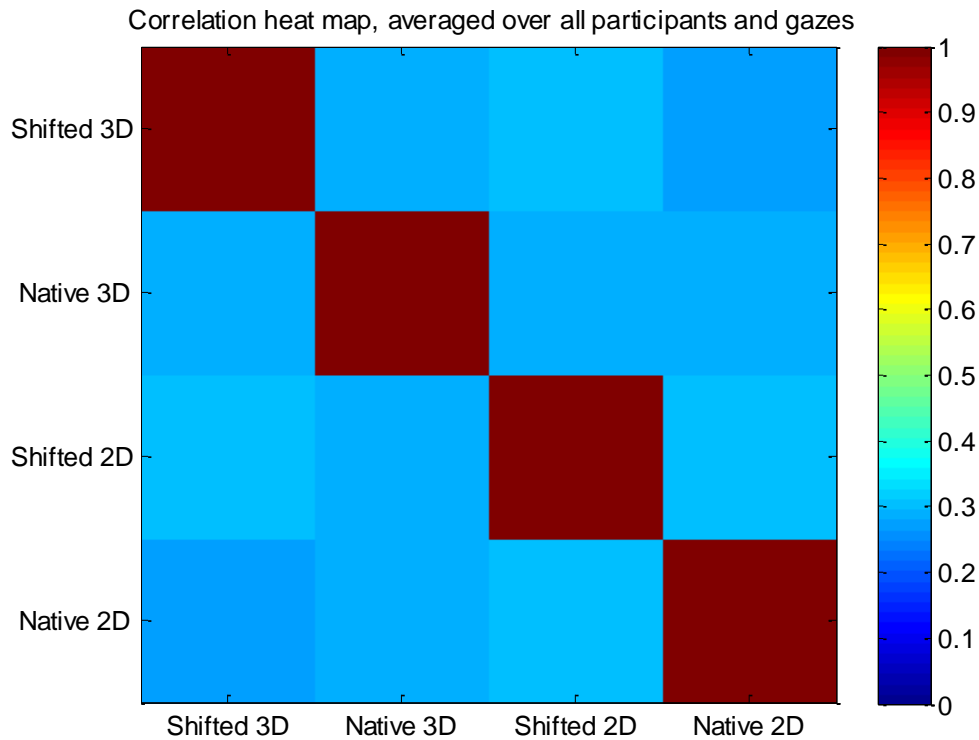
2795



2796

2797 **Fig. 5.8.** Example correlation heat maps for participant one in the horizontal and vertical gaze

2798 positions. Red squares indicate a high correlation, blue squares a low correlation.



2799

2800 **Fig. 5.9.** Heat map for correlation values averaged across all participants and gaze positions.

2801 Table 5.1. shows the values of the correlations (R) for each participant and configuration, as
 2802 well as the respective number of gaze positions compared (N) for both the x and y gaze
 2803 positions. The p-values are not listed as all correlations were highly significant ($P < 10^{-20}$).
 2804 Table 5.2. shows the mean correlation when all subjects' horizontal and vertical correlations
 2805 are averaged. As can be seen, all participants had positive correlations significant from zero
 2806 when considering the different configurations. This is not surprising, as it would stand to
 2807 reason that gaze is mostly driven by the content on the screen, and hence clips are fairly well
 2808 correlated. However from the heat maps and the R values there is no higher correlation
 2809 between different configurations. For example, it is not the case that gaze position on the
 2810 two different presentations of S3D content are more highly correlated than gaze position on
 2811 the two different presentations of 2D content. This suggests that participants look in similar
 2812 places regardless of whether the content is displayed in S3D or in 2D, indicated by the similar
 2813 R values and heat map colours.

Participant 1 N = 184646		Shifted S3D	Native S3D	Shifted 2D	Native 2D
X gaze position	Shifted S3D	R = 1	R = 0.33	R = 0.37	R = 0.27
	Native S3D		R = 1	R = 0.28	R = 0.30
	Shifted 2D			R = 1	R = 0.34
	Native 2D				R = 1
Y gaze position	Shifted S3D	R = 1	R = 0.32	R = 0.35	R = 0.24
	Native S3D		R = 1	R = 0.38	R = 0.27
	Shifted 2D			R = 1	R = 0.32
	Native 2D				R = 1
Participant 2 N = 194623		Shifted S3D	Native S3D	Shifted 2D	Native 2D
X gaze position	Shifted S3D	R = 1	R = 0.23	R = 0.26	R = 0.21
	Native S3D		R = 1	R = 0.17	R = 0.20
	Shifted 2D			R = 1	R = 0.29
	Native 2D				R = 1
Y gaze position	Shifted S3D	R = 1	R = 0.21	R = 0.25	R = 0.12
	Native S3D		R = 1	R = 0.25	R = 0.24
	Shifted 2D			R = 1	R = 0.32
	Native 2D				R = 1
Participant 3 N = 161897		Shifted S3D	Native S3D	Shifted 2D	Native 2D
X gaze position	Shifted S3D	R = 1	R = 0.42	R = 0.42	R = 0.40
	Native S3D		R = 1	R = 0.33	R = 0.34
	Shifted 2D			R = 1	R = 0.40

	Native 2D				R = 1
Y gaze position	Shifted S3D	R = 1	R = 0.19	R = 0.21	R = 0.21
	Native S3D		R = 1	R = 0.19	R = 0.18
	Shifted 2D			R = 1	R = 0.18
	Native 2D				R = 1
Participant 4 N = 106364		Shifted S3D	Native S3D	Shifted 2D	Native 2D
X gaze position	Shifted S3D	R = 1	R = 0.27	R = 0.27	R = 0.22
	Native S3D		R = 1	R = 0.29	R = 0.32
	Shifted 2D			R = 1	R = 0.22
	Native 2D				R = 1
Y gaze position	Shifted S3D	R = 1	R = 0.19	R = 0.32	R = 0.34
	Native S3D		R = 1	R = 0.22	R = 0.23
	Shifted 2D			R = 1	R = 0.28
	Native 2D				R = 1
Participant 5 N = 171213		Shifted S3D	Native S3D	Shifted 2D	Native 2D
X gaze position	Shifted S3D	R = 1	R = 0.23	R = 0.31	R = 0.29
	Native S3D		R = 1	R = 0.21	R = 0.31
	Shifted 2D			R = 1	R = 0.21
	Native 2D				R = 1
Y gaze position	Shifted S3D	R = 1	R = 0.24	R = 0.29	R = 0.21
	Native S3D		R = 1	R = 0.38	R = 0.28
	Shifted 2D			R = 1	R = 0.30
	Native 2D				R = 1

Participant 6 N = 182152		Shifted S3D	Native S3D	Shifted 2D	Native 2D
X gaze position	Shifted S3D	R = 1	R = 0.37	R = 0.38	R = 0.36
	Native S3D		R = 1	R = 0.38	R = 0.39
	Shifted 2D			R = 1	R = 0.41
	Native 2D				R = 1
Y gaze position	Shifted S3D	R = 1	R = 0.30	R = 0.35	R = 0.29
	Native S3D		R = 1	R = 0.31	R = 0.31
	Shifted 2D			R = 1	R = 0.37
	Native 2D				R = 1
Participant 7 N = 90698		Shifted S3D	Native S3D	Shifted 2D	Native 2D
X gaze position	Shifted S3D	R = 1	R = 0.38	R = 0.31	R = 0.32
	Native S3D		R = 1	R = 0.29	R = 0.28
	Shifted 2D			R = 1	R = 0.29
	Native 2D				R = 1
Y gaze position	Shifted S3D	R = 1	R = 0.32	R = 0.31	R = 0.32
	Native S3D		R = 1	R = 0.27	R = 0.36
	Shifted 2D			R = 1	R = 0.33
	Native 2D				R = 1
Participant 8 N = 131823		Shifted S3D	Native S3D	Shifted 2D	Native 2D
X gaze position	Shifted S3D	R = 1	R = 0.31	R = 0.28	R = 0.38
	Native S3D		R = 1	R = 0.30	R = 0.33
	Shifted 2D			R = 1	R = 0.36

	Native 2D				R = 1
Y gaze position	Shifted S3D	R = 1	R = 0.24	R = 0.28	R = 0.27
	Native S3D		R = 1	R = 0.37	R = 0.31
	Shifted 2D			R = 1	R = 0.37
	Native 2D				R = 1
Participant 9 N = 132277		Shifted S3D	Native S3D	Shifted 2D	Native 2D
X gaze position	Shifted S3D	R = 1	R = 0.42	R = 0.24	R = 0.27
	Native S3D		R = 1	R = 0.27	R = 0.33
	Shifted 2D			R = 1	R = 0.28
	Native 2D				R = 1
Y gaze position	Shifted S3D	R = 1	R = 0.32	R = 0.14	R = 0.33
	Native S3D		R = 1	R = 0.22	R = 0.32
	Shifted 2D			R = 1	R = 0.22
	Native 2D				R = 1

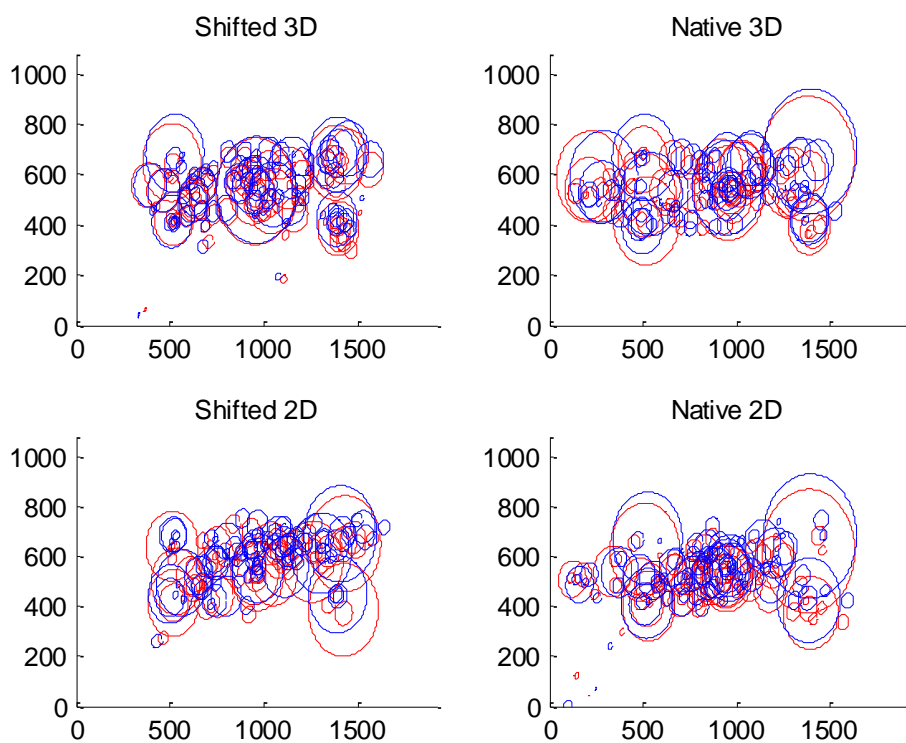
2814 **Table 5.1.** Computed R values for each participant in both the X (horizontal) and Y (vertical)
2815 gaze position. Respective N values also shown for each participant (representing number of
2816 datapoints compared in each correlation). P-values not shown as all were calculated to as P
2817 < 10⁻²⁰.

	Shifted S3D	Native S3D	Shifted 2D	Native 2D
Shifted S3D	R = 1	R = 0.294	R = 0.298	R = 0.281
Native S3D		R = 1	R = 0.284	R = 0.295
Shifted 2D			R = 1	R = 0.306
Native 2D				R = 1

2818 **Table 5.2.** Correlation R values averaged over all 9 participant and 2 gaze positions.

2819 5.3.4 Fixation analysis

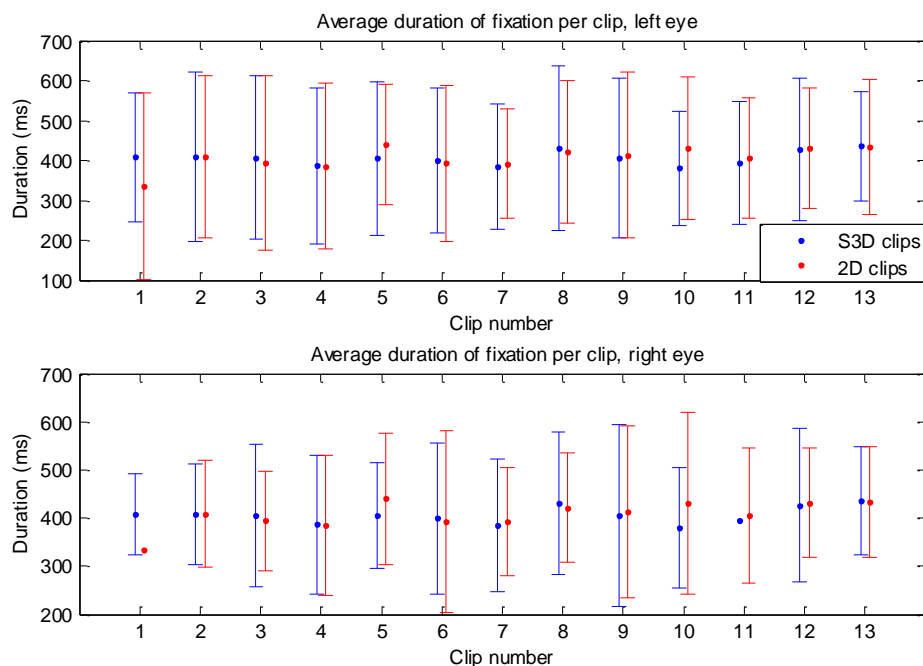
2820 I conducted further analysis of the eyetracking data by looking at the fixations
2821 involved during the viewing of the clips. The eyetracker software saves fixations in the EDF
2822 files so I used these to consider if there was any significant difference in fixation data
2823 between S3D and 2D data. I plotted figures that showed the location of each fixation on the
2824 screen in pixel coordinates. These are represented by circles, centred on the x and y
2825 coordinates of the fixation in Fig 5.10. below. Each fixation has a different duration in ms
2826 and this is represented by the relative size of the circle, with a larger radius indicating a
2827 longer fixation period. Both left and right eye fixations are shown, with the left eye
2828 fixations shown in red and the right eye fixations shown in blue. Here I keep the different
2829 configurations separate (2D, S3D shifted and native respectively).



2830
2831 **Fig 5.10.** Eyetracking fixations for an example participant in one clip. Red circles are centred
2832 on where the left eye fixated and have a radius relative to the length of that respective
2833 fixation. The same applies for the right eye with the blue circles.

2834 As can qualitatively be seen there does not appear to be a difference in the type and
 2835 position of fixations over the length of the clip. I opted not to include a pooled figure as the
 2836 number of fixations became too large to discern any patterns or indeed identify any circles
 2837 separately.

2838 I considered the number of fixations and average duration of fixations when watching 2D
 2839 compared to S3D. For this, after considering the subjective results from section 5.3.1 and
 2840 the eyetracking results thus far, I decided to combine the shifted and native S3D as S3D
 2841 data and the shifted and native 2D as 2D data. For each participant I therefore had two
 2842 separate repetitions from each participant for each clip. From these trials I calculated the
 2843 mean fixation duration and the total number of fixations in each. Fig 5.11. shows the mean
 2844 fixation duration averaged across participants and repetitions, error bars show the
 2845 standard deviations of the mean durations



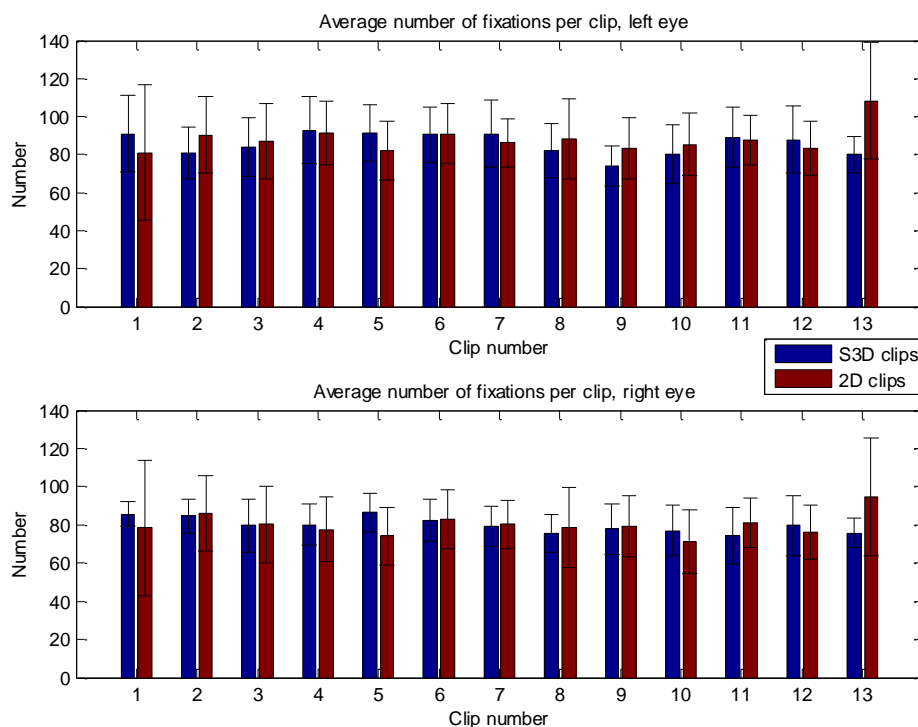
2846

2847 **Fig. 5.11.** Mean fixation duration for each clip, pooled across participants and repetitions.

2848 Blue data is for S3D and red data is for 2D clips. Errorbars show standard deviations of the

2849 mean clip durations. Each point represents 16 datapoints

2850 As can be seen there appears to be no difference between the 2D and S3D fixation
 2851 durations. This result is echoed in the number of fixations, shown below in fig. 5.12. where I
 2852 have conducted a similar analysis, taking the total number of fixations per trial and
 2853 averaging that over participant and repetition for each clip. As can be seen the driving
 2854 factor of the number of fixations appears to be the clip number (and hence the content
 2855 shown) rather than whether it was shown in 2D or S3D.



2856
 2857 **Fig. 5.12.** Mean number of fixations per trial, averaged across participants and repetitions,
 2858 for left and right eye. Blue bars represent S3D clips, red bars represent 2D clips. Each bar
 2859 represents 16 datapoints.

2860 I conducted a 2 X 2 repeated measures ANOVA analysis on the average duration of the clips
 2861 and the average number of fixations, averaging over clip and shift to establish a value for
 2862 fixation and duration for left and right eye, S3D and 2D respectively for each subject. The
 2863 results of this analysis are reported below in table 5.3. As can be seen, there is no

2864 significant effect of type of viewing (S3D or 2D), eye (left or right) or interaction on fixation
2865 duration and total number of fixations.

Fixation	Factor or Interaction	F	DF	P
Duration	Eye	0.251	8	0.632
	Viewing Type	0.285	8	0.610
	Interaction	1.209	8	0.308
Number	Eye	1.631	8	0.242
	Viewing Type	0.387	8	0.554
	Interaction	1.117	8	0.326

2866

2867 **Table 5.3.** 2 X 2 repeated measures ANOVA analysis on fixation duration and number, with
2868 eye (left and right) and viewing type (S3D and 2D) as factors. The interaction is also
2869 considered. Data recorded by the eyetracker.

2870 5.4 Discussion

2871 5.4.1 Subjective ratings

2872 My results confirm previous findings and support my prediction, showing,
2873 unsurprisingly, that viewers experience a more impressive illusion of depth with S3D as
2874 compared to 2D content. The results are very much in line with Bohr & Read 2014, who also
2875 used a 7-point Likert scale to investigate depth realism (their Figure 6), this time across
2876 different groups who viewed the film 'Toy Story' in either 2D or S3D (Read & Bohr, 2014).
2877 The mean 'depth realism' rating was 5.40 for their two S3D groups, compared to 4.26 for
2878 their three 2D groups ($P < 10^{-10}$, Mann-Whitney rank sum test; there were no significant
2879 differences within their S3D or 2D groups). The slightly smaller difference between 2D and
2880 S3D in Bohr & Read (2014) may reflect their between-subjects comparison; no participants

2881 had the opportunity to compare 2D and S3D directly as in the present study, where all
2882 participants viewed all clips in all conditions.

2883 I found no evidence that shifting 2D content behind the screen plane produces a stronger
2884 illusion of depth. Depth ratings were very slightly higher for shifted 2D content, and this
2885 difference might possibly have become significant if I had more statistical power. However,
2886 for practical purposes this is immaterial. The question being considered was whether the
2887 depth shift could simulate the depth of true stereoscopic 3D, and here the answer is clear: it
2888 does not come close. Even if the increase could be shown to be real by a more powerful
2889 study, it would still be too small to be of interest as a practical way of substituting for true
2890 S3D. Apparently, the binocular disparity cues indicating that the content is flat still dominate,
2891 even when the image is shifted behind the physical screen plane.

2892 This conclusion is necessarily limited to the particular clips I used. These were all taken from
2893 the same S3D programme 'Micro Monsters', they were all similar in nature (wildlife
2894 documentary), and they did not differ in the strength of the depth illusion they created. The
2895 logic of the shift manipulation is that weakening cues to the screen plane enables monocular
2896 depth cues to dominate perception (Ames Jr, 1925). This would predict that the effect of the
2897 shift should be stronger for content with more powerful monocular depth cues. As
2898 Koenderink et al write, "A photograph of a brick wall in frontoparallel attitude is not going to
2899 reveal any 'zograscopic effect'". More subtly, a 2D photograph of several frontoparallel
2900 surfaces at different distances may also display very little zograscopic effect, simply because
2901 the monocular depth cues are weak, even though the binocular disparity cues make the
2902 surfaces appear clearly separated in depth when viewed in S3D. Could this explain my
2903 results? The highly significant increase in depth ratings when the content was displayed in
2904 S3D proves that, unlike the brick wall example, my scenes did depict a wide range of depth,
2905 but I have not assessed their monocular depth cues objectively. The literature on 'monocular

2906 stereopsis' to date contains only cursory discussion of how the nature of the content might
2907 affect the strength of the effect (Koenderink et al., 2013). I cannot identify any particular
2908 reason why the content I used should be particularly ineffective at producing a 'zograscopic
2909 effect'. The literature reviewed in the Introduction used static images, which cannot contain
2910 depth cues such as structure-from-motion and looming, whereas my 'Micro Monsters' clips
2911 regularly contained these cues. Additionally, the clips were typically of insects filmed in
2912 extreme macro, meaning that they contained depth-of-field (blur) cues to three-dimensional
2913 structure. Rather than consisting of sets of frontoparallel surfaces with little depth structure
2914 within each surface, the clips typically depicted undergrowth, bark and so on extending in
2915 depth. Thus, while I cannot rule out that the shift manipulation would have produced a more
2916 compelling depth impression with other content, my chosen examples seem likely to have
2917 had monocular depth cues at least as strong as other commercial S3D content.

2918 The shift I applied may simply have been too small to produce the intended effect. My images
2919 were 48cm across and shifted by 0.75cm in each eye, resulting in a screen parallax of 1.44cm.
2920 This means that the binocular geometry specified the content as being at a viewing distance
2921 of 128cm, 28cm behind the physical screen plane at 100 cm. This is a substantial parallax,
2922 representing 3% of the image. BSKyB's Technical Guidelines for Plano Stereoscopic (3D)
2923 Programme Content (3D) specify that parallax behind the screen "should not exceed 2% for
2924 majority of shots" (the limit for parallax in front of the screen is even smaller at 1%). Thus,
2925 the parallax I applied is substantial by the standards of commercial S3D content and it is
2926 unlikely to be practical to apply larger amounts. However, Ames recommends a much larger
2927 disparity for images viewed at a distance of 100cm. To achieve this disparity, I would have to
2928 shift each eye's image 3cm on the screen, for a total parallax of 6cm or 13%. Thus, the most
2929 likely reason for my failure to see a 'zograscopic effect' is simply that much larger disparities
2930 are required. Prisms, as used by Ames, probably also distort the image and create much
2931 greater uncertainty about the location of the picture in space than in my set-up, where I

2932 retain all the usual cues indicating that the monitor is physically 100cm in front of the
2933 observer.

2934 Due to the content that we were showing (small insects, clips made available by BskyB) the
2935 footage was very close up and 'zoomed in', showing the focus of the clips a lot larger than
2936 they typically would be seen if in a real world environment. This resulted in close up shots
2937 with lots of depth, which was an intention of the footage, as disparity driven depth is greatest
2938 at short to medium distances. As chapter 4 has already concluded, familiar size is an
2939 important clue to depth, both relatively and absolutely. So there was potential that the clips
2940 may mislead participants on the amount of depth in the footage. I believe that, were this the
2941 case, our question removes the problem posed, as we were asking people to compare the
2942 clips on depth perception and immersion, and all the four manipulations of the clips
2943 contained the same content. It would be interesting to consider 'more natural' stimuli (i.e.
2944 stimuli not zoomed so far in).

2945 In summary, I have found no evidence that shifting 2D content behind the screen produces
2946 a depth illusion that is at all comparable to true S3D, at least not without the use of
2947 unacceptably large parallaxes. I conclude that the technique is not viable as a cheap way of
2948 making 'fake' S3D.

2949 [5.4.2 Eyetracking discussion](#)

2950 The eyetracking results support my prediction that gaze position and depth
2951 impression is very much content driven, and that fixations depend more on the content being
2952 shown than on the configuration (2D or S3D) that it is shown in.

2953 My results are necessarily limited to the content used in my study. In general, disparity-
2954 defined depth certainly can influence salience (Lang, Nguyen, Kattie, Yadati, Kankanhalli &
2955 Yan, 2012). For example Jansen et al. use natural disparity to create white and pink noise
2956 artificial stimuli with different depths, and found that participants fixated closer locations

2957 earlier than locations at a greater distance (Jansen, Onat & König, 2009). However, my study
2958 wished to focus on the 'real life' scenario and I used commercial content for this reason.

2959 The fixation analysis that was conducted suggested that the number of fixations and the
2960 duration of the fixations were very content driven. This again makes sense, as content where
2961 the focus of the scene moved would have less fixations than a scene where the shot was still
2962 and the content was centred around one point for an extended period of time. This analysis
2963 looked at the total number of fixations and the duration of these fixations and found no
2964 overall difference between S3D and 2D content. It would be interesting to consider the data
2965 as a time series and see if there was still no significant difference between the S3D and 2D
2966 versions of the clips for fixations. I predict that there would be no significant difference
2967 found, based on the results that I have gathered in this chapter.

2968 The clips that were used for this study, like typical commercial content, had many different
2969 cues to depth and many layers of colour and detail, not used in typical psychophysical or
2970 clinical studies. Thus, disparity was probably a less distinctive cue to salience than in a more
2971 impoverished stimulus. Furthermore, commercial content typically confounds disparity and
2972 other cues to salience, i.e. the director uses disparity as well as other cues such as blur and
2973 lighting to direct the viewer's attention to the object of interest. Both the author and
2974 participants, upon successfully completing the experiment, made the observation that the
2975 '3D' elements of the clips (i.e. the elements that were at a different depth to the screen
2976 plane) were mostly also elements that were the attentional focus point of the scene in
2977 question. This would obviously tend to reduce any difference between where people fixate
2978 in 2D versus S3D viewing of this content. I conclude that for typical commercial content, S3D
2979 does not tend to alter where people fixate in a scene.

2980 A potential limitation of the experiment is that despite a rigorous calibration and validation
2981 procedure measurements taken with an eyetracker can be subject to a large amount of

2982 noise. This is somewhat reflected in my results (as an example, one participant appeared to
2983 be verging to 10,000m away based on the parallax between the left and right eye positions,
2984 both horizontally and vertically). The vertical difference in left and right eye position was my
2985 first indication that the measurements were not going to be 100% accurate. However the
2986 Eyelink recommended levels were matched and the data did give us significance and a good
2987 indication of gaze, once enough data was collected. Once the technology is more refined, it
2988 would be interesting to revisit this study to see if the results hold with less noise.

2989 As this is the final chapter in my PhD thesis there are open questions left that could be
2990 considered. One of them has been touched on above, with analysis of the eyetracking data
2991 as a time series to consider if there is a significant difference between the S3D and 2D data.
2992 It would also be interesting to attempt to remove the influence of the content by considering
2993 artificially generated stimuli. This wasn't done here as I wished to consider the situation with
2994 content as accessible and similar to commercial content as possible. The impact of viewing
2995 distance could also have been considered, by changing and moving participants closer and
2996 further away, to consider whether the data was influenced by any accommodation cues to
2997 depth (these should be lessened at larger viewing distances).

2998 [5.5 Conclusion](#)

2999 Producers of S3D sometimes use a shortcut to mimic true S3D content by shifting 2D
3000 images behind the screen plane using a uniform disparity across the image, and rely on the
3001 monocular cues to depth to generate an immersive illusion. This is a poor substitute for
3002 commercial S3D content, as users note a significantly better quality of depth when true S3D
3003 content is used compared to the shifted 2D content. Eyetracking technology has indicated
3004 that, as expected, fixation length and population is very much content driven. In this study
3005 the introduction of S3D content did not significantly change where viewers attended to
3006 during the clips. This conclusion holds for typical commercial S3D content where disparity is

- 3007 generally used to reinforce 2D cues to salience, and might fail to hold for content where
- 3008 disparity and 2D cues to salience are different.

3009 6. Conclusion

3010 The overall goal of this thesis was to improve and build on the understanding of the
3011 factors affecting viewer experience in S3D cinema and television. S3D is a still growing area
3012 of interest that is being ever more refined and improved with the introduction of new
3013 technologies, such as autostereoscopic and volumetric displays. I conclude that there are
3014 subtle differences between S3D and 2D content that should be more carefully considered in
3015 the viewing of S3D media, such as the viewing angle and distance the content should be
3016 viewed at, with a potential to recommend that viewing angle should not exceed more than
3017 20° away from normal. Consideration should also be applied to the construction of S3D
3018 content, in particular the inter-axial value should be carefully considered in an effort to
3019 minimise the risk of the 'puppet theatre effect', despite the findings of this paper that
3020 suggest the familiar size of an object is considered preferably to the depth suggested by
3021 vergence eye movements. Clearly the introduction of disparity using S3D technology does
3022 bring a more immersive sense of depth to the viewer, as I have shown by comparing the
3023 native S3D content with 'fake' 3D generated by shifting a 2D image backwards using disparity
3024 and finding a significant difference between the impressions of depth for participants ($P <$
3025 10^{-17} , paired T-test analysis on 117 pairs of data). A consequence of this is that producers of
3026 S3D content should be aware of shifting the 2D image backwards, as the result is not as
3027 immersive as true, proper S3D content, and this might have a detrimental effect on the
3028 overall perception of how 'good' stereoscopic content is. More consideration should be
3029 given to attention and gaze position when viewing S3D content; an interesting avenue of
3030 research would be to build on my eyetracking study by also considering accommodation
3031 measurements, as well as vergence movements and gaze position, and to consider time
3032 series fixations, advancing further than the average duration and number of fixations as has
3033 been considered here.

3034 The studies conducted in these chapters can have different impacts depending on
3035 the way that the S3D content is being viewed (i.e. the medium used to view the content).
3036 There are many differences between watching S3D in the home compared to at the cinema,
3037 and my results from different chapters may impact on the viewing experience in different
3038 ways because of this. An obvious example is that in chapter 4, when considering the
3039 interaction of familiar size and vergence. In my study viewing distance was found to be a
3040 significant factor ($P=0$ in an omnibus ANOVA considering all factors and interactions),
3041 however the viewing distances I covered were 50cm, 100cm and 200cm. While chapter 2
3042 (television viewing distance in British households) would support these distances as being
3043 fairly typical to watch television on, this may not be, and typically is not, the case when
3044 considering a cinema auditorium, with most seats a greater distance than 200cm away from
3045 a screen. While the screen itself is bigger, and hence some of the interactions between
3046 familiar size and stereo will probably still have an effect, this increased viewing distance
3047 could have an impact on other cues to depth which may impact on experience (such as
3048 accommodation, a cue to depth that is considerably weaker as viewing distance increases).
3049 Some of the interactions between my studies would be interesting to consider. For example,
3050 I have concluded that viewing S3D from an oblique angle can have an adverse effect on what
3051 the participant or viewer perceives on a screen. In the eyetracking study (chapter 5) I had
3052 participants situated directly perpendicular to the screen. If the participant was to be moved
3053 to a more oblique angle would that have an impact on the fixations during the clip?
3054 Particularly considering that the fixations were very content driven? It would be an
3055 interesting question to consider in the future. Other potential confounds can exist when
3056 comparing S3D viewing in cinema or on a television screen in the home, such as distractions,
3057 and personal attitude. Typically viewers pay specifically to sit and watch content in a cinema,
3058 compared to watching the television as a pastime in the home. This small but important
3059 detail could have an impact, with viewers wanting to talk amongst themselves and

3060 potentially check other devices like smartphones (in which case the glasses may have an
3061 adverse effect) or to get up and do something else while the content is playing (which may
3062 impact the attention paid to the content).

3063 The results from this thesis contribute to the overarching study of the visual system by
3064 helping to further understand some of the intricacies involved in stereoscopic vision,
3065 particularly those associated with S3D technology and viewing an image on a flat viewing
3066 medium while disparity between the images is injected and manipulated using technology
3067 such as active and passive filters. The experiments help to further understand some of the
3068 cue conflicts that can occur while viewing S3D content generated by this introduction of
3069 illusory depth, such as the Subbuteo player effect. It also contributes to the understanding
3070 of the warping that can occur if the stereoscopic content is not viewed from a position
3071 desired by the producer of the S3D content in question. It would be interesting in future
3072 studies to consider the neural pathways and which areas of the visual cortex are influencing
3073 these perceptive phenomena. I believe the more advanced areas of the visual cortex would
3074 be the areas that would be of most interest to observe with stimuli similar to that described
3075 in this thesis.

3076 There are many results from this thesis that could be relevant and useful in the industry of
3077 S3D content production, and will also help to alleviate some of the concerns that are raised
3078 when considering S3D production. The eyetracking results suggest that introducing disparity
3079 to content does not in fact alter where the viewer fixates, and hence producers don't need
3080 to have special consideration when creating S3D content. It stands to reason that producers
3081 of content can use the same considerations that they use during creation of 2D material
3082 when creating scenes to attract attention. The subjective results from the same study
3083 suggest that it is a waste of producers' resources to replicate the 2D image in both eyes with
3084 a uniform disparity, as the artificial depth created does not generate a more immersive sense

3085 of depth than 2D content. If some scenes are displayed in 2D in an overall S3D production it
3086 would have as immersive an experience without altering the 2D images, and hence would be
3087 more economic. However the viewing distance that the content is shown at is a contributing
3088 factor to the viewers impression of size, confirming that the puppet theatre effect is an issue
3089 that producers of content need to consider when creating S3D content. A positive result from
3090 this study is that when considering the oblique angles chapter. Contrary to what may have
3091 been thought, the viewing angle of S3D content reveals no distortions for viewing angles of
3092 up to 20° from normal. And angles further than that have very similar distortions as that for 2D
3093 content, so the same considerations as applied to 2D content can be used, echoing the
3094 eyetracking study.

3095 Two of my experiments considered cue combinations in their analysis. In these I constructed
3096 models which combined the two different cues in a statistically sound fashion. The results
3097 from the fitted models suggested that the cues were combined optimally, echoing previous
3098 studies as completed by Ernst and Banks (2002). The cues in my modelling analyses weren't
3099 explicitly independent of one another as in other studies (i.e. both cues in both studies were
3100 visual cues, and hence couldn't be separated completely from one another, compared to e.g.
3101 haptic and visual cues). Hence some of the modelling was based on assumptions that the
3102 standard model (with independent cues) could be applied. I believe the results justify the
3103 assumptions as my models fit the data well, and further support the consensus in the cue
3104 combination framework that humans integrate cues in a way that is statistically optimal in
3105 detecting a correct signal if the cues provide different signal points, both for robust cues and
3106 smaller discrepancies.

3107 It is worth noting that S3D content in the cinema is still very popular, with most mainline
3108 cinema releases having an S3D release and a 2D release. The same cannot be said for
3109 television, and with the only S3D channels (SKY 3D, and before that, BBC 3D) being removed

3110 from streaming content directly, it could be that the home television medium for viewing
3111 S3D is not yet a viable option. It may improve with advancing technology and the creation
3112 and viewing of S3D content decreasing in price. The conclusions drawn from my thesis may
3113 provide some details to help improve the creation, production and viewing of S3D content.

3114 7. References

- 3115 3D, S. Bskyb technical guidelines for plano stereoscopic (3d) programme content: BSkyB.
- 3116 Ames Jr, A. (1925). The illusion of depth from single pictures. *J. Opt. Soc. Am*, 10, 137-148.
- 3117 Ardito, M. (1994). Studies of the influence of display size and picture brightness on the
3118 preferred viewing distance for HDTV programs. *SMPTE journal*, 103(8), 517-522.
- 3119 Azuma, R., Bailiot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent
3120 advances in augmented reality. *IEEE computer graphics and applications*, 21(6), 34-
3121 47.
- 3122 Banks, M. S., Held, R. T., & Girshick, A. R. (2009). Perception of 3-D Layout in Stereo Displays.
3123 *Inf Disp (1975)*, 25(1), 12-16.
- 3124 Banks, M. S., Read, J. C., Allison, R. S., & Watt, S. J. (2012). Stereoscopy and the Human Visual
3125 System. *SMPTE Motion Imaging J*, 121(4), 24-43. doi: 10.5594/j18173
- 3126 Bereby-Meyer, Y., Leiser, D., & Meyer, J. (1999). Perception of artificial stereoscopic stimuli
3127 from an incorrect viewing point. *Perception & Psychophysics*, 61(8), 1555-1563.
- 3128 Bishop, P. & Pettigrew, J. D. (1986). Neural mechanisms of binocular vision. *Vision Research*,
3129 26 (9): 1587–1600. doi:10.1016/0042-6989(86)90177-x
- 3130 Brady, M. P. (2009). Picture-Perfect HDTV. *Forbes*.
- 3131 Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433-436.
- 3132 Brenner, E., & van Damme, W. J. M. (1999). Perceived distance, shape and size. *Vision*
3133 *Research*, 39(5), 975-986. doi: [http://dx.doi.org/10.1016/S0042-6989\(98\)00162-X](http://dx.doi.org/10.1016/S0042-6989(98)00162-X)
- 3134 Brumm, K., Walenski, M., Haist, F., Robbins, S. L., Granet, D. B., & Love, T. (2010). Functional
3135 magnetic resonance imaging of a child with Alice in Wonderland syndrome during an
3136 episode of micropsia. *Journal of American Association for Pediatric Ophthalmology*
3137 *and Strabismus*, 14(4), 317-322. doi:
3138 <http://dx.doi.org/10.1016/j.jaapos.2010.03.007>

- 3139 Burks, R., Harper, C., & Bartha, M. C. (2014). Examining 3-D technologies in laptop displays.
3140 *Ergonomics in Design: The Quarterly of Human Factors Applications*, 22(3), 17-22.
- 3141 Cavanagh, P. (1987). Reconstructing the third dimension: Interactions between color,
3142 texture, motion, binocular disparity, and shape. *Computer Vision, Graphics, and*
3143 *Image Processing*, 37(2), 171-195. doi: [http://dx.doi.org/10.1016/S0734-](http://dx.doi.org/10.1016/S0734-189X(87)80001-4)
3144 [189X\(87\)80001-4](http://dx.doi.org/10.1016/S0734-189X(87)80001-4)
- 3145 Claparède, E. (1904). *Stéréoscopie monoculaire paradoxale*.
- 3146 Cooper, E. A., Piazza, E. A., & Banks, M. S. (2012). The perceptual basis of common
3147 photographic practice. *J Vis*, 12(5), 8.
- 3148 Cullen, B., Galperin, D., Collins, K., Kapralos, B., & Hogue, A. (2012). *The effects of audio on*
3149 *depth perception in S3D games*. Paper presented at the proceedings of the 7th audio
3150 mostly conference: a conference on interaction with sound.
- 3151 Cutting, J. E. (1987). Rigidity in cinema seen from the front row, side aisle. *J Exp Psychol Hum*
3152 *Percept Perform*, 13(3), 323-334.
- 3153 Da Vinci, L. (2012). *A treatise on painting*: Courier Dover Publications.
- 3154 Darlin, D. (2005). Falling Costs of Big-Screen TVs to Keep Falling, *NY Times*.
- 3155 De La Gournerie, J. (1859). *Traité de perspective linéaire contenant les tracés pour les*
3156 *tableaux, plans & courbes, les bas reliefs & les décorations théâtrales, avec une*
3157 *théorie des effets de perspective*: Dalmont et Dunod.
- 3158 Deregowski, J. B. (1969). A pictorial perception paradox. *Acta Psychol (Amst)*, 31(4), 365-374.
- 3159 Devernay, F., & Beardsley, P. (2010). Stereoscopic cinema *Image and Geometry processing*
3160 *for 3-D cinematography* (pp. 11-51): Springer.
- 3161 Di Luca, M. Domini & F. Caudek, C. (2006). Depth cues do not specify a unique Affine or
3162 Euclidean shape representation [Abstract]. *Journal of Vision*, 6(6):340, 340a
- 3163 Dodgson, N. A. (2004). *Variation and extrema of human interpupillary distance*. Paper
3164 presented at the Stereoscopic Displays and Virtual Reality Systems XI, San Jose, CA.

3165 Duane, A. (1900). The effect of converging prisms upon our notions of size and distance—an
3166 experimental study. *Ophthalmic Record*, 9, 595-607.

3167 Duchowski, A. T., Pelfrey, B., House, D. H., & Wang, R. (2011). *Measuring gaze depth with an*
3168 *eye tracker during stereoscopic display*. Paper presented at the Proceedings of the
3169 ACM SIGGRAPH symposium on applied perception in graphics and visualization.

3170 Earnshaw, R. A. (2014). *Virtual reality systems*: Academic press.

3171 Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a
3172 statistically optimal fashion. *Nature*, 415(6870), 429-433.

3173 Fernández-Caballero, A., López, M. T., & Saiz-Valverde, S. (2008). Dynamic stereoscopic
3174 selective visual attention (DSSVA): Integrating motion and shape with depth in video
3175 segmentation. *Expert Systems with Applications*, 34(2), 1394-1402.

3176 Foley, J. M. (1985). Binocular distance perception: egocentric distance tasks. *Journal of*
3177 *Experimental Psychology: Human Perception and Performance*, 11(2), 133.

3178 Foley, J. M. (1968). Depth, size and distance in stereoscopic vision. *Perception &*
3179 *Psychophysics*, 3(4), 265-274.

3180 Footprint, S. t. (2013) How big is a house? Average house size by country. from
3181 <http://shrinkthatfootprint.com/how-big-is-a-house>

3182 Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention.
3183 *Perception & Psychophysics*, 65(7), 999-1010.

3184 Fröhlich, B., Hochstrate, J., Hoffmann, J., Klüger, K., Blach, R., Bues, M., & Stefani, O. (2005).
3185 Implementing multi-viewer stereo displays.

3186 Gibson, J. J. (1950). *The perception of the visual world*.

3187 Golden, G. S. (1979). The Alice in Wonderland syndrome in juvenile migraine. *Pediatrics*,
3188 63(4), 517-519.

3189 Gombrich, E. H. (1972). *Art and illusion: a study in the psychology of pictorial representation*
3190 (4th ed.). London,: Phaldon.

3191 Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1974):
3192 Wiley New York.

3193 Gregory, R. L. (2015). *Eye and Brain: The Psychology of Seeing: The Psychology of Seeing*:
3194 Princeton university press.

3195 Hagen, M. A., & Elliott, H. B. (1976). An investigation of the relationship between viewing
3196 condition and preference for true and modified linear perspective and adults. *J Exp*
3197 *Psychol Hum Percept Perform*, 2(4), 479-490.

3198 Hagen, M. A., Elliott, H. B., & Jones, R. K. (1978). A distinctive characteristic of pictorial
3199 perception: the zoom effect. *Perception*, 7(6), 625-633.

3200 Hands, P., Khushu, A., & Read, J. C. A. (2014, 9-10 Dec. 2014). *Interaction between size and*
3201 *disparity cues in distance judgements*. Paper presented at the 3D Imaging (IC3D),
3202 2014 International Conference on.

3203 Hands, P., Khushu, A., & Read, J. C. (2015). The interaction between relative, familiar object
3204 size and binocular vision cues when perceiving stereoscopic 3D content. *J Vis*, 15(12),
3205 1082-1082.

3206 Hands, P., & Read, J. C. (2013). *Perceptual compensation mechanisms when viewing*
3207 *stereoscopic 3D from an oblique angle*. Paper presented at the 3D Imaging (IC3D),
3208 2013 International Conference on Liege, Belgium.
3209 <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6732086>

3210 Hands, P., Smulders, T. V., & Read, J. C. A. (2015). Stereoscopic 3-D content appears relatively
3211 veridical when viewed from an oblique angle. *J Vis*, 15(5), 6-6.

3212 Hands, P., & Read, J.C. (2015). *Television viewing distance in British households*. Paper
3213 presented at the 3D imaging (IC3D), 2015 International conference on Liege,
3214 Belgium.

3215 Held, R. T., & Banks, M. S. (2008). Misperceptions in Stereoscopic Displays: A Vision Science
3216 Perspective. *ACM Trans Graph*, 2008, 23-32. doi: 10.1145/1394281.1394285

- 3217 Hillaire, S., Lécuyer, A., Cozot, R., & Casiez, G. (2008). *Using an eye-tracking system to*
3218 *improve camera motions and depth-of-field blur effects in virtual environments.*
3219 Paper presented at the Virtual Reality Conference, 2008. VR'08. IEEE.
- 3220 Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity
3221 cues: Optimal cue combination. *Journal of vision*, 4(12), 1-1.
- 3222 Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence-accommodation
3223 conflicts hinder visual performance and cause visual fatigue. *J Vis*, 8(3), 33.31-30. doi:
3224 10.1167/8.3.33
- 3225 Hopf, K. (2000). An autostereoscopic display providing comfortable viewing conditions and a
3226 high degree of telepresence. *Circuits and Systems for Video Technology, IEEE*
3227 *Transactions on*, 10(3), 359-365. doi: 10.1109/76.836280
- 3228 Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*: Oxford University Press,
3229 USA.
- 3230 Howarth, P. A. (2011). Potential hazards of viewing 3D stereoscopic television and cinema: a
3231 review. *Ophthalmic & Physiological Optics*, 31, 111-122.
- 3232 Hudson, W. (1967). The study of the problem of pictorial perception among unacculturated
3233 groups. *International Journal of Psychology*, 2(2), 89-107.
- 3234 Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid
3235 scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*,
3236 20(11), 1254-1259.
- 3237 Jahoda, G., & McGurk, H. (1974a). Development of pictorial depth perception: cross-cultural
3238 replications. *Child Dev*, 45(4), 1042-1047.
- 3239 Jahoda, G., & McGurk, H. (1974b). Pictorial depth perception: a developmental study. *Br J*
3240 *Psychol*, 65(1), 141-149.
- 3241 Jansen, L., Onat, S., & König, P. (2009). Influence of disparity on fixation and saccades in free
3242 viewing of natural scenes. *Journal of Vision*, 9(1), 29-29.

- 3243 Julesz, B. (1971). Foundations of cyclopean perception.
- 3244 Julesz, B. (1986). Stereoscopic vision. *Vision Research*, 26(9), 1601-1612.
- 3245 Karajeh, H., Maqableh, M., & Masa'deh, R. e. (2014). A Review on Stereoscopic 3D: Home
3246 Entertainment for the Twenty First Century. *3D Research*, 5(4), 1-9.
- 3247 Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new
3248 in Psychtoolbox-3. *Perception*, 36(14), 1.
- 3249 Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict
3250 studies with stereoscopic and figure cues to slant. *Journal of vision*, 7(7), 5-5.
- 3251 Koenderink, J., Wijntjes, M., & van Doorn, A. (2013). Zograscopic viewing. *i-Perception*, 4(3),
3252 192-206.
- 3253 Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2004). Pointing out of the
3254 picture. *Perception*, 33, 513-530. doi: 10.1068/p3454
- 3255 Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. L. (1994). On so-called paradoxical
3256 monocular stereoscopy. *Perception*, 23(5), 583-594.
- 3257 Kubovy, M. (1988). *The Psychology of Perspective and Renaissance Art*: Cambridge University
3258 Press.
- 3259 Kuo, Y.-T., Chiu, N.-C., Shen, E.-Y., Ho, C.-S., & Wu, M.-C. (1998). Cerebral perfusion in
3260 children with Alice in Wonderland syndrome. *Pediatric Neurology*, 19(2), 105-108.
3261 doi: [http://dx.doi.org/10.1016/S0887-8994\(98\)00037-X](http://dx.doi.org/10.1016/S0887-8994(98)00037-X)
- 3262 Kurtz, H. F. (1937). Orthostereoscopy. *Journal of the Optical Society of America*, 27(10), 323-
3263 339.
- 3264 Lambooj, M., Ijsselsteijn, W., Fortuin, M., & Heynderickx, I. (2009). Visual Discomfort and
3265 Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Science and
3266 Technology*, 53(3), 1-14.

3267 Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling
3268 of depth cue combination: in defense of weak fusion. *Vision research*, 35(3), 389-
3269 412.

3270 Lang, C., Nguyen, T. V., Katti, H., Yadati, K., Kankanhalli, M., & Yan, S. (2012). Depth matters:
3271 Influence of depth cues on visual saliency. *Computer Vision - ECCV 2012*, 101 - 115

3272 Legge, G. E., & Yuanchao, G. (1989). Stereopsis and contrast. *Vision Research*, 29(8), 989-
3273 1004.

3274 Lund, A. M. (1993). The influence of video image size and resolution on viewing-distance
3275 preferences. *SMPTE journal*, 102(5), 406-415.

3276 Mayhew, J. E., & Longuet-Higgins, H. C. (1982). A computational model of binocular depth
3277 perception. *Nature*, 297(5865), 376-378.

3278 McCloy, R., & Stone, R. (2001). Virtual reality in surgery. *Bmj*, 323(7318), 912-915.

3279 McIntosh, R. D., & Lashley, G. (2008). Matching boxes: Familiar size influences action
3280 programming. *Neuropsychologia*, 46(9), 2441-2444.

3281 Meesters, L. M. J., Ijsselsteijn, W. A., & Seuntjens, P. J. H. (2004). A survey of perceptual
3282 evaluations and requirements of three-dimensional TV. *Circuits and Systems for*
3283 *Video Technology, IEEE Transactions on*, 14(3), 381-391. doi:
3284 10.1109/TCSVT.2004.823398

3285 Mendiburu, B. (2009). 3D movie making. *Focal Express*.

3286 Mishkin, M., Ungerleider, L. G. (1982) "Contribution of striate inputs to the visuospatial
3287 functions of parieto-preoccipital cortex in monkeys". *Behav. Brain Res.* 6 (1): 57-77.
3288 doi:10.1016/0166-4328(82)90081-X

3289 Mylonas, G. P., Darzi, A., & Yang, G.-Z. (2004). Gaze contingent depth recovery and motion
3290 stabilisation for minimally invasive robotic surgery *Medical Imaging and Augmented*
3291 *Reality* (pp. 311-319): Springer.

3292 Nave, R. (2014). Light and Vision. *Hyperphysics*

3293 Necker, L. A. (1832). Observations on some remarkable optical phaenomena seen in
3294 Switzerland; and on an optical phaenomenon which occurs on viewing a figure of a
3295 crystal or geometrical solid. *London and Edinburgh Philosophical Magazine and*
3296 *Journal of Science*, 1(5), 329-337.

3297 Noland, K., & Truong, L. (2015). A Survey of UK Television Viewing Conditions. BBC white
3298 paper.

3299 O’leary, A., & Wallach, H. (1980). Familiar size and linear perspective as distance cues in
3300 stereoscopic depth constancy. *Perception & Psychophysics*, 27(2), 131-135. doi:
3301 10.3758/BF03204300

3302 Ogle, K. N. (1938). Induced size effect I: A new phenomenon in binocular vision associated
3303 with the relative size of the images in the two eyes. *Archives of Ophthalmology*, 20,
3304 604.

3305 Olson, R. K., & Boswell, S. L. (1976). Pictorial depth sensitivity in two-year-old children. *Child*
3306 *Dev*, 47(4), 1175-1178.

3307 Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nat Rev Neurosci*,
3308 8(5), 379-391. doi: 10.1038/nrn2131

3309 Pastoor, S., & Wöpking, M. (1997). 3-D displays: A review of current technologies. *Displays*,
3310 17(2), 100-110.

3311 Pelli, D. G. (1997). The Video Toolbox softwar for visual psychophysics: transforming
3312 numbers into movies. *Spatial Vision*, 10(4), 437-442.

3313 Perkins, D. N. (1973). Compensating for Distortion in Viewing Pictures Obliquely. *Perception*
3314 *& Psychophysics*, 14(1), 13-18. doi: Doi 10.3758/Bf03198608

3315 Pirenne, M. H. (1970). *Optics, painting & photography*. London,: Cambridge U.P.

3316 Plumert, J. M., Kearney, J. K., Cremer, J. F., & Recker, K. (2005). Distance perception in real
3317 and virtual environments. *ACM Transactions on Applied Perception (TAP)*, 2(3), 216-
3318 233.

3319 Read, J. C. (2014). Viewer experience with stereoscopic 3D television in the home. *Displays*,
3320 *in press*.

3321 Read, J. C., & Bohr, I. (2014). User experience while viewing stereoscopic 3D television.
3322 *Ergonomics*, 57(8), 1140-1153. doi: 10.1080/00140139.2014.914581

3323 Read, J. C., Phillipson, G. P., & Glennerster, A. (2009). Latitude and longitude vertical
3324 disparities. *J Vis*, 9(13), 11 11-37. doi: 10.1167/9.13.11

3325 Read, J. C. A. (2014). Viewer experience with stereoscopic 3D television in the home.
3326 *Displays*, 35(5), 252-260.

3327 Ringach, D. L., Hawken, M. J., & Shapley, R. (1996). Binocular eye movements caused by the
3328 perception of three-dimensional structure from motion. *Vision Research*, 36(10),
3329 1479-1492.

3330 Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and
3331 binocular stereopsis. *Nature*, 361(6409), 253-255.

3332 Rosas, P., Wagemans, J., Ernst, M. O., & Wichmann, F. A. (2005). Texture and haptic cues in
3333 slant discrimination: reliability-based cue weighting without statistically optimal cue
3334 combination. *J Opt Soc Am A Opt Image Sci Vis*, 22(5), 801-809.

3335 Rosinski, R. R., Mulholland, T., Degelman, D., & Farber, J. (1980). Picture perception: an
3336 analysis of visual compensation. *Perception & Psychophysics*, 28(6), 521-526.

3337 Rushing, K. (2004). *Home Theatre Design*: Rockport Publishers.

3338 Rushton, S. K., & Riddell, P. M. (1999). Developing visual systems and exposure to virtual
3339 reality and stereo displays: some concerns and speculations about the demands on
3340 accommodation and vergence. *Applied Ergonomics*, 30(1), 69-78.

3341 Rushton, S. K., & Wann, J. P. (1999). Weighted combination of size and disparity: a
3342 computational model for timing a ball catch. *Nat Neurosci*, 2(2), 186-190. doi:
3343 10.1038/5750

- 3344 Saunders, J. A., & Backus, B. T. (2007). Both parallelism and orthogonality are used to
3345 perceive 3D slant of rectangles from 2D images. *J Vis*, 7(6). doi: 10.1167/7.6.7
- 3346 Scarfe, P., & Hibbard, P. B. (2011). Statistically optimal integration of biased sensory
3347 estimates. *J Vis*, 11(7). doi: 10.1167/11.7.12
- 3348 Schild, J., & Masuch, M. (2011). Fundamentals of stereoscopic 3d game design *Entertainment*
3349 *Computing–ICEC 2011* (pp. 155-160): Springer.
- 3350 Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011). The zone of comfort: Predicting
3351 visual discomfort with stereo displays. *J Vis*, 11(8), 1-29.
- 3352 Smith, M. D., & Malia, J. (2015). Controlling Miniaturization in Stereoscopic 3D Imagery.
3353 *SMPTE Motion Imaging J*, 124(3), 50-58.
- 3354 Snowden, R., Snowden, R. J., Thompson, P., & Troscianko, T. (2012). *Basic vision: an*
3355 *introduction to visual perception*: Oxford University Press.
- 3356 SurveyMonkey. Link to survey. from <https://www.surveymonkey.com/r/RZMNFHH>
- 3357 Tanton, N. (2004). Results of a survey on television viewing distance. BBC White paper.
- 3358 Thouless, R. H. (1931). Phenomenal Regression to the Real Object. *Nature*, 131, 261-263.
- 3359 THX recommended viewing distances. from [http://www.thx.com/test-bench-blog/what-](http://www.thx.com/test-bench-blog/what-size-tv-should-i-buy-tech-tips-from-thx/)
3360 [size-tv-should-i-buy-tech-tips-from-thx/](http://www.thx.com/test-bench-blog/what-size-tv-should-i-buy-tech-tips-from-thx/)
- 3361 Todd, J. (1955). Alice in Wonderland" syndrome. *Can Med Assoc J*, 73, 701-704.
- 3362 Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling
3363 visual attention via selective tuning. *Artificial intelligence*, 78(1), 507-545.
- 3364 Ullman, S. (1979). The interpretation of structure from motion. *Proc R Soc Lond B Biol Sci*,
3365 203(1153), 405-426.
- 3366 Verhoeff, F. H. (1935). A new theory of binocular vision. *Archives of Ophthalmology*, 13(2),
3367 151-175.
- 3368 Vinci, L. d. (2013). *A treatise on painting*: Read Books Ltd.

- 3369 Vishwanath, D., Girshick, A. R., & Banks, M. S. (2005). Why pictures look right when viewed
3370 from the wrong place. *Nature Neuroscience*, 8(10), 1401-1410.
- 3371 Vishwanath, D., & Hibbard, P. B. (2013). Seeing in 3-D with just one eye stereopsis without
3372 binocular vision. *Psychological science*, 0956797613477867.
- 3373 Wallach, H., Frey, K. J., & Bode, K. A. (1972). The nature of adaptation in distance perception
3374 based on oculomotor cues. *Perception & Psychophysics*, 11(1), 110-116.
- 3375 Walsh, V., & Kulikowski, J. (1998). *Perceptual constancy: Why things look as they do*:
3376 Cambridge University Press.
- 3377 Watt, S. J., Akeley, K., Girshick, A. R., & Banks, M. S. (2005). Achieving near-correct focus cues
3378 in a 3d display using multiple image planes. In *Electronic Imaging 2005* (pp. 393-401).
3379 International Society for Optics and Photonics.
- 3380 Welchman, A. E., Deubelius, A., Conrad, V., Bülthoff, H. H., & Kourtzi, Z. (2005). 3D shape
3381 perception from combined depth cues in human visual cortex. *Nat Neurosci*, 8(6),
3382 820-827.
- 3383 Wheatstone, C. (1838). Contributions to the Physiology of Vision. Part the First. On Some
3384 Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision.
3385 *Philosophical Transactions of the Royal Society of London*, 128, 371-394. doi:
3386 10.1098/rstl.1838.0019
- 3387 Wilks, S. S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing
3388 Composite Hypotheses. 60-62. doi: 10.1214/aoms/1177732360
- 3389 Woodgate, G. J., & Ezra, D. (1995). Autostereoscopic display device: Google Patents.
- 3390 Woods, A., Docherty, T., & Koch, R. (1993). Image Distortions in Stereoscopic Video Systems.
3391 *Stereoscopic Displays and Applications Iv, 1915*, 36-48. doi: Doi 10.1117/12.157041
- 3392 Woods, A. J. (1993). *Image distortions in stereoscopic video systems*. Paper presented at the
3393 SPIE 1915, Stereoscopic Displays and Applications IV, 36 (September 23, 1993).

- 3394 Woods, A. J. (2011). *How are crosstalk and ghosting defined in the stereoscopic literature?*
- 3395 Paper presented at the Stereoscopic Displays and Applications XXII, San Francisco
- 3396 Airport, California, USA.
- 3397 www.mathworks.com.
- 3398 Yamanoue, H., Okui, M., & Okano, F. (2006). Geometrical analysis of puppet-theater and
- 3399 cardboard effects in stereoscopic HDTV images. *Circuits and Systems for Video*
- 3400 *Technology, IEEE Transactions on*, 16(6), 744-752. doi: 10.1109/TCSVT.2006.875213
- 3401 Zone, R. (2007). *Stereoscopic cinema & the origins of 3D film*: University Press of Kentucky.
- 3402 Zorin, D., & Barr, A. H. (1995). *Correction of geometric perceptual distortions in pictures*.
- 3403 Paper presented at the Proceedings of the 22nd annual conference on Computer
- 3404 graphics and interactive techniques.
- 3405