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

List of tables and figures

1. Introduction

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

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

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

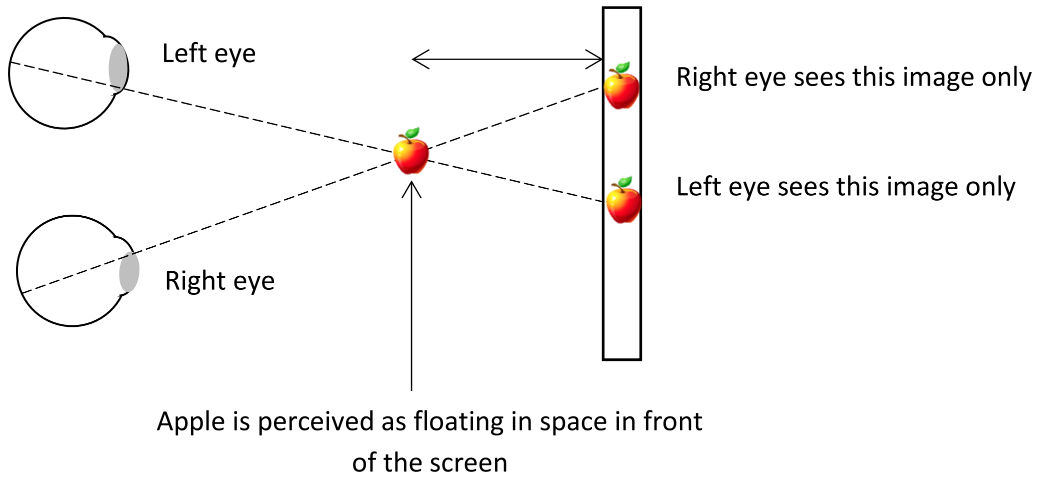


Fig. 1.1. *A diagram of the basic geometry of S3D displays and stereopsis. Using disparity, the apple can be perceived as floating in front of the screen. In this instance estimating the physical distance the apple is away from the viewer would be absolute disparity. Being able to ascertain (via disparity)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

1.1 S3D display technologies

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

Passive S3D displays show the left and right eye images at the same time on the screen. The two most common forms of passive display use colour anaglyph (red/blue or red/green) or polarising filter technology. With polarising filters, the images are typically presented row interleaved (i.e. the odd pixel rows show the left eye image and the even show the right eye, or vice versa), and a polarising filter is used to separate the two images and display only the desired pixel rows to each eye. Polarising filter is better for colour preservation but there is still a loss in vertical resolution with the row-interleaving technique.

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

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

A developing display technology is that of virtual reality (VR) and augmented reality (AR). Typically in these technologies the display is head mounted, and allows for a much more immersive experience. Many different elements of the display can be manipulated, such as using optics to choose at what depth to display the content, despite the screens being physically very close to the eyes of the viewer. Each viewer has their own display set while using head-mounted VR and AR, and this allows content producers to be confident on the viewing distance and angle the viewer is sat at. However this technology is not without flaws. It is still relatively expensive, although models such as the Oculus Rift are decreasing in price, and there are issues with lag and conflict with vestibular cues (with VR), which is a contributing factor to the regularly reported dizziness and motion sickness associated with these technologies (Azuma et al., 2001).

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

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

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

1.2 Filming S3D content

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

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

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

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

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

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

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

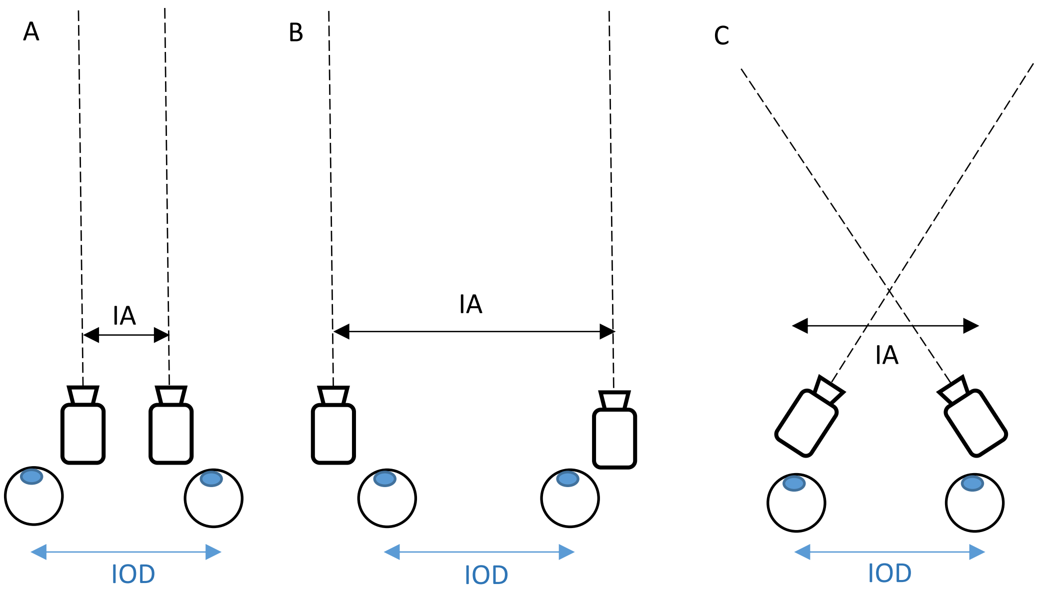


Fig. 1.3. *Explanatory diagram explaining the different camera configurations possible and the most likely perceptual outcome. I show a standard interocular distance (IOD) and change the interaxial (IA) distance. A) IA too small for IOD, which typically results in gigantism. B) IA too large for IOD,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

1.3 The different cues to depth

It is important to note that in the commercial world it is very rare for content to be created where stereoscopic binocular disparity is the only depth cue. There are almost always other depth cues, e.g. such as perspective, shading, texture, and motion parallax. However this is enough to generate a vivid, clear impression of change in depth of a 2D image, as shown using the random dot stereograms designed by Julesz (Julesz, 1986). In the more clinical and controlled lab environment that most experiments are conducted in, it is important to control for anything that could also affect the perception of the stimulus. Hence most experimental stimuli used in research are less like natural environments and very much controlled images, such as sine wave gratings (Legge & Yuanchao, 1989) and random dot stereograms. In these experiments participants are regularly positioned using something to limit any movement, such as a bite bar or a chin rest. This also enables the experiment to control for any motion parallax cues that might be available (Ames Jr, 1925). Pictorial cues, such as shading, texture and motion, are affine cues to depth, whereas cues to depth from binocular vision are metric, allowing the viewer (from the disparity information) to establish definitely a distance between two points in the picture, relative to the rest of the depth in the image. This isn't possible for affine cues from pictures, as the cues only provide a suggestion as to relative depths in the image scene, allowing the viewer to estimate, for example, local curvature of an object using shading (Di Luca, M., Domini, F. & Caudek, C., 2006). It is an important consideration that while disparity driven depth is considered metric, this is only in relation to the estimate of fixation depth that the user has at that time. Because of this the actual depth information provided by disparity is necessarily relative, and transformed into metric, based on this assumption of fixation depth (Foley, J. M., 1985). Humans typically underestimate depth (Plumert, J. et al., 2005) which would reduce the values of the estimated metric. It is because of the physical, measurable distances between objects in depth, provided by the horizontal offset in both eyes images, that S3D works so well in cinema and television. However this fixation distance judgement (and hence the resulting disparity driven metrics being estimated) could potentially lead to problems, such as incorrectly estimating the fixation distance. Hence the disparity driven metrics would also be incorrect, leaving the viewer perceiving something that doesn't look as intended by the producer of the S3D content (a good example of this is miniaturisation, which is studied in more depth in chapter 4). Depth from motion parallax is only defined up to scale unless the viewer knows for certain the distance that they (or the object) has moved, and is hence another form of relative depth perception.

Like binocular disparity, information regarding the depth of different objects can be detected directly using relative motion (Gibson, 1950). In this instance the geometry applied is exactly the same as stereopsis; however rather than considering two views at once the visual system considers two views at two different times. If these two cues to depth provide different depth information about the scene, some people may perceive the scene with the disparity defined depth; some may perceive the depth in the scene as defined by the relative motion cue, and some may perceive the depth in the scene as some combination of the two cues. This is a situation known as a cue conflict, where two elements provide different information to the brain (in this case specifically, the visual cortex). The way that the different cues to depth are combined in the visual cortex is an interesting area of research, and the mathematics behind it can be quite complex. A key consideration in this area of research is that conducted by Hillis et al. in which they establish that cues are combined in a statistically optimal fashion, in so much that as the reliability of the information provided by the cue decreases, the weighting of that cue in the combination also decreases (Hillis, J. M. et al., 2004). This study provided further support to the theory that cues are combined using a maximum likelihood estimating model, combining the information in an attempt to maximise the accuracy of the signal detection from the cues. A confounding factor of this theory is that reliability estimates of the cues are not constants, but vary according to the viewing geometry the observer has at that time. This means that as the viewing distance increases, the reliability of the different signals will not change in an identical fashion, resulting in a change in viewing distance changing the weighting of the two cues. Mathematically, in the context of depth, if you consider combining two different cues there are three outcomes that are possible: Each one of the two cues can be preferred over the other cue explicitly, or a weighted summation of the two cues occurs. In a weighted summation the explicit cue preference case is still possible, as the weighting for that cue is reduced to zero. If the discrepancies between the cues are large (known as robust cue integration) then the less reliable cue is vetoed, and effectively ignored. In the case of smaller cue combinations a weighted summation is optimised to get the most precise estimate to the depth from the cues provided. Research by Knill (2007) considered robust cue integration with a maximum likelihood estimator model and found that the while the less reliable cue's weighting does decrease significantly, it never has a weight of zero in the sum. This would question whether the cue can be completely ignored in combination of cues when the discrepancy is large (Knill, 2007), or whether a weighted summation would be a more accurate consideration, with the less reliable cues weight being considerably smaller than the more reliable cue. Ernst and Banks (2002) consider a combination of cues in their study (in their case the cues are visual and haptic cues). They found that the nervous system in general combines cues in a statistically optimal fashion, very much like the maximum likelihood estimating models considered above. The model attempts to minimise the variance by combining the variances from the two cues. This is directly transferable to combining the reliability of the cues to get the most precision in the estimate of the metric being considered. In cue combination usually the two cues can be approximated as coming from separate Gaussian distributions, in which the more reliable cue has a smaller variance (and hence standard deviation). The resulting combination of the cues assigns more weight to the consideration of that respective signal, so as to maximise the precision and minimise the variance. Fig. 1.4. below gives a brief example of the cue combination concept. In the top figure both Gaussian curves, intended to represent the different signals being detected, have equal variance and hence are equally reliable on the precision of the cue, and so the mean value that the cue combination model settles on (represented by the black dotted line) is directly in the middle of them. The combined Gaussian distribution is shown for the optimal signal estimation. In the bottom figure the means are the same as before but now the red curve has half the variance of the blue curve, and is hence considered the more accurate and reliable cue. Because of this more weight is assigned to the more reliable cue, and hence the estimate for the true value perceived is more similar to this cue's value. This results in the black dotted line moving closer to the more reliable signal. Landy et al. (1995) developed a model for cue combination known as modified weak fusion, based on the Bayesian theories of cue combination used elsewhere (Ernst and Banks, 2002). They consider three key issues of cue combination and cue conflict: Robustness, weighting and promotion of different depth cues (Landy, Maloney, Johnston & Young, 1995). The idea of combining cues in this manner, as well as the Bayesian technique, is an interesting one and is considered in chapters of this paper.

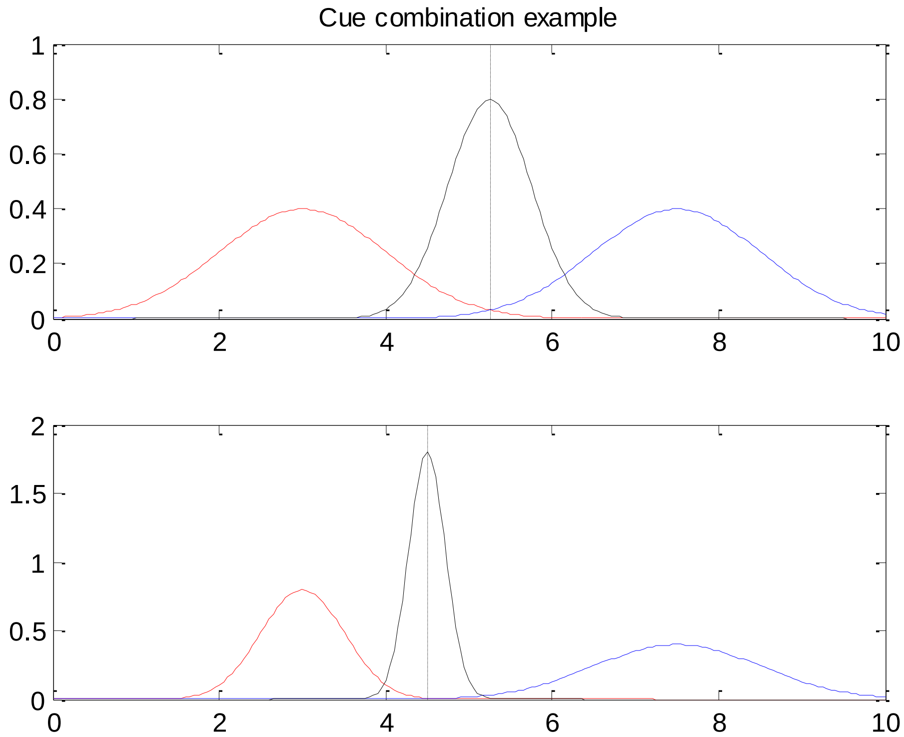


Fig 1.4. *Example figures to show cue combination optimisation. The red and blue curves are example Gaussian curves representing two different signals that are being detected. The width of the Gaussian (the variance associated with that Gaussian) is an indication of the reliability of the signal - a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

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

1.4 The vergence/accommodation conflict

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

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

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

In principle, when viewed with one eye through a pinhole from the centre of projection used to generate the image, a 2D picture could recreate the same retinal image as a real scene. However we never view 2D content in such a way. We are almost always viewing it from a position other than the centre of projection. Based on this it is surprising that we do not regularly perceive 2D pictures as distorted.

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

1.5 Aims

The overall aim of this thesis is to consider how different viewing conditions can affect the perception of S3D content, and hence improve understanding of the different factors that can affect viewer experience in S3D cinema and television. To achieve this, Chapter 2 provides a report on a survey conducted on 559 UK members of the public considering viewing habits in relation to viewing distance, and consumption of S3D content. This was published as (Hands & Read, 2015). Chapter 3 considers an experiment designed to assess the effect of oblique viewing of S3D content, and whether it can create geometrical distortions in the structure of the images perceived. This work was published as (Hands, Smulders & Read, 2015). Chapter 4 explores the interaction between vergence eye movements from disparity and perceived size, with application to understanding the "puppet theatre" effect that can be sometimes perceived in S3D content when the interaxial distance (the distance between the left and right eye cameras) exceeds the interocular distance. Producers sometimes attempt to 'fake' S3D by introducing a uniform parallax between previously identical left and right images, hence shifting 2D content backwards. This is because S3D content is expensive and difficult to generate, and producers therefore attempt to substitute true S3D, and instead show the 2D clips as if through a window, in the hope that this will be perceived as true S3D content. This work was presented at the Vision Sciences Society meeting 2016. Finally, I am interested in the effect that showing content in S3D may have on fixation. As an illustration it is possible that a particular region of the image is not salient in 2D, however, in S3D, disparity in the image may cause that region to be more likely to attract the viewer's attention. In Chapter 5, I investigate viewer experience with 'fake S3D' and show that while the impression of depth is marginally better than 2D content, it is no substitute for genuine S3D content, and I use eyetracking technology to examine whether viewers fixate on different video features when content is viewed in S3D as compared to 2D.

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

2.1. Introduction

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

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

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

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

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

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

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

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

Thus:

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

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

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

V = 2.5*D [Eq 2.6.]

implying an angle $\theta=20^{o}$.

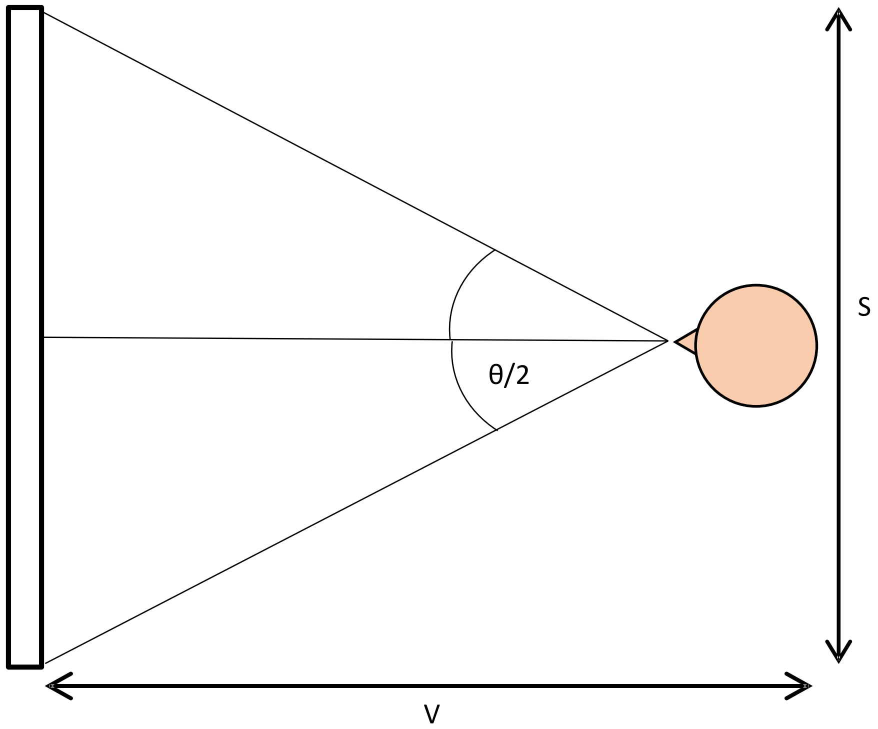


Fig. 2.1. *The relationship between viewing distance V and screen width S. The full visual angle subtended by the screen width is* ϑ.

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

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

$D = 0.615*V$ – SMPTE [Eq 2.8.]

$D = 0.746*V$ – THX [Eq 2.9.]

This implies that people should increase their viewing distance if they buy a larger television. Lab studies indicate that people do indeed prefer to view larger images from further away (Ardito, 1994; Cooper et al., 2012; Lund, 1993) and survey studies conducted by both Noland and Truong in 2015 and Tanton in 2004 seems to echo this (Noland & Truong, 2015; Tanton, 2004). Ardito considered the relationship between viewing distance and picture height, as reported subjectively by participants. He found that the viewing distance was preferred to be between 4 and 5 times the height of the image. This is considerably larger than the relationships now recommended, which could be a reflection of how much screen sizes have grown since the study was conducted in 1994. Lund conducted a similar study, to establish a relationship, and found similar results. Both results found relationships that were hyperbolic and non-linear, which raises the question as to why the guidelines are still linear in structure. It could possibly be in an attempt to make the advice more understandable for the average consumer. Viewing distance is often constrained by the size of the living room, especially in British houses which tend to be small compared to US ones (Footprint, 2013). Thus, it seems unlikely that real viewers are following these recommendations.

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

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

2.2. Materials and methods

2.2.1. Participants

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

2.2.2. Procedure

Participants were asked to complete a survey of 10 questions, based on their general viewing habits and S3D consumption. These questions are shown below, with a brief 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.

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

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

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

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

2.3. Results

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

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

2.3.1 S3D viewing

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

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

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

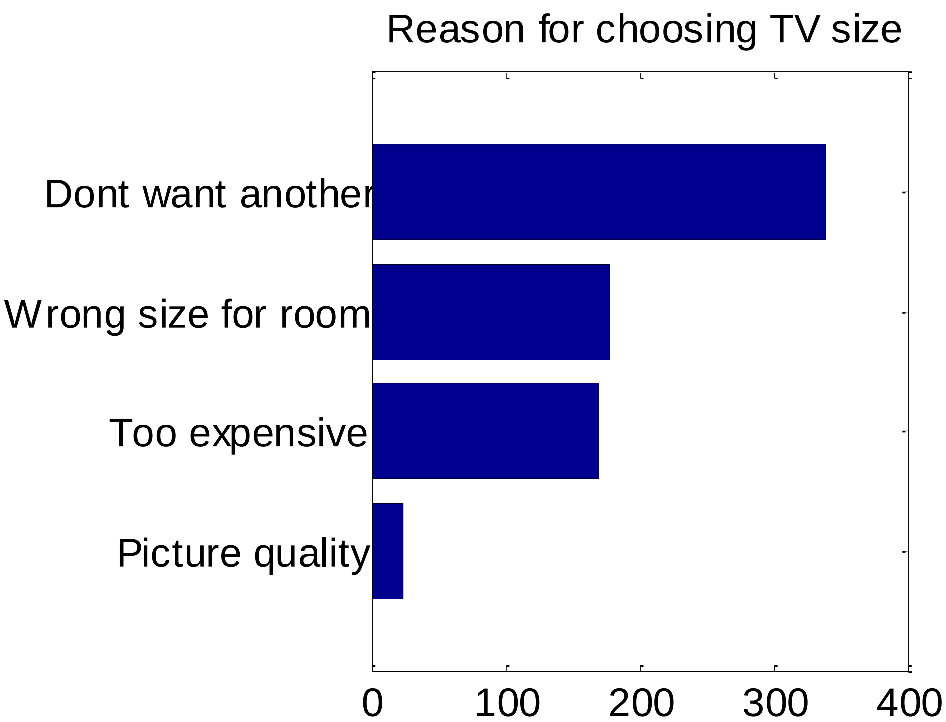


Fig. 2.2. *Histogram of responses to question number 3.*

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

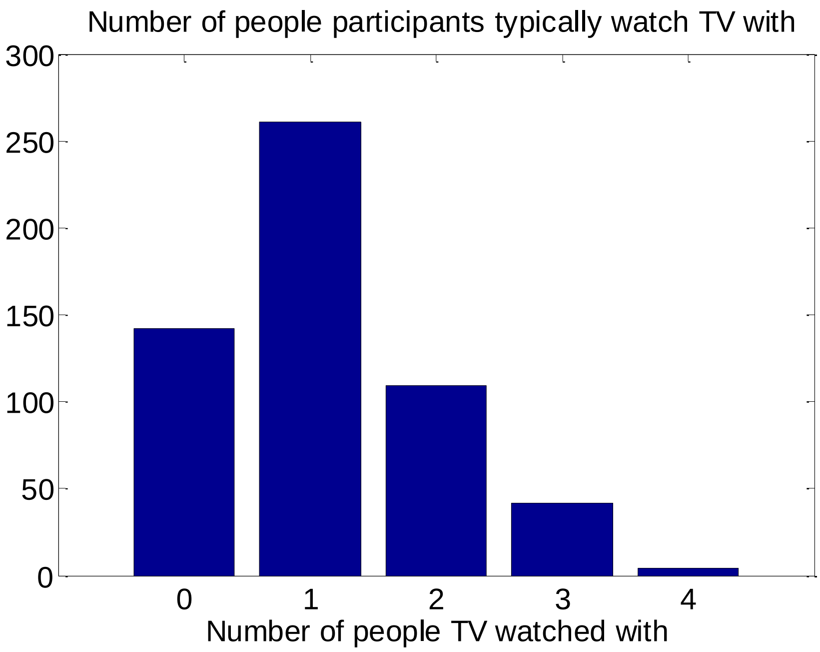


Fig. 2.3. *Histogram to show the number of people participants typically watched television with. As can be seen the most responses are from the lower numbers, with 1 other person being the most common number to watch television with.*

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

2.3.2 Viewing distance and screen size

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

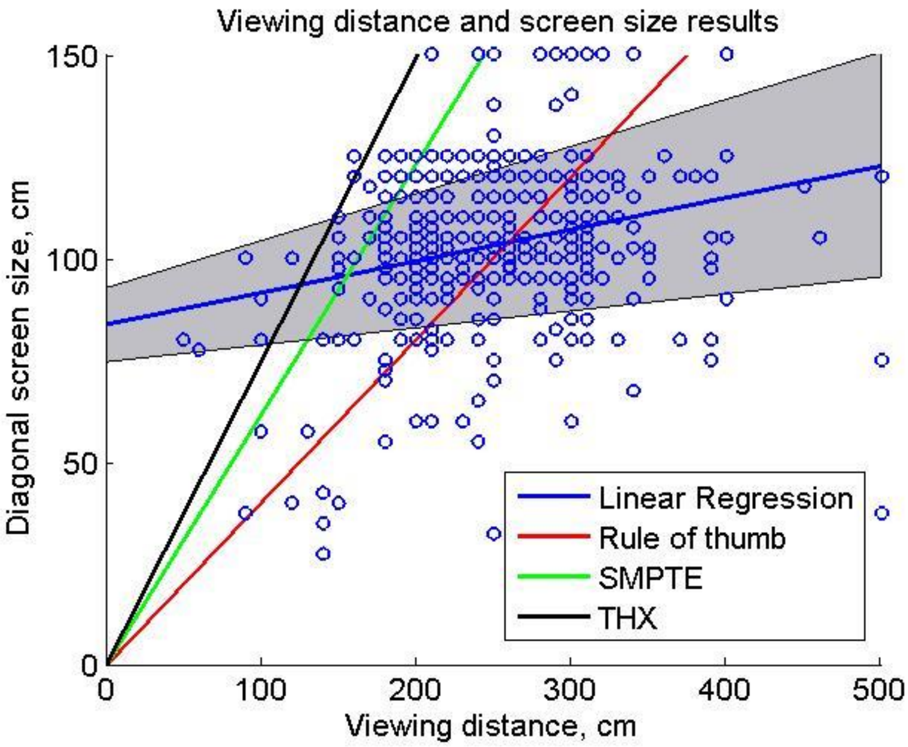


Fig. 2.4. *Scatter graph of responses from n = 558 participants. The horizontal axis is the reported distance from the television that participants sit, and the vertical axis is the reported diagonal screen size of the television, both in cm. Also shown are the regression line of best fit, and the three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

The fitted regression line is:

$$D = 0.077*V + 83.705 \qquad \text{[Eq 2.10.]}$$

This curve has a significantly shallower slope than the 3 recommended relationships (the 95% confidence interval for the slope were 0.038 to 0.114, much lower than the lowest recommended slope of 0.4 from ROT) and also has an intercept, which none of the other three guidelines have. It is possible that the guidelines don't have an intercept because the guidelines would then be more complicated for consumers to implement themselves. Fig. 2.5. shows the same fitted regression forced to have no intercept. As can be seen the regression line is very similar to the rule of thumb, and the confidence interval is considerably smaller.

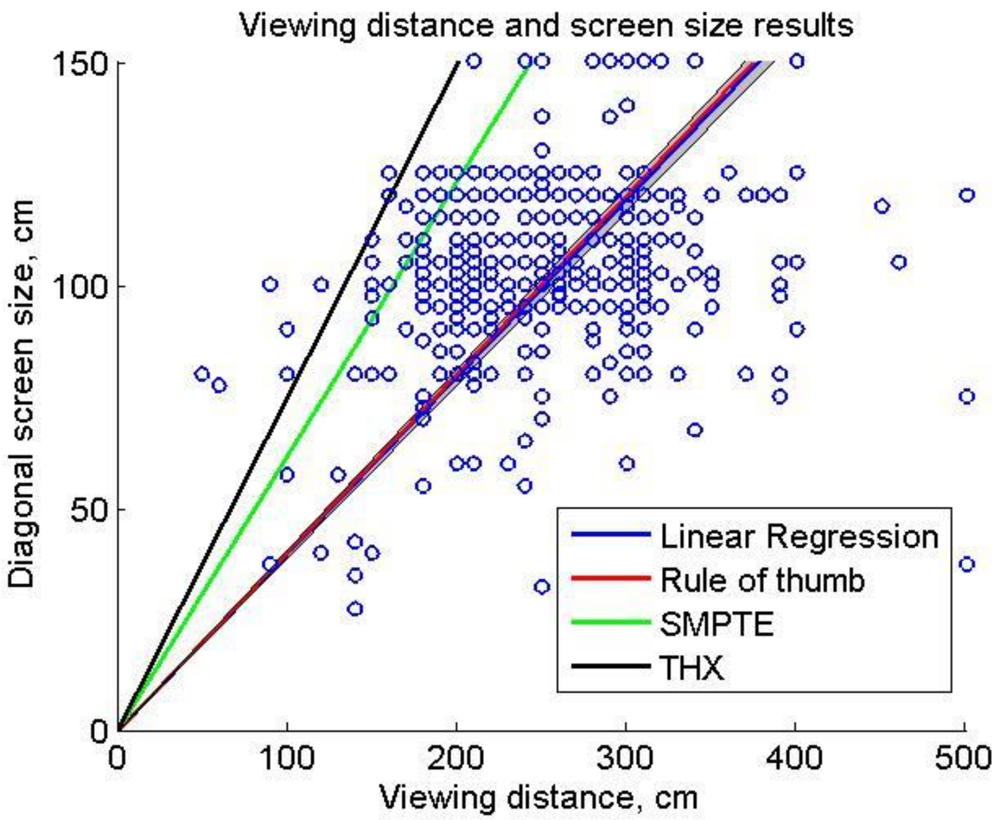


Fig. 2.5. *Results plotted as for Fig. 2.4. but now the regression line is forced to have no intercept. As can be seen the linear regression and the rule of thumb are very similar in slope.*

Viewing distance was only weakly correlated with screen size (R = 0.2483, P < 10^{-8}, Pearson's correlation coefficient), suggesting that the relationship between viewing distance and screen size is not an important consideration when buying a television. This correlation increased when only the subset of participants who owned an S3D television was considered (R = 0.4576, P < 10^{-4}). This subset is shown below in Figs. 2.6. and 2.7. As can be seen the subset has not got as many larger outliers, and hence the correlation is more positive. The confidence intervals for the regression lines are larger, due to a smaller population. The values selected suggest that those with S3D televisions purchase larger televisions and sit further away than those with televisions that do not have S3D capacity. One way this could be interpreted is that people who are more affluent can afford to purchase the larger, 3D capable televisions to put in their houses.

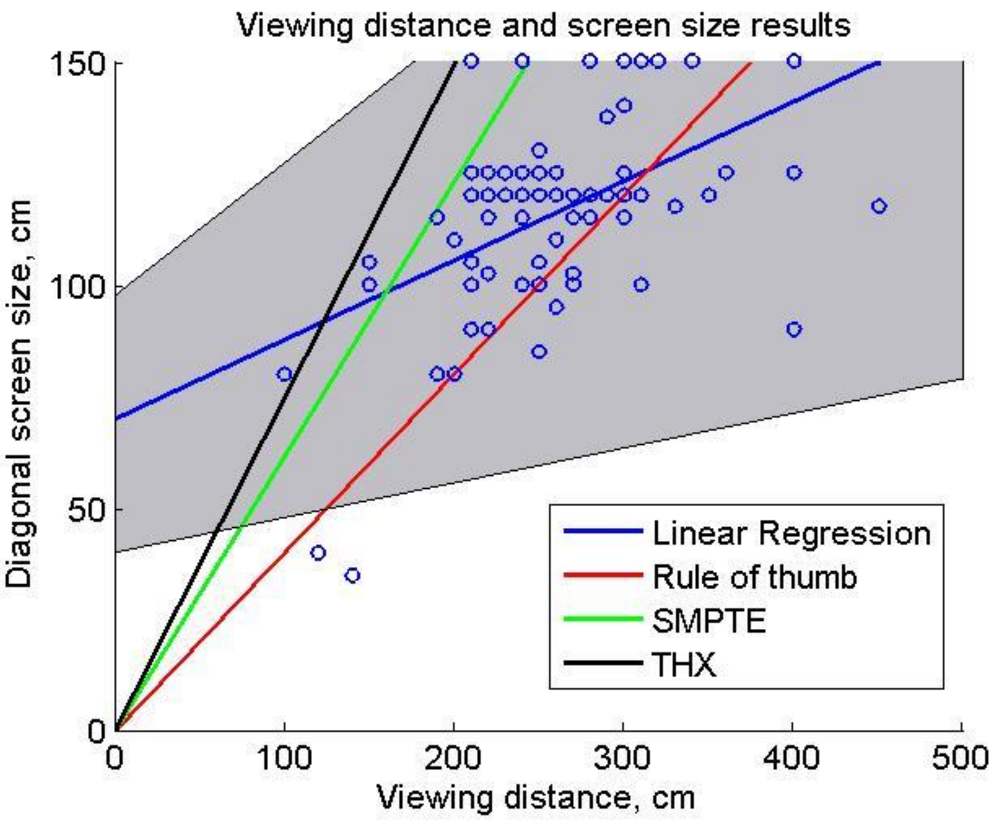


Fig. 2.6. *Scatter graph of responses from n = 86 participants whom owned an S3D capable television. The horizontal axis is the reported distance from the television that participants sit, and the vertical axis is the reported diagonal screen size of the television, both in cm. Also shown are the regression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

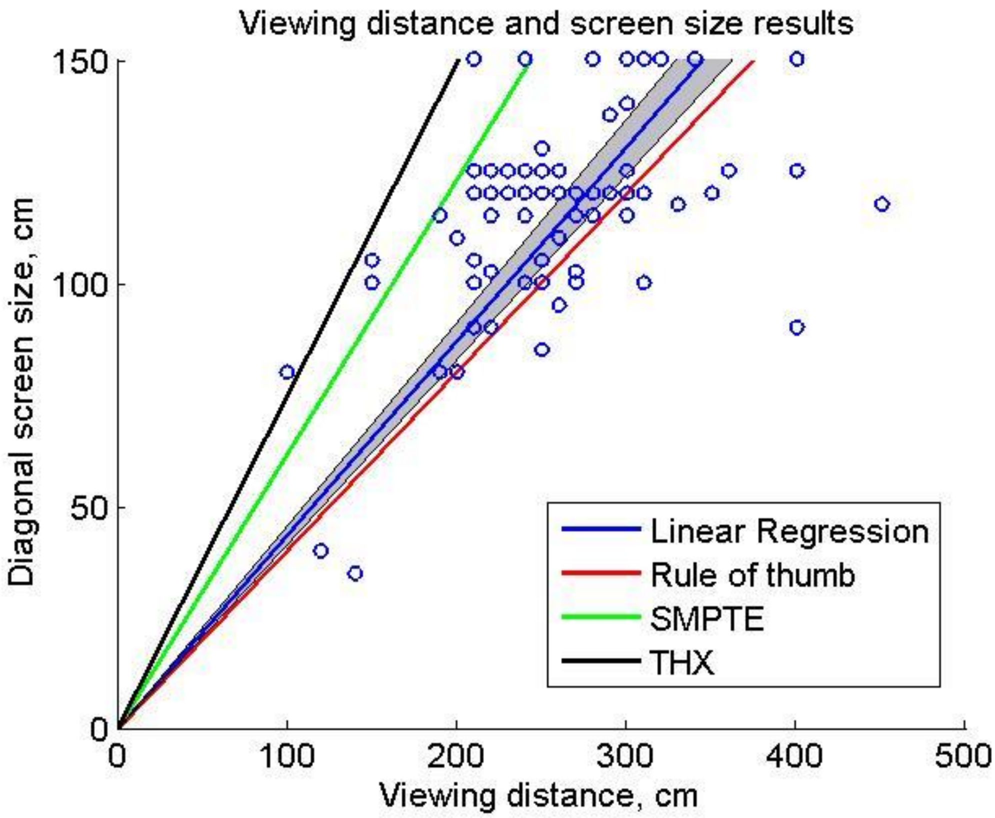


Fig. 2.7. *Results plotted as for Fig. 2.6. but now the regression line is forced to have no intercept. As can be seen the linear regression and the rule of thumb are still very similar in slope.*

There was a significant correlation between viewing distance and screen size. This further increased for the subset of participants in the study that had an S3D capable television. This suggests that viewers, whether consciously or subconsciously, consider the space available for them when they choose what size television to purchase, i.e. they either buy a television that is suitable for the size of the room, or possibly rearrange furniture to accommodate as larger a viewing distance as possible, or both. While the correlation values are only weak, there is clearly a relationship between screen size and viewing distance. If the intercept is forced to be zero then the rule of thumb appears to be the guideline most consistently followed by participants in this experiment. However the fit of the data is much better if a non-zero intercept is permitted, suggesting that in reality the increase of the screen size with viewing distance is much less steep than any of the guidelines specify. The limited range of television sizes (mainly 80-120cm) and viewing distances (mainly 150-300cm) mean that most viewers end up sitting at viewing distances broadly consistent with all the guidelines (Figs. 2.4 and 2.5.). Indeed in feedback once the survey was completed participants regularly stated they did not know the guidelines even existed. This could be a reflection on the fact that it appears the recommendations made by THX and SMPTE are successful in capturing people's instinctive preference about where is the best place to sit with regards to screen size.

2.4. Discussion

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

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

A further reason could in principle be that it is harder to watch S3D in groups. For an optimal experience, the viewer must be seated at correct positions for both viewing distance and viewing angle (Hands & Read, 2013), and the need to wear S3D technology glasses means only a set number of people can watch the S3D content. However, in my survey most people reported that they usually viewed television with just 1 or 2 other people. This suggests that the number of glasses that come with S3D televisions should not be a limiting factor. Additionally, it should not be hard for two or even three people to find a good viewing angle to watch S3D together. The need to wear S3D glasses, however, may be off-putting in itself, compared to sitting down to watch regular 2D television content. Further study is needed to establish what factors limit the popularity of S3D content.

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

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

An interesting observation from Ardito and Lund was that the relationship between screen size and viewing distance was not a linear one, as I have modelled with regards to my data and the guidelines recommended by THX, SMPTE and the rule of thumb. Indeed, Ardito suggests the relationship is hyperbolic, while Lund simply states the relationship is decidedly not linear. I wished to consider the guidelines in place already, and because they are linear I opted to force my results to be linear also. There is an important distinction in the studies also, as the experiments by Lund and Ardito were constructed in a lab and participants changed the screen size until happy, compared to the method of survey to record already established viewing habits. It would be interesting in a further study to consider if a more complicated model could better encapsulate the data, while still being accessible to consumers. Fig. 2.8. shows my data along with the fits from both Ardito and Lund, instead of from my own regression. The equation Ardito derived was

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

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

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

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

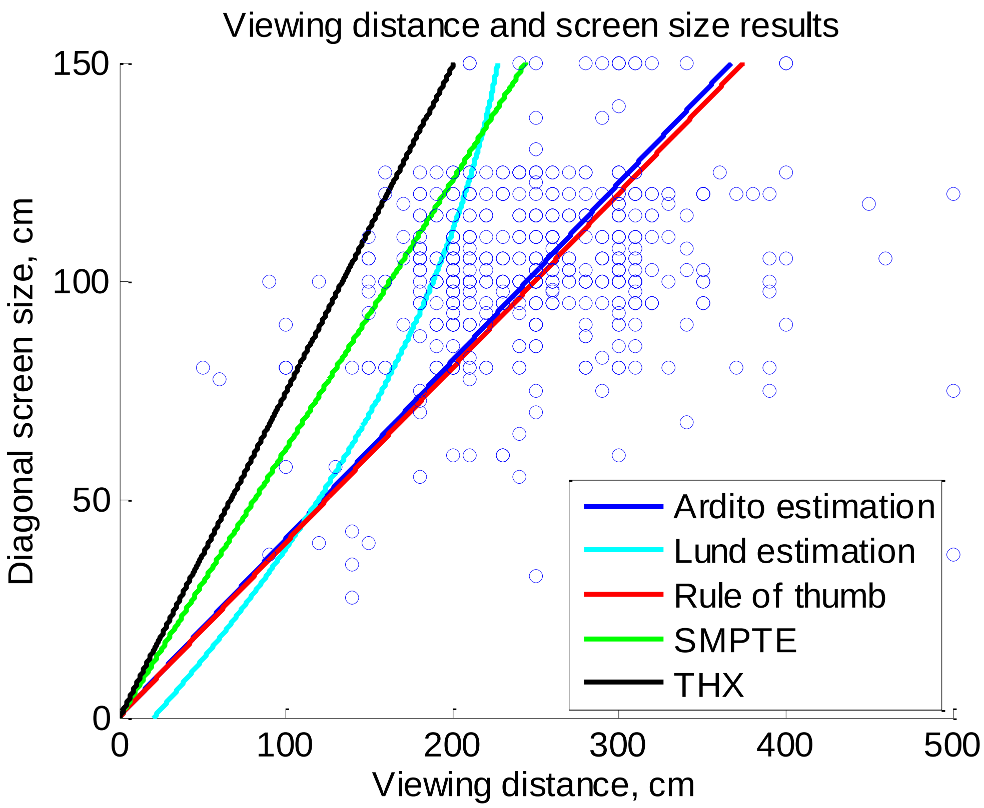


Fig. 2.8. *Datapoints from N = 558 participants along with the fitted equations from Ardito (1994) and Lund (1993), and the recommended guidelines. Ardito has a linear relationship very similar to that of the rule of thumb (blue and red lines respectively), while Lund's equation is hyperbolic (cyan line).*

A further potential issue that this paper faces is that while the overall number of participants was high and there were many surveys collected, only a very small subsection (n = 5) watched S3D content regularly. This might question the confidence in the results based on this subsection. A potential way to avoid this would be to target S3D viewers specifically (e.g. to wait outside an S3D movie theatre to ask the questions). However this would put bias into the survey, and I wanted to assess the prevalence of S3D viewing in a general population. A potential other solution would be to increase the total survey size to try to get an increased population of S3D viewers. Unfortunately, if only 0.9% of the population regularly watch S3D content, to get a larger subset I would need a considerably larger population. As an illustration to get 20 participants reporting regular S3D viewing I would need to question over 2200 participants.

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

2.5. Conclusion

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

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

3.1 Introduction

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

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

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

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

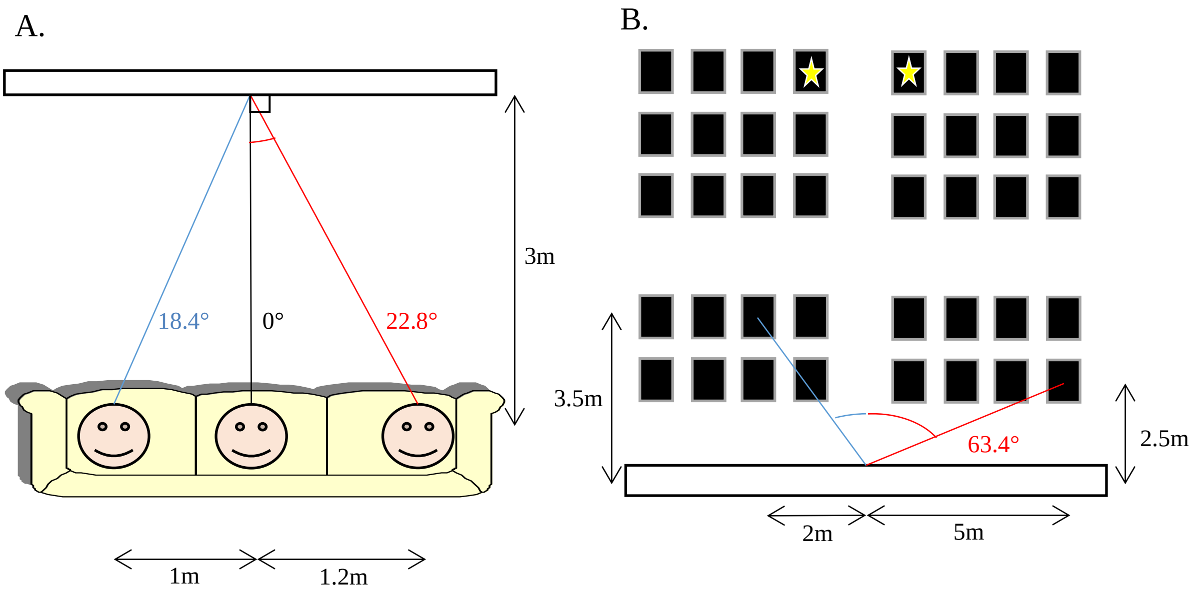


Fig. 3.1. *(A) A diagram to illustrate how viewing angles can change. If three people sit in the home watching television together, 3 metres away from the screen and between 1 and 1.2 metres apart, the viewing angle can be up to 23°. Angles are measured from perpendicular viewing (here I show a top down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

Of course, these problems also apply to 2D images, in the sense that the image projected onto the retina varies as a function of viewing angle. The problem of why, nevertheless, images appear veridical from a range of viewing angles has fascinated researchers since the Renaissance (Kubovy, 1988; Pirenne, 1970). Several factors seem to contribute. One is that humans are simply not very sensitive to the distortion introduced by oblique viewing (Cutting, 1987; Gombrich, 1972). Additionally, images usually depict familiar objects, so the viewer's perceptions can be influenced by their expectations (Thouless, 1931). However, it is also clear that observers are capable of compensating for the oblique viewing, so that perception is based not on the image actually projected onto the retina, but on the image which would have been seen if viewed perpendicularly (Perkins, 1973; Rosinski, Mulholland, Degelman, & Farber, 1980; Vishwanath, Girshick, & Banks, 2005). This compensation could work by recovering the true centre of projection and reinterpreting the retinal image accordingly. The true centre of projection could be estimated from cues present within the depicted scene (De La Gournerie, 1859; Saunders & Backus, 2007), such as the location of vanishing points, and/or from external cues regarding the orientation of the picture plane combined with simplifying assumptions such that the true centre of projection lies on the central surface normal. Presumably, such a mechanism would have to reflect experience with 2D pictures (Deregowski, 1969; Jahoda & McGurk, 1974a, 1974b; Olson & Boswell, 1976). Vishwanath et al. (2005) have recently argued for a simpler heuristic, whereby the retinal image is reinterpreted locally based on local surface slant. They argue that this may reflect a more general heuristic which is useful when interacting with real objects viewed obliquely, not a specific mechanism for interpreting pictures. External cues to local surface slant include binocular disparity, vergence, accommodation, the position of specular highlights relative to external light sources, and perspective cues provided by a frame surrounding the screen plane. If you consider the perceptual mechanisms involved it would make sense that the initial slant and orientation of the picture impact heavily on the compensation mechanism and are a cue used. These lines are the initial deciphering of the visual scene in the visual cortex of V1, and hence it would be consistent with the deeper areas of the neural pathway deciding 'what' the image was showing after the angular orientation of the viewing medium has been calculated and compensated for. Accordingly, occluding the frame of the display, viewing monocularly or viewing through a pinhole all tend to make the compensation less effective, so that images appear warped when viewed at oblique angles (Bereby-Meyer, Leiser, & Meyer, 1999; Perkins, 1973; Vishwanath et al., 2005).

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

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

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

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

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

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

3.2 Material and methods

3.2.1 Participants

Participants were recruited via an internal volunteer scheme at Newcastle University Institute of Neuroscience, on the basis that they had no visual problems other than wearing glasses or contact lenses. The work was approved by Newcastle University Faculty of Medical Sciences Ethics Committee. Ten participants (9 female, 8 naïve; 1 male, PH) were used in the study. Only one voluntary participant and the author took part in both the previous study (Hands & Read, 2013) and the current one due to availability of the other participants. Naïve participants were not informed of the experimental aims or hypotheses, but due to the random order of blocks will have been able to work out that the viewing angle was changing. Participants were paid £10 for completing the study.

3.2.2 Apparatus

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

Participants sat at a viewing distance of 120cm, measured perpendicularly from the center of the screen to the midpoint of the eyes, with their eyes at the same height as the center of the screen. They wore passive 3D glasses throughout the experiment, enabling us to interleave S3D, 2D and monocular stimuli. The monitor sat on a turntable, which allowed it to be accurately rotated between ±45° about a vertical axis passing through the midline of the screen. I define the viewing angle, θ_{view}, to be the angle between the plane normal to the screen and the viewer's line of sight to the center of the screen (Fig 3.2A). In different experimental blocks, the turntable was rotated so that angle θ_{view} was either 0^{o}, -45^{o} (closer to the viewer on their right) or $+20^{o}$. It was convenient to alter the viewing angle by moving the display screen rather than the participant (see Fig 3.2A). A chinrest was used to ensure the subject's eyes were at the correct position, and the chair was adjustable to ensure the participant was comfortable. In some experimental blocks, a fabric curtain with a hole was pulled across which occluded all four screen edges from the participant's view, while allowing them to see the stimuli.

3.2.3 Stimulus generation

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

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

In a previous study (Hands & Read, 2013) using wire-frame cubes, I wrote my own Matlab software to calculate where to render the vertices of each cube. I checked my calculations by drawing a square onto a sheet of acetate and mounted it on a sheet of Perspex in front of the screen, representing one face of the virtual cube. I supplied my code with the physical coordinates of this square, and rendered it for different viewing angles. I verified that, in each case, the image drawn on the screen lined up with the physical square drawn on the acetate, confirming that my code was rendering the virtual code correctly, whether the screen was perpendicular to the viewer or viewed obliquely at the specified angle. In the experiments reported here, I used Psychtoolbox with the OpenGL library to draw solid, textured cubes. I confirmed that this produced the same vertex positions by using Psychtoolbox to draw dots on top of the rendered cubes at the locations of the vertices as calculated by my own code, and checking that these dots lay on the vertices of the rendered cubes.

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

3.2.4 Experimental design

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

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

In each trial, one of the two cubes was rendered for frontoparallel viewing in the normal way, i.e. for a line of sight perpendicular to the screen. The other was rendered for an oblique viewing angle that varied between θ_{rend}=-45° and θ_{rend}=+45°. I will refer to these as the "normal-rendered" and "obliquely-rendered" cube, respectively. When $\theta_{rend}=\theta_{view}$, the obliquely-rendered cube was rendered for the actual viewing angle of the participant. I will refer to this as "geometrically correct". In the S3D condition, the geometrically-correct stimulus is orthostereoscopic, i.e. each eye ideally saw the retinal image which would have been projected by a physical cube in front of the viewer, apart from accommodation effects.

On each trial, the orientation of each cube was random: each virtual cube was rotated through a random angle about all three axes in succession before being rendered.

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

In static trials both objects remained stationary on the screen; in motion trials, both objects rotated at a constant speed of 18 deg/s about all three axes. This rotation speed was chosen as being slow enough to be comfortable for the participant to follow, while fast enough to produce rapid changes in the on-screen image and thus powerful structure-from-motion cues (Ullman, 1979). In both types of trials, the objects remained on screen until the participant indicated whether the top or bottom object appeared most cube-like.

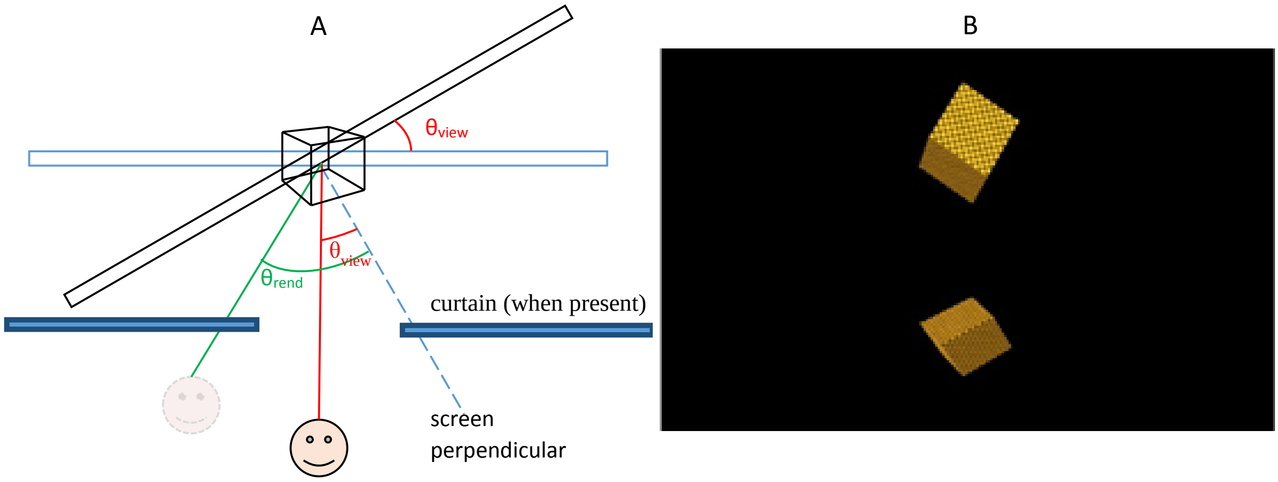


Fig. 3.2. *Experimental setup seen from above (A) and example stimulus (B). A: The heads of two possible viewers are sketched. The head drawn with solid lines represents the actual position of the participant, whose line of sight to the screen (red line) is at angle* θ_{view} *to the perpendicular (blue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

3.2.5 Experimental parameters

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

1. The angle θ_{rend} used to project the obliquely-rendered cube (8 possible values: ±45°, ±35°, ±20° and ±10°; see Fig 3.3B)

2. Whether the normal-rendered cube was at top or bottom of the screen (2 possible values)

3. Object motion (2 possible values: static or rotating)

4. Binocularity (4 possible values: S3D (binocular; each eye sees a different image), B2D (binocular; each eye sees the same image on the screen) or M2D (monocular, left or right)).

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

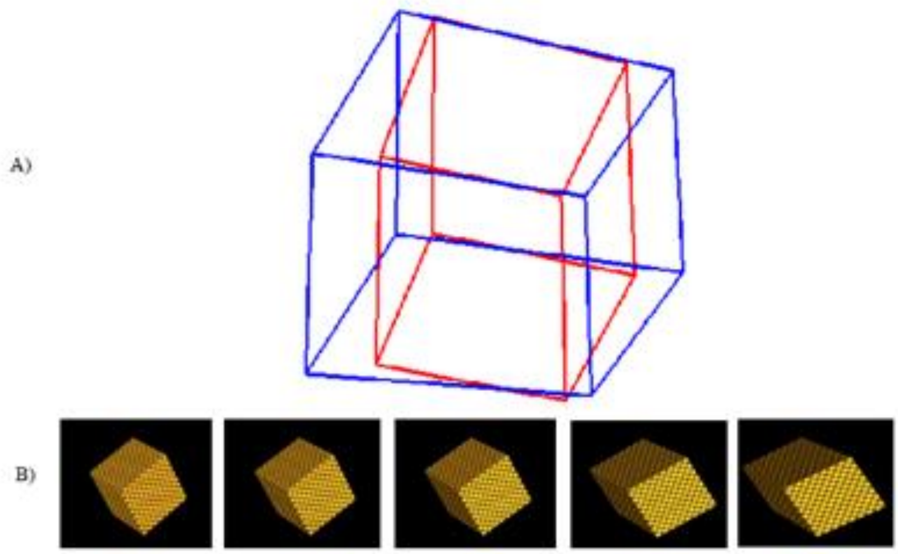


Fig. 3.3. *(A) Rendered image on screen for two different render angles. The two cubes are projections from exactly the same virtual cube in space, but the red cube is rendered for perpendicular viewing while the blue cube is rendered for a viewing angle of 45°. If you view the image with one eye from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

3.2.6 Modelling

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

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

where the free parameters *s* and *r* determine each factor's sensitivity to θ_{rend}, and *A* and *B* determine the relative weight of each factor. *A* is the weight given to normal rendering, and *B* the weight given to geometrical correctness. In my experiments, one of the cubes was always rendered for perpendicular viewing, θ_{rend}=0°. The difference in "perceived veridicality" between this normal-rendered cube and the obliquely-rendered cube is therefore

$$\Delta V = A - Aexp\left(-\frac{\theta_{rend}^2}{2s^2}\right) + Bexp\left(-\frac{\theta_{view}^2}{2r^2}\right) - Bexp\left(-\frac{[\theta_{rend}-\theta_{view}]^2}{2r^2}\right).$$

[Eq3.2]

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

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

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

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

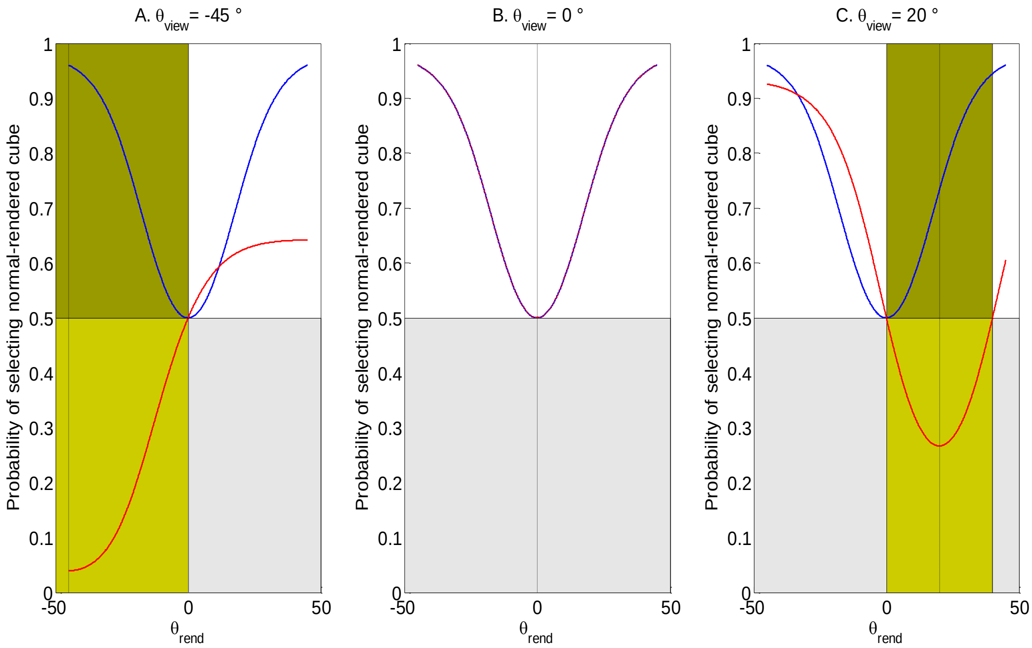


Fig. 3.4. *Model predictions with perfect compensation for oblique viewing (blue) and no compensation (red). Curves show probability that the model selects the normal-rendered cube as being more veridical, plotted as a function of render angle θ_{rend} for 3 different viewing angles θ_{view}. Model parameters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

3.2.7 Fitting

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

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

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

3.3 Results

Figs 3.5 and 3.6 show the proportion of trials on which the normal-rendered cube was selected as being "more cube-like", pooled over all observers. I plot this as a function of θ_{rend}, the viewing angle for which the obliquely-rendered cube was drawn (Fig 3.2A). For θ_{rend}=0, both cubes would be rendered for perpendicular viewing, so performance would necessarily be at chance. Figs 3.5 and 3.6 show results for the frame-visible and frame-occluded conditions, respectively. The three panels in each row of Figs 3.5 and 3.6 show results for the 3 different viewing angles, θ_{view}. The different colours and symbols show different binocularity conditions: red squares = binocular viewing in stereoscopic 3D; blue triangles = binocular viewing in 2D (same image on screen for both left and right eyes); green disks = monocular viewing (pooled left and right monocular results). The upper panels (ABC) show data for rotating stimuli and the lower (DEF) for static.

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

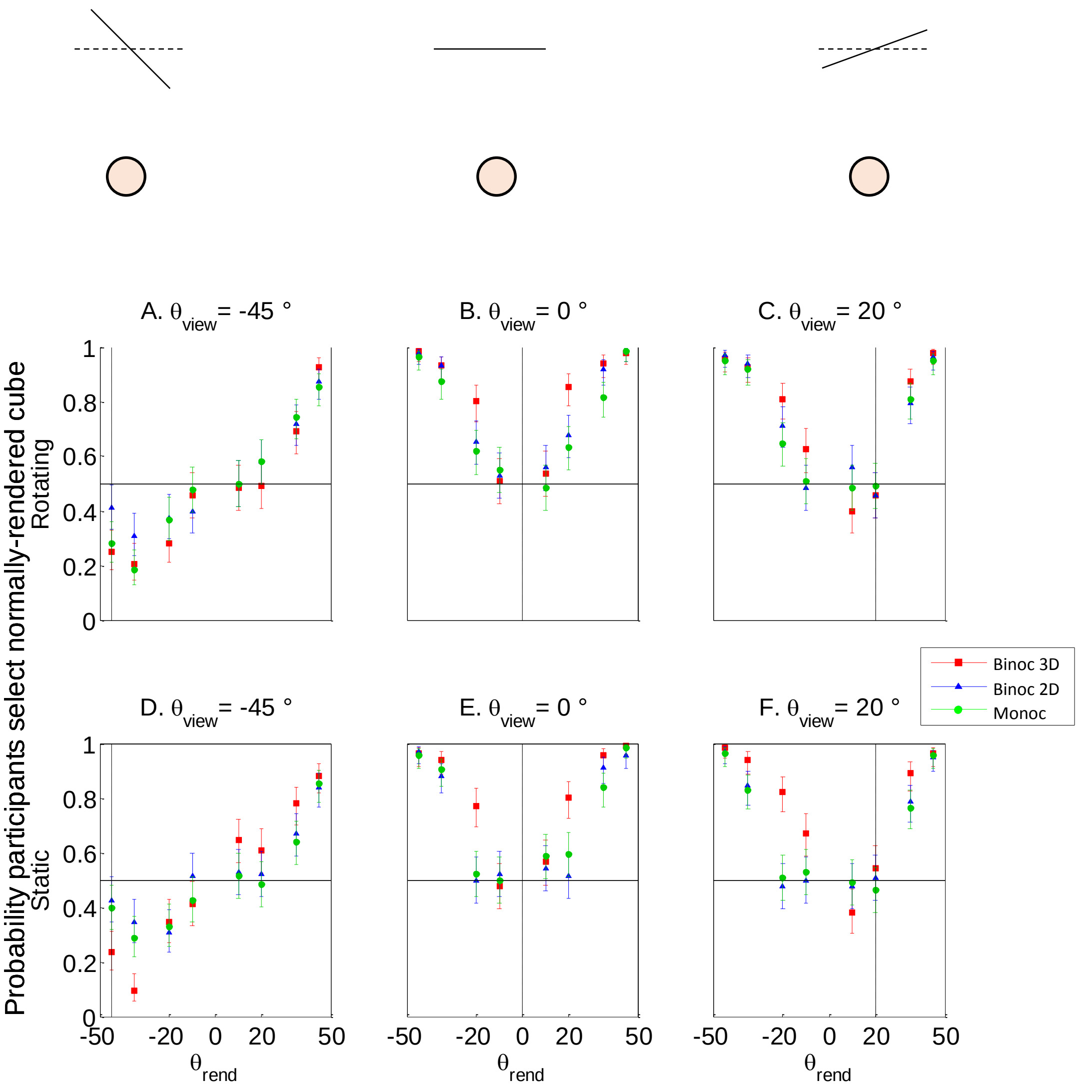


Fig. 3.5. *Data from the frame-visible condition. The vertical axis displays the proportion of trials on which subjects reported the normal-rendered cube as appearing "more cube-like". Results are plotted as a function of* θ_{rend}*, the viewing angle for which the obliquely-rendered cube was drawn.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

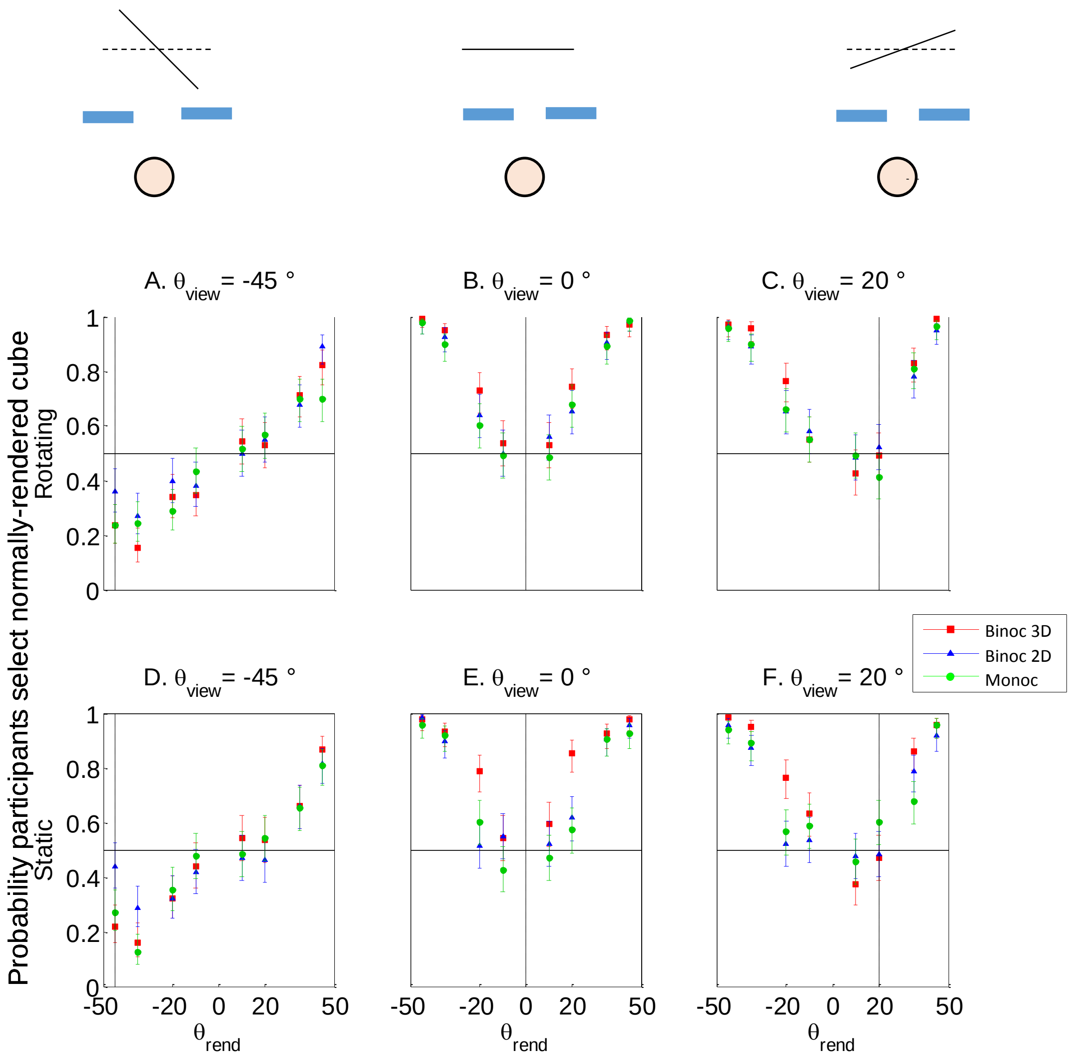


Fig. 3.6. *Frame-occluded data, presented as described in Fig. 3.5.*

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

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

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

Figs 3.7. and 3.8. have the model fitted over the data. The model is fitted to all participants' results pooled together. If objects look veridical when rendered for normal viewing, even when viewed obliquely, data-points should lie above this line. If objects look veridical when they are geometrically correct on the retina, where data-points should lie depends on rendering and viewing angle. The white regions in each panel show where the normal-rendered cube is closer than the obliquely-rendered cube to being geometrically correct for the particular viewing angle. Here the normal-rendered cube should look more veridical, so subjects should select it whenever they can detect a difference between the two render angles (probability ≥0.5). The fact that data-points do lie in the white regions, rather than in the gray regions below them, confirms this, but does not enable us to distinguish between a preference for normal rendering and a preference for geometrical correctness.

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

3.3.1 Statistical analysis

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

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

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

In the above statistical analysis I considered all the data collected. This makes it difficult to assess the effect of different factors on the two different components identified in my model. As argued above, my data imply that two factors affect whether an object appears distorted: whether it is geometrically correct on the retina ($\theta_{rend}=\theta_{view}$), but also whether it would be correct if viewed perpendicularly ($\theta_{rend}=0^o$). Much of my data confound these two effects, because often, both factors imply that the user should select the normal-rendered cube. This situation corresponds to the white regions in Figs 3.5 and 3.6. To assess how the different experimental conditions (occlusion, binocularity, rotation) affected the competition between the two model components, I also repeated this statistical analysis using only data where the two components pulled in opposite directions, i.e. the yellow regions in Figs 3.5 and 3.6. Here there is no overlap in the values of θ_{view} and θ_{rend} so the statistical significance of θ_{view} cannot be determined. I therefore only consider the main factor influences and the interactions between frame occlusion/visibility, binocularity, rotation and θ_{rend}. Table 3.2 shows the main effects and interaction terms for these 4 factors.

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

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

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

3.3.2 Sensitivity to rendering angle θ_{rend}

I first consider the central panel, Fig 3.7B, where θ_{view}=0°, i.e. the screen was frontoparallel in the usual way. If θ_{rend}=0°, both cubes would have the same projection, so performance would be at chance. As the obliquely-rendered cube is drawn at ever more extreme angles, it appears progressively more distorted and subjects become more likely to choose the normal-rendered cube. The rendering angle θ_{rend} is significant when considering only this subset of the data (χ^2 = 42080.1, $P < 0.0005$). In agreement with previous studies (Cutting, 1987), subjects are fairly insensitive to incorrect rendering. At $|\theta_{rend}|=10^o$, results do not differ significantly from chance for any binocularity conditions (95% confidence intervals in Fig 3.7 overlap chance). Even when θ_{view} is as large as 20^o, the results are not significantly different from chance for a static cube viewed without S3D. For a rotating cube, or a static cube viewed in S3D, subjects are significantly more likely to choose the normal-rendered cube, but do so only about 75% of the time. Even when the obliquely-rendered cube is drawn for a viewing angle as extreme as 45^o, subjects still choose it as being "more cube-like" on nearly 10% of trials when viewing a static cube in 2D. This is surprising, given that a rendering angle of $\theta_{rend}=45^o$ produces a very different image on the screen from one of 0^o (Fig 3.3A).

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

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

However, oblique viewing clearly had a strong effect on perception, even when the retinal image had been designed to take oblique viewing into account. For example in Fig 3.7A, the viewing angle was $\theta_{view}=-45^o$. Thus at $\theta_{rend}=-45^o$, the obliquely-rendered cube produced the geometrically correct image of a cube on the retina, whereas the normal-rendered cube was distorted. Fig 3.5B shows subjects are quite capable of detecting a 45° error in rendering angle when the display is frontoparallel: they reject the erroneous rendering over 80% of the time. However, when viewing obliquely at $\theta_{view}=-45^\circ$ (Fig 3.5A), subjects did not show a comparably strong preference for the geometrically correct cube: they chose it only 25% of the time for the S3D stimulus at $\theta_{rend}=-45^\circ$, while for the 2D stimuli, they picked both cubes equally often. This cannot be explained simply by a lack of sensitivity to distortion (Cutting, 1987; Gombrich, 1972), but must reflect a mechanism favoring normal rendering.

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

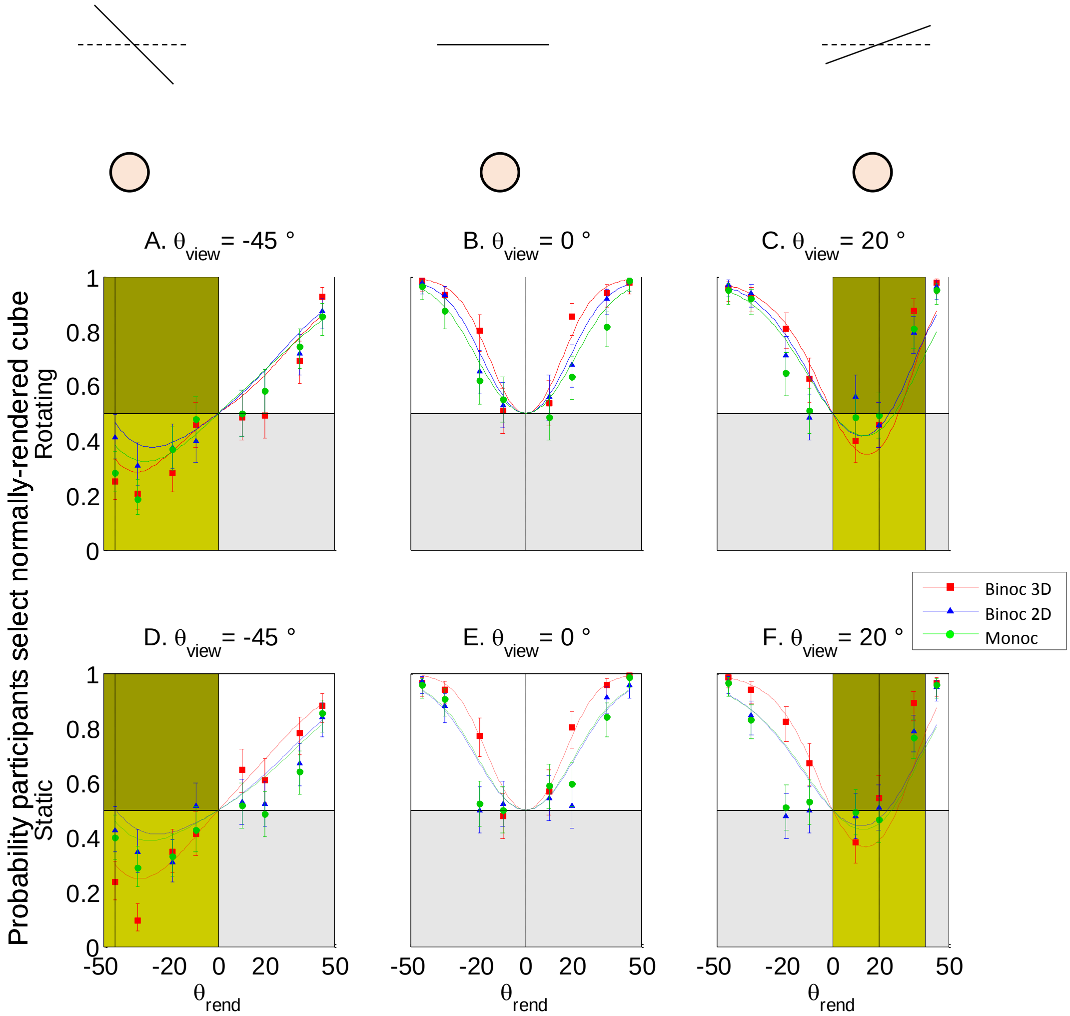


Fig. 3.7. *Results from the frame-visible condition as in Fig. 3.5, with the model fitted to the data. The model is fitted to all participants as a pooled group. The top row (solid lines) are for trials with rotating cubes, the bottom row (dashed lines) are for trials with static cubes. Shading as in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

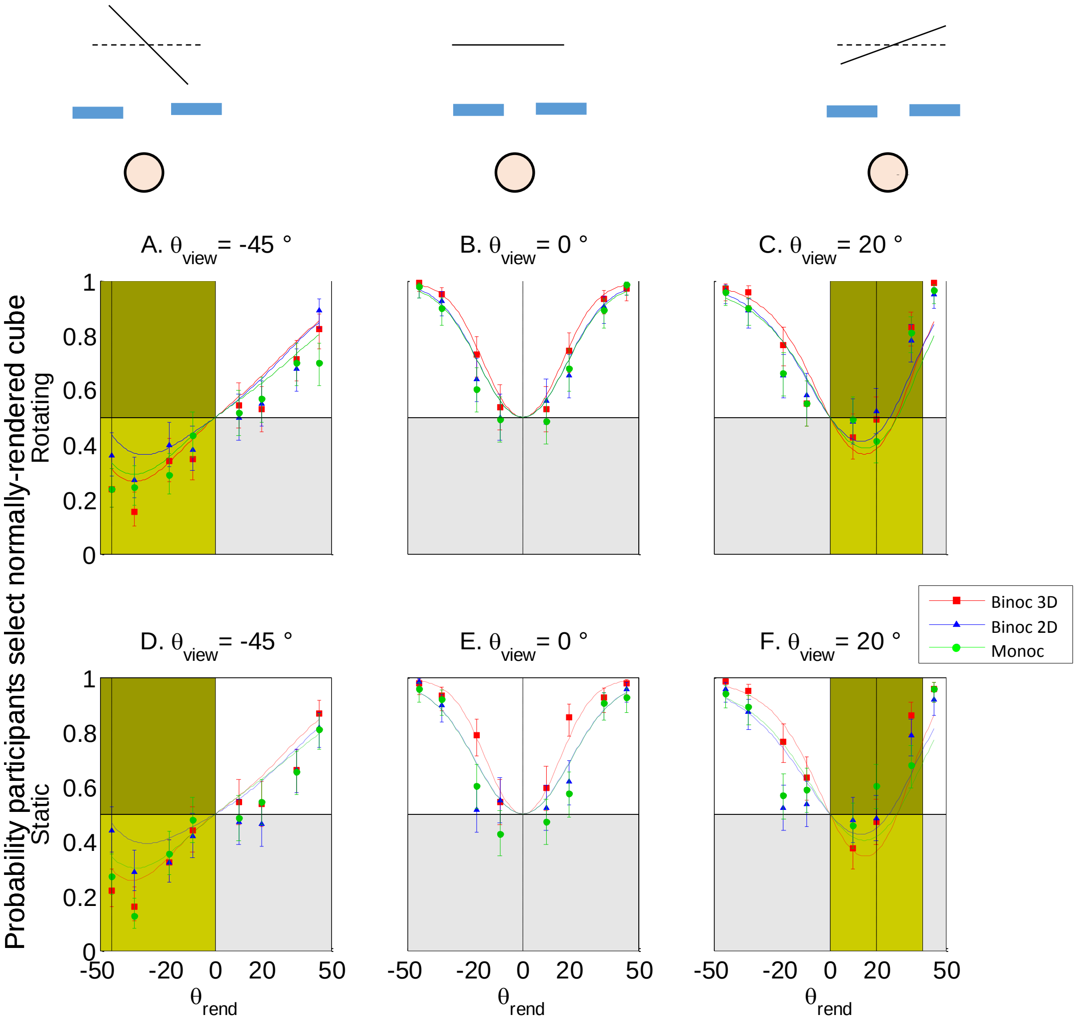


Fig. 3.8. *Results from the frame-occluded condition. As for Fig 3.7, except here a curtain prevented the subject from seeing the edges of the screen.*

3.3.4 Model fitting

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. Table 3.3 gives fitted model parameters and percentage variance explained for the different conditions. The parameters were fitted simultaneously to all data in a given object motion/binocularity/frame-occlusion condition, i.e. across all 3 panels in each row, the same parameters are used for all curves of a given colour. Table 3.3 gives the values of these parameters along with the percentage of variance explained. In every case, the model explains >80% of the variance. Interestingly the fits are generally better for the binocular S3D and B2D conditions, where they explain >93% and >85% of the variance respectively, than for the monocular conditions, even though the fit parameters are fitted independently for each binocularity condition. In the monocular conditions, subjects tend to choose the obliquely-rendered cube slightly more often than my model can capture, especially when the obliquely-rendered cube is close to being geometrically correct. However, the generally successful performance of the model confirms the qualitative argument developed above, that objects tend to look less distorted if they are rendered *either* for the geometrically correct viewing angle *or* for normal, perpendicular viewing. An advantage of the model is that it also allows us to make quantitative comparisons between the two mechanisms, as follows.

		Fitted model parameters			**Compensation index, C=A/(A+B)**	**% variance explained**
		Weights	**Sensitivity**			
		Geometrical-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%

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

3.3.5 Quantifying the preference for normal rendering vs geometrical correctness

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

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

[Eq3.5]

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

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

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

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

Table 3.4. *How the individual factors affected the compensation index (main effects) and the interactions between the factors, denoted by *. Results are from a linear generalized estimating equation done in SPSS, and return the χ2 value, along with the degrees of freedom and associated P-value. Significant effects are highlighted green, non-significant results in red.*

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

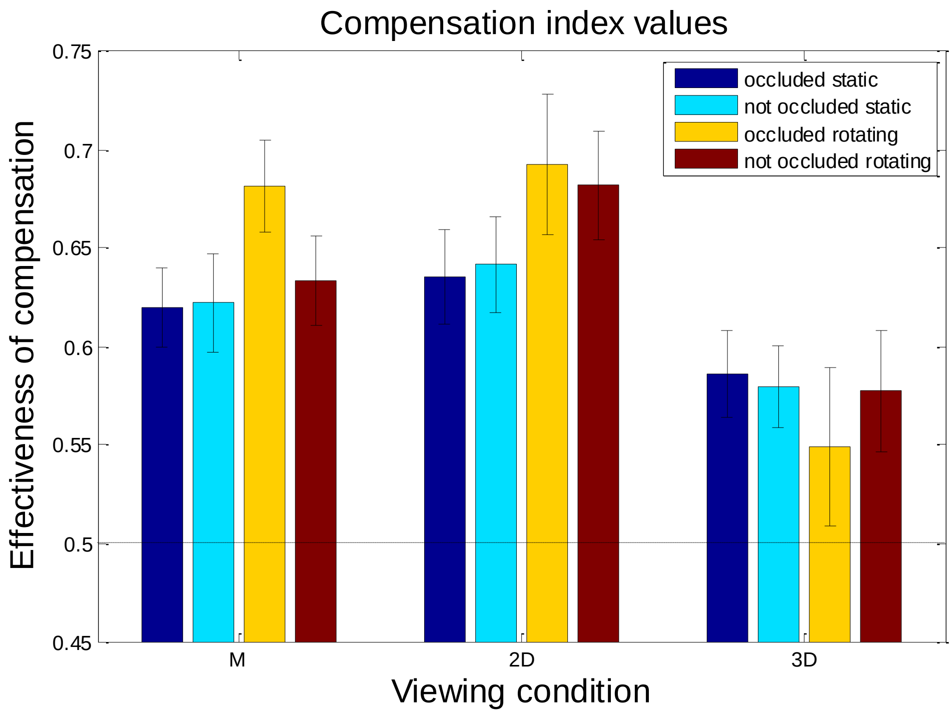


Fig. 3.9. *Compensation index, defined as the ratio A/(A+B). A and B are model parameters modelling the strength of the preference for normal rendering and for geometrical correctness, respectively. Errorbars show standard error values generated during the linear general estimating equation. Higher values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

3.3.6 Effect of S3D

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

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

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

3.3.7 Effect of frame occlusion

As my statistical tables show, the effect of frame occlusion was one of the weakest in my statistical analysis. Comparing Figs 3.7 (frame-visible) and 3.8 (frame-occluded), little difference is apparent. In Fig 3.9, the compensation index is barely affected by frame occlusion, moving from C=0.63 when the frame was visible to 0.62 when it was occluded (averaged over other viewing conditions), which is not statistically significant. However, occluding the frame does produce a substantial – and significant – drop in compensation for the monocular static condition (Fig 3.9). This is in qualitative agreement with Vishwanath et al (2005). These authors found some compensation with monocular viewing when the picture frame was visible, but none for monocular viewing through an aperture. I also saw a significant effect of occlusion when restricting my analysis to data where the normal-rendering and geometrical-correctness preferences make opposite predictions ($P < 0.0005$, Table 3.3).

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

3.3.8 Effect of object motion

Our statistical analysis of the raw data indicates that object motion is a significant factor in both the full data set and the important subset where the preference for geometrical correctness is pitted against the preference for normal rendering ($P = 0.011$ and $P = 0.023$ respectively). However, object motion did not have the effect I had expected. I had speculated that structure-from-motion cues might contribute to the compensation mechanism, increasing the preference for normal-rendered objects. In fact, object motion decreased the compensation index for both monocular and binocular 2D cubes (Fig 3.9); this was significant in the monocular condition. In stereoscopic 3D, object motion did tend to increase compensation index, but the increase was not significant. As table 3.2 shows, when considering the full set of raw data, there is a significant interaction between object motion and binocularity (P<0.0005). Pairwise comparison shows that object motion has a significant effect even when considering the individual binocularity conditions (P=0.011 for all three conditions of S3D, B2D and monocular) However, this interaction was not significant when I restricted my analysis to the subset of data in table 3.3. I conclude that overall, object motion has little consistent effect on perception.

3.4 Discussion

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

There are good reasons to imagine that this compensation mechanism might be weaker for stereoscopic 3D content, mainly because disparity is now not a reliable cue to the location and orientation of the screen plane. Informally, one can experience this by moving from left to right in front of an S3D image. The image appears to move in synchrony with you, as when a portrait's eyes appear to follow the viewer around the room (Koenderink, van Doom, Kappers, & Todd, 2004; Perkins, 1973) but now extended to the whole depicted object. S3D content is often already affected by a number of distortions, such as the puppet theater effect or cardboard cut-out effect (Banks et al., 2012; Yamanoue, Okui, & Okano, 2006) . If oblique viewing produces further distortions in perceived depth or shape, this would be a further problem for creators of S3D content. It would be particularly difficult to address in applications such as 3D cinema, where content must be viewed by large numbers of people simultaneously.

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

3.4.1 Sensitivity to viewing angle

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

3.4.2 Range over which compensation operates

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

3.4.3 Regression to expected shape

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

3.4.4 Differences between stereoscopic 3D and 2D

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

Banks et al (2009) also compared shape distortions in oblique viewing for 2D and S3D stimuli and found that viewing in S3D abolished compensation almost completely. In contrast, I find that compensation still dominates even in S3D, though it is less effective than in 2D. I highlight three differences in protocol which may contribute to this difference. First, perceptual invariance depends on the stimulus, and particularly the depth variation in the stimulus (Banks et al., 2009). My stimuli were solid cubes, with a sidelength from 6 to 14cm, viewed from a distance of 120cm. The stimuli of Banks et al were hinged wireframe squares with a sidelength of 30cm, viewed from 45cm. my stimuli thus contained relatively less depth variation. I opted for this stimulus rather than one which would be more closely related to Banks et al. as in this thesis I wish to consider stimuli and setups as close to natural viewing in the home as possible. In that respect sitting participants further away made more sense. Second, the longer viewing distance used in my study may have enhanced the preference for normal rendering. Artists since the Renaissance have discussed the recommended distance at which to capture the perspective projection of an object, in order for it to look pleasing and natural. Hagen and colleagues have argued for a distance at least ten times the mean object size along its various dimensions, very close to that used in my experiments (Hagen & Elliott, 1976; Hagen, Elliott, & Jones, 1978); Leonardo da Vinci recommends a smaller distance of 3 times the height of the object (Da Vinci, 2012). The longer viewing distance reduces the amount of perspective convergence, making the projection closer to orthographic. Viewers report such projections as appearing more veridical, even when they are geometrically incorrect for the given viewing distance (Hagen & Elliott, 1976; Hagen et al., 1978). This reflects the fact that viewers do not compensate for wrong viewing distance as they do for oblique viewing (Cooper, Piazza, & Banks, 2012). Thus, when viewed and rendered normally, my cubes should have looked veridical, whereas the hinge stimuli of Banks et al may still have looked “wrong” because of the short viewing distance relative to the size of the object. This may have weakened the effectiveness of the compensation for oblique viewing. Finally, my stimuli were renderings of cubes, where there is a clear canonical form which may have influenced perception, whereas those of Banks et al were wireframe hinges, with no clear expectation regarding hinge angle. One might expect the “regression” mechanism of Thouless (1931) and Gombrich (1972) to operate more strongly on cubes than on hinges. In the terms of my model, this would be expected to boost the parameter *r*, i.e. make subjects more tolerant of departures from geometrical correctness. It might also boost the weight of *B* relative to *A*, thus reducing the compensation index *C*. If so, this could potentially be one reason I found less compensation with cubes than Banks et al did with hinges.

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

3.4.5 Effect of frame visibility and object motion

A new contribution of my study was that I investigated the effect of object motion. This is particularly relevant for entertainment applications of S3D, where content is generally dynamic. I had speculated that structure-from-motion cues, together with the rigidity heuristic, might enable the visual system to compensate more effectively for oblique viewing. In fact, object motion had little effect in S3D, and tended to weaken compensation in the 2D and monocular conditions. This suggests that the difference between S3D and 2D TV and movies may be even less than that for S3D and 2D static images.

3.4.6 Limitations

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

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

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

This study has only considered one effect of oblique viewing: distortions in perceived shape. Another approach would be to consider whether viewing stereoscopic content from inappropriate viewing angles is a source of viewer discomfort (Howarth, 2011; Lambooij, Ijsselsteijn, Fortuin, & Heynderickx, 2009). It would be interesting to look at the various different definitions of a “zone of comfort”, the range of depth allowed in 3D content before discomfort begins to adversely affect the viewing, and see if changing the viewing angle has any effect on the zone of comfort (Shibata, Kim, Hoffman, & Banks, 2011).

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

3.5 Conclusion

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

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

4.1 Introduction

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

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

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

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

This change in the appearance of the scene is known as miniaturisation (or the 'puppet-theatre' effect) (Banks et al., 2012). The effect makes the 3D object in question appear noticeably smaller compared to what the viewer would expect to see naturally (Hopf, 2000; Meesters, Ijsselsteijn, & Seuntiens, 2004; Yamanoue et al., 2006). The effect can be in the other direction (gigantism) where the object of focus appears too large. This occurs when the interaxial value of the two camera lenses is too small, compared to too large for miniaturisation. However gigantism occurs considerably less often than miniaturisation. The effect cannot be physically measured as it is subjectively assessed, therefore each person may perceive the amount of miniaturisation differently (Yamanoue et al., 2006). The level of effect is heightened when using the 'toed-in' camera configuration (Banks et al., 2012). Furthermore, the probability of the effect occurring increases with larger viewing distances (Yamanoue et al., 2006). Some of these concepts were shown in Fig 1.2. in the introduction.

Solutions have been studied to minimise the effect, with some techniques allowing greater stereoscopic distances to be presented (Hopf, 2000). The size-stereo miniaturisation conflict is often reported in the viewing of football matches in S3D. As the matches are filmed a great distance away from the players, there is little sense of depth for the viewer. As mentioned previously, to enhance this sense of depth the interaxial values are vastly increased and 'toed-in' configuration often used, which in turn enhances the depth of the images presented. This can lead to the players appearing unnaturally small compared to the pitch and stadium, for this reason losing the sensation of realism (Yamanoue et al., 2006). Another good example of this is in the S3D feature film 'Gravity', which set the interaxial distance of the two 'virtual cameras' in creating the CGI footage of space to be at a distance of 'infinity', so as to give the impression of vastness to space and 'smallness' to the characters in the movie. I intend to assess this conflict as although it is well known, it has not been studied in great detail in a research environment.

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

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

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

Familiar size as a cue to depth is inherently tied to the stereoscopic cues to depth. Wallach and O'Leary considered these two with physically created stimuli on cardboard and deemed that the interaction between the two cues was of interest. In their study, they did not separate the familiar size and vergence cues, but recorded mathematically that both must be involved in the perception of depth (O'leary & Wallach, 1980).

In this chapter I intend to look at the cue conflict that arises between familiar size and stereo information in S3D content and whether one cue is preferable to the other in visual perception. Without the combination of cues, images can appear unrealistic for the viewer (Scarfe & Hibbard, 2011). However, when two cues conflict with one another in S3D content, one of them will more often than not violate the perception of what one expects the original natural scene to be viewed as and the other one will take preference (Cavanagh, 1987). Studies have demonstrated that different weightings can be put on cues depending on the environment or situation the subject was in (Rushton & Wann, 1999). There also appears to be some prior bias depending on how the cues have been used during developmental years, when the system is still considered plastic (Rushton & Riddell, 1999). It is known that the human visual system is better at detecting relative disparity over absolute disparity (Parker, 2007). This could somewhat explain some of the reactions to familiar size and binocular disparity cues in the scenes seen, as most of the S3D scenes will be displaying relative disparity, along with various other cues to depth such as occlusion, shading, and texture. It could be that the introduction of disparity in S3D media means that different considerations need to be given to the setting of the scenes in, production and filming of S3D content. I conducted an experiment to consider whether the chosen structure of my experiment would allow us to compare the familiar size and vergence cues. Once this was confirmed I continued with my full experiment. Additionally, the amount by which each cue is weighted is thought to be judged on the reliability of the cue itself in the given situation (Rosas, Wagemans, Ernst, & Wichmann, 2005). Therefore, in the main experiment, I also intend to numerically assess the conflict between size information and disparity to see if it is possible to assign weighting bias to the size and stereo information cues via signal detection modelling. I ran three experiments to do this. Work in this chapter has been presented at the Vision Science Society conferences 2015 (Hands, Khushu & Read, 2015), and the other experiment study has also been published (Hands et al. 2014)

4.2. Initial experiment

4.2.1 Material and methods - Participants

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

4.2.2 Material and methods - Equipment

Subjects were shown a computer generated image of a standard credit card (ISO/IEC 7810 identity card), as this object is well known and of a particular size (8.56cm x 5.398cm). The images were presented on an LG passive 3D TV (LG 47LD920-ZA) using the computer programming environment Matlab ("www.mathworks.com,") and the Psychtoolbox extension (Brainard, 1997; Kleiner et al., 2007). Participants sat at different viewing distances of 50cm, 100cm and 200cm in different blocks of the experiment, measured perpendicularly from the centre of the screen to the midpoint of the eyes, with their eyes at the same height as the centre of the screen. They wore passive 3D glasses throughout the experiment, so as to not be able to tell when there was no disparity between the left and right eye images. The three different viewing distances were used to attempt to assess how different viewing distances might affect the responses. I believe that viewing over these three distances will allow us to discuss the vergence-accommodation conflict with regards to this cue combination, as at the 200cm distance, accommodation cues are somewhat eliminated (Banks et al., 2012; Hoffman et al., 2008). Therefore, only vergence should influence the perceived depth of the stimuli at that distance. The subject was sat on an adjustable chair to allow for height change. This meant each subject had their eyes in the correct position both horizontally and vertically. Participants were given a reference credit card that they could hold and look down at, but were asked to not raise the card up to the screen to aid in choosing whether the card was bigger or smaller, as this would have enabled participants to make relative comparison judgements purely on the basis of retinal size, independent of perception, which I wanted to avoid. Finally, a keyboard was used for the subjects to make their choices using the up and down arrow keys.

4.2.3 Material and methods - Procedure

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

For the black background, 7 sizes and 7 parallaxes gave a total of 49 different parameter options as each size was shown with each disparity. The noisy background only had 28 parameter options as only negative and zero parallax disparities from the above 7 were used (4 values). This was to ensure that the disparity of the noisy background could be kept constant on the screen plane, hence having zero disparity itself. Therefore, a total of 77 parameter options were shown to the subject. At times the card would appear the exact same size as the reference card and in the screen plane; in these instances the subject was forced to make a decision as no 'same' option was available, which I would expect to be simply a chance selection. Each parameter permutation was repeated 20 times. The 1540 (77 x 20) trials were displayed over 5 blocks to allow the subject to have a few minutes rest between each block. Each subject saw a random order of the different trial parameters.

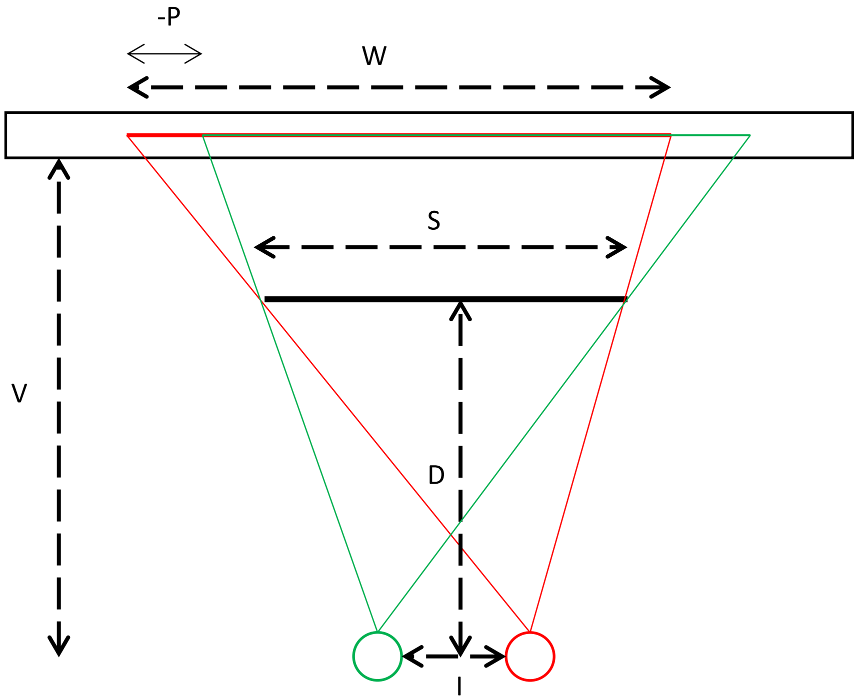


Fig. 4.1. *How the on-screen size of the card would change depending on its depth. Geometrically, in order to depict a virtual object of physical size S, I must draw the image with width W on screen. I will refer to W as “on-screen size” and to S as “physical size”. Here I = Interocular distance, D =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

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

Fig. 4.2.*The stimulus used was a credit card, shown above, at different widths and heights. Here it is displayed at a correct width and height (width 8.56cm, height 5.398cm). Image taken from http://www.psdgraphics.com/psd/credit-card-template/.*

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

4.2.4 Material and methods - Parameters

7 widths W were chosen to be displayed as the width values to be measured. The 'true' correct size of a credit card is 8.56cm. The stimulus used was therefore chosen to be 8.56cm ± 1, 2 and 3cm respectively. Eq. 4.1. below shows how the interaction of W and I can be used to generate P such that the size of S is the same as a natural card. This means that

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

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

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

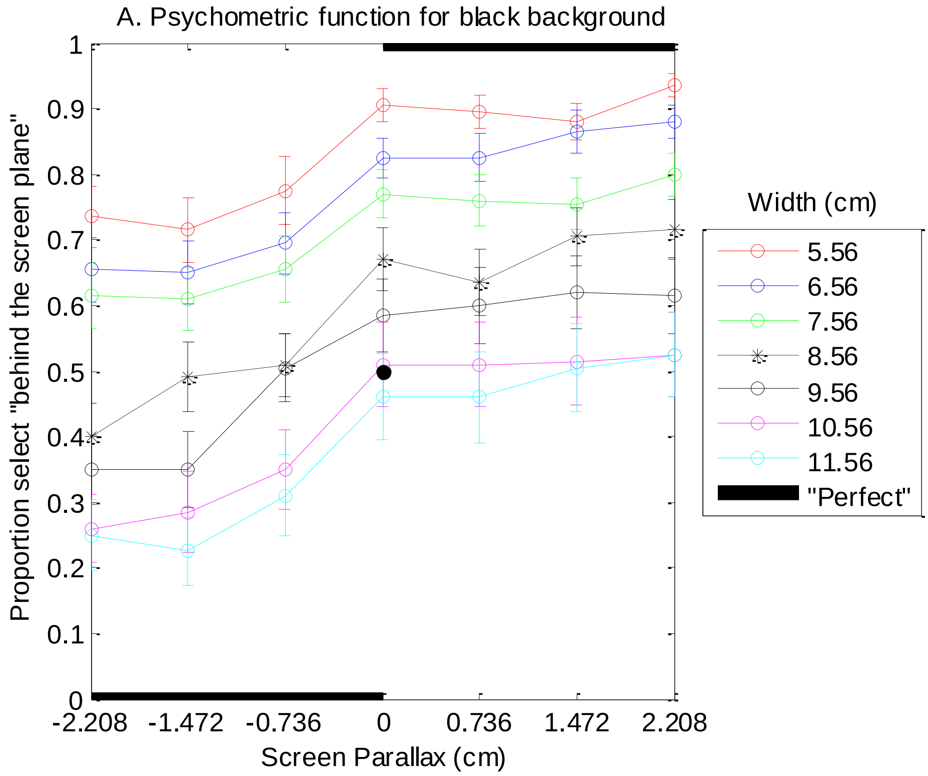
Width parameter, W, cm	Parallax parameter, P, cm
5.56	2.208
6.56	1.472
7.56	0.736
8.56	0
9.56	- 0.736
10.56	- 1.472
11.56	- 2.208

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

4.2.5. Results

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

I evaluated statistical significance using a generalised estimating equation in SPSS, using inter-subject and global comparisons of the raw data with stimulus width, disparity, and type of disparity (relative or absolute, determined by the background) as variables. As expected, participants do use screen parallax to perform the task. Additionally, my results confirm the increased ability to distinguish depth with relative disparity compared to absolute disparity. In absolute disparity, the ability to distinguish the card being in front or behind the screen is significantly worse. On average across all non-zero parallax values, participants are correct on 60% of trials with a black background, but 89% with the noisy background, where "correct" is defined based on the sign of parallax. For example, consider the case where the card has the most negative parallax, -2.21cm, but has the smallest width, 5.56cm. Then, disparity suggests the card is 'in front of' the screen plane, but the familiar size cue indicates to the participant the card is 'behind' the screen, as it is smaller. With the noisy background Fig 4.4.B shows that participants were correct 84% of the time, while with the black background they were sucessful only 28% of the time. My statistical analyses confirmed this significant difference (paired T-test on n=9 different participant observations, P = 0.001). Participants also clearly base their depth judgment partly on the on-screen size of the credit card. That is, they have a tendency to report that smaller stimuli are further than the screen while larger stimuli are in front. This tendency is particularly strong for the black background condition, Fig. 4.4.A. In this condition, the stimulus was presented in isolation, so essentially only absolute disparity cues were available. The human visual system is particularly sensitive to relative-disparity cues between nearby objects, and is not very sensitive to absolute disparity. Accordingly, in this condition participants gave significantly more weighting to the depth cue implied by the size of the familiar object, Cards that were displayed smaller than the familiar size are deemed to be ‘behind’ the screen plane significantly more than cards displayed larger than the familiar size, regardless of the disparity that they were displayed at. However it is clear that disparity still makes some contribution to the decision about depth, or else the gradient of the different width lines would be flat, and each line would be a perfectly horizontal line, going down in proportion from red to cyan as the size increased from 5.56cm to 11.56cm.



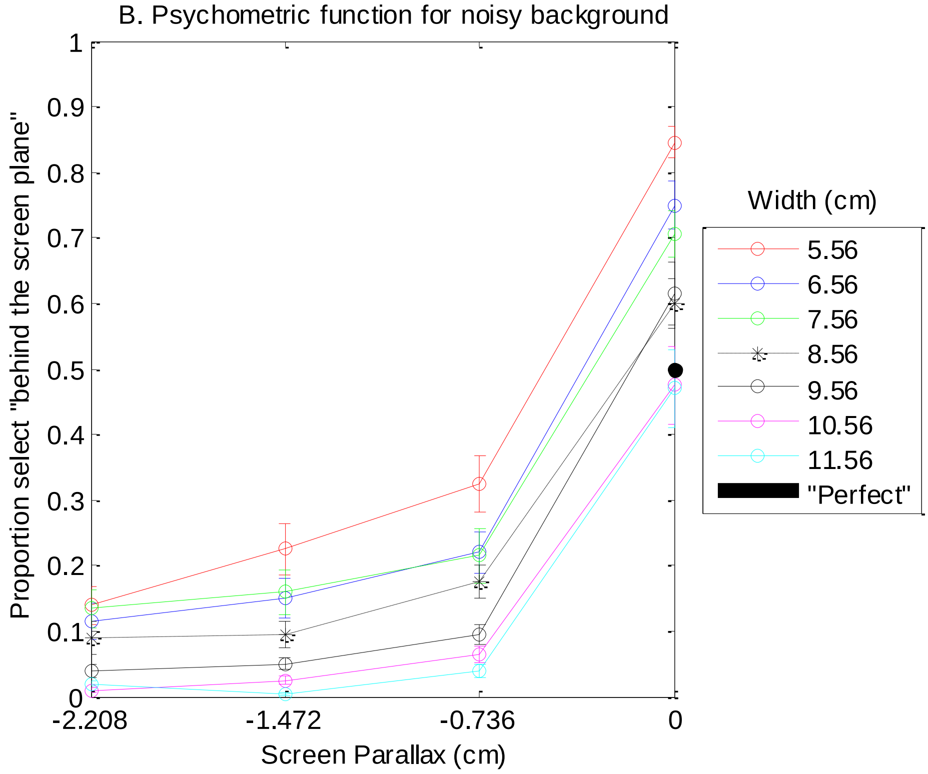


Fig. 4.4. *Psychometric functions for black (A, absolute screen parallax) and noisy (B, relative screen parallax) backgrounds. The horizontal axis shows screen parallax, and each line type*

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

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

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

A limitation of the noisy-background condition is that no positive parallaxes were presented. This was because there was no unproblematic way of doing this. A stimulus with positive parallax must either be occluded by the zero-parallax background surface, or be seen through a transparent surface, or be seen through a hole cut in the surface. Fortunately, there does not appear to be any significant difference between the zero disparity cases for the black and the noisy backgrounds. In both cases, participants show a bias towards reporting that objects are further than the screen plane. This is a helpful indication that the noisy background only provided a reference surface for comparison and did not otherwise affect the results. As can be seen, once disparity information is removed (i.e. once the display reverts back to 2D, rather than S3D) the reliance on the familiar size cue increases, as displayed by the increased gradient of the slope in Fig.4.4.B. between the negative parallax value of -0.736cm and the 2D case of 0cm parallax.

4.2.6. Discussion

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

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

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

4.3 General methods

4.3.1 Participants

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

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

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

4.3.2 Equipment

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

4.3.3 Procedure

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

In addition to the main experimental setups, each cue was analysed separately with further experiments. For size, the subject was presented images that changed in width and height but had a constant disparity of zero (i.e. flat on the screen). The subject still wore the S3D passive glasses to maintain consistency and once again was asked to decide whether the image was bigger or smaller than the reference card. I call this experiment the size judgements in 2D experiment. For vergence, the size of the card remained the same as the reference (8.56cm width, 5.398cm height) whilst the parallax changed. The subject was required to decide whether the card was appearing in front of the screen (negative parallax) or behind the screen (positive parallax), using the keyboard. I refer to this as the vergence-based depth judgement experiment. The experiments were conducted in the order of: main experiment, size judgements in 2D experiment, vergence based depth judgement experiment. There is potential for practice effects as all participants completed the experiment in this manner, which I consider more in section 4.8 (discussion). This data, combined with my data for the 'chance' instances in the experiment, will help us in the mathematical modelling in assessing any bias that occurred on an individual basis.

4.3.4 Parameters

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

4.3.5 Mathematical modelling

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

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

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

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

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

Combining these two I get the relationship that:

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

To quantify my data I developed a model based on signal detection theory (Ernst & Banks, 2002; Green & Swets, 1966) which assumes the perceived depth of familiar object relies on two competing mechanisms: That of the perceived size of the object (i.e. a size signal), and the binocular vergence to correctly assess the depth of the object (i.e. a parallax signal). I use modelling based on log values, as it allows for easier mathematical calculations in combinations, and is justified by Fechner's law. Hence:

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

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

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

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

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

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

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

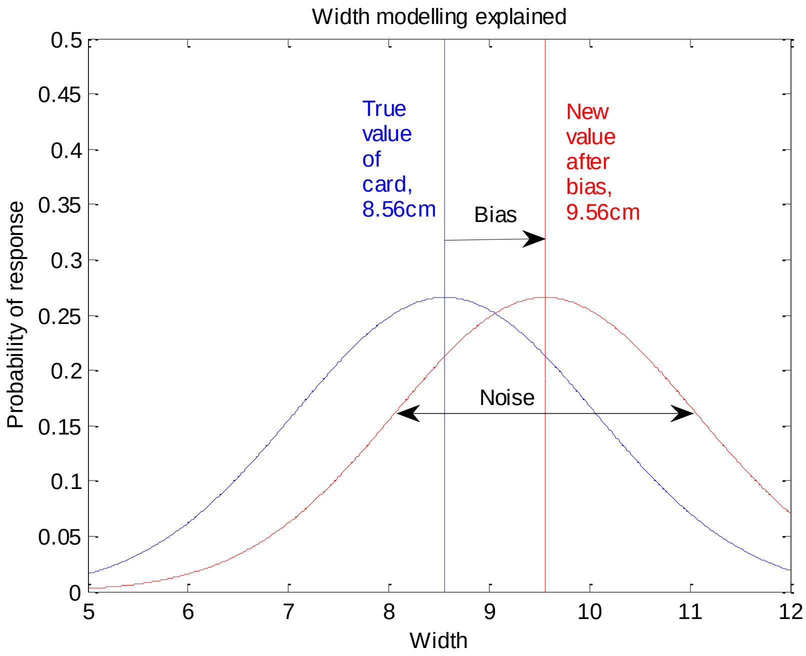


Fig. 4.5. *How the bias and noise contribute to the signal detection for the familiar size cue. In this case the physically displayed width W is equal to the 'true' value of 8.56cm. The size bias is a value of +1.0cm in this example and the noise is a value of 1.5cm. This means in this case my size signal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

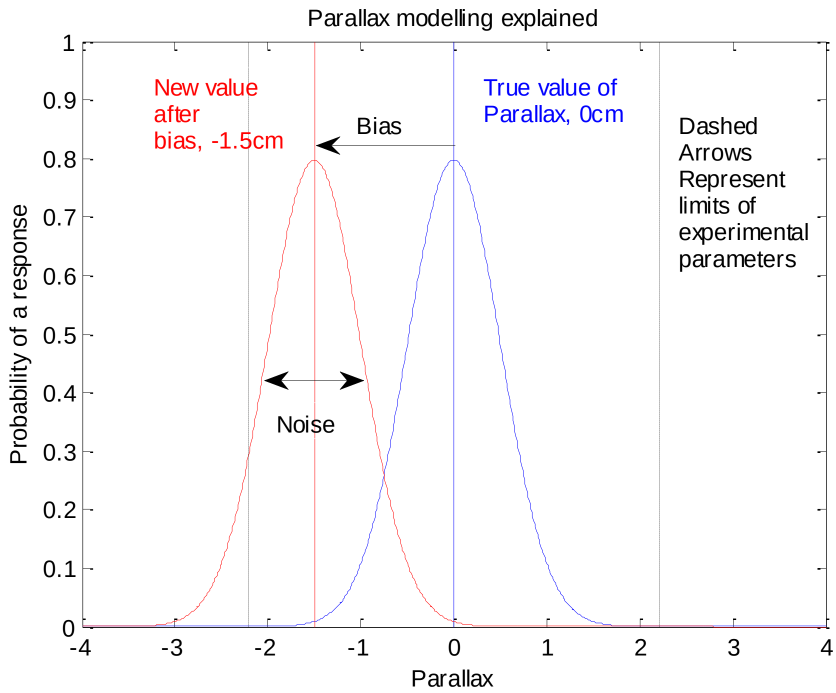


Fig. 4.6. *The parallax signal is slightly more complicated than the size signal model but the premise is very similar. In this example the parallax shown is 0cm, but the bias moves the signal in front of the screen plane, and in this example the noise is considerably less so the peak is higher. Here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

P ~ N(-1.5,0.5)

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

4.4 Experiment 1: Vergence-based depth judgments

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

4.4.1 Methods

In experiment 1 the methods are exactly as in the general methods, as the two backgrounds were used, and the parallaxes associated with them. In this experiment however the displayed width was set to be the true displayed value of 8.56cm, and did not change. Thus the experiment was a shorter one than experiment 3, with a total of 220 trials in the experiment (1 width x (7 absolute parallaxes + 4 relative parallaxes) x 20 repeats). In this experiment the question asked to participants was slightly different also; participants were asked to answer if the displayed card was ‘in front of or behind the screen plane’, as they had done for the initial experiment. This was because I was in effect attempting to establish a value for the participants’ stereoacuity.

4.4.2 Model

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

Eq. 4.9.

$$P(infront) = \frac{1}{\sigma_p\sqrt{2\pi}}\int_{-\infty}^{0} dp\, exp\left(-\frac{\left(p - lnI + \ln(I-P) - \theta_p\right)^2}{2\sigma_p^2}\right)$$

$$= \frac{1}{2}\left(1 + erf\left(\frac{-lnI + \ln(I-P) - \theta_p}{\sigma_p\sqrt{2}}\right)\right)$$

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

4.5 Experiment 2: Size judgments in 2D

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

4.5.1 Methods

In experiment 2, the methods are exactly as in the general methods, as the two backgrounds were used, and the full range of the widths of the displayed cards. In this experiment however the displayed parallax was set to always be a value of 0cm, hence in the screen plane, and did not change. Thus the experiment was a shorter one than experiment 3, with a total of 280 trials in the experiment (1 parallax x 7 widths each for both black and noisy backgrounds x 20 repeats). In this experiment the question was as it was in the general methods.

4.5.2 Model

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

Eq.4.10.

$$P(too\ big) = \frac{1}{2}\left(1 + erf\left(\frac{lnW + \theta_w - lnS_{cc}}{\sigma_w\sqrt{2}}\right)\right)$$

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

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

4.6 Experiment 3: Size/vergence interaction

The original aim of this study was to consider the interaction between the two pieces of depth information when they were conflicting with one another. Experiment 3 is the experiment that really considers the depth cue combination problem. Experiments 1 and 2 could be considered supplementary to this one, and the models for experiments 1 and 2 are smaller cases of the experiment 3 model (where one of the weights was considered to be zero). I believe that the model fits for experiments 1 and 2 will help us to justify my experiment 3 model. All 6 parameters are considered in the fitting here (θ_p, θ_w, σ_p, σ_w, g_p, and g_w).

4.6.1 Methods

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

4.6.2 Model

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

Eq. 4.11.

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

$$P(too\ big) = \frac{1}{2}\left(1 + erf\left(\frac{g_w(lnW + \theta_w) + g_p(lnI - ln(I - P)\ +\ \theta_p) - lnS_{cc}}{\sqrt{2(g_w^2\sigma_w^2 + g_p^2\sigma_p^2)}}\right)\right)$$

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

4.7 Results

I used an omnibus ANOVA in matlab to assess significance of factors and interactions between them for my experimental results at all viewing distances, pooled across participants. These results are shown in table 4.3 below, where I report the viewing condition (black or noisy) and then consider different factors (experiment number, virtual size of credit card, parallax the card was displayed at and the viewing distance of the participant) as well as the respective 2-way interactions. I report the F-statistics and also the respective P-values, as well as the DFN and DFD (degrees of freedom in the numerator and denominator 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	0
	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

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

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

I consider the results and fitted models in Figs. 4.7 to 4.11. These are fits from different participants in either the black or noisy background condition. These are typical of the population of fits and patterns. In doing this I try to establish whether my model with 6 free parameters is effective, and what the underlying trends might be for the experimental results. As can be seen from Figs. 4.7. to 4.11. my model fits the data well, and captures the pattern of the 63 different datapoints (absolute disparity) or 42 datapoints (relative disparity) with only 6 free parameters. Figures show individual participants data and is not averaged or pooled across participants here, as the results were all very similar. Participants are more likely to say the virtual card displayed is ‘too big’ more as the vergence depth information moves the card further away from the participant (i.e. from negative to positive parallax) for any given size of the virtual card. Participants also consider the card ‘too big’ more often as the virtual size increases, regardless of disparity. This trend is exhibited in all three experimental setups (where applicable).

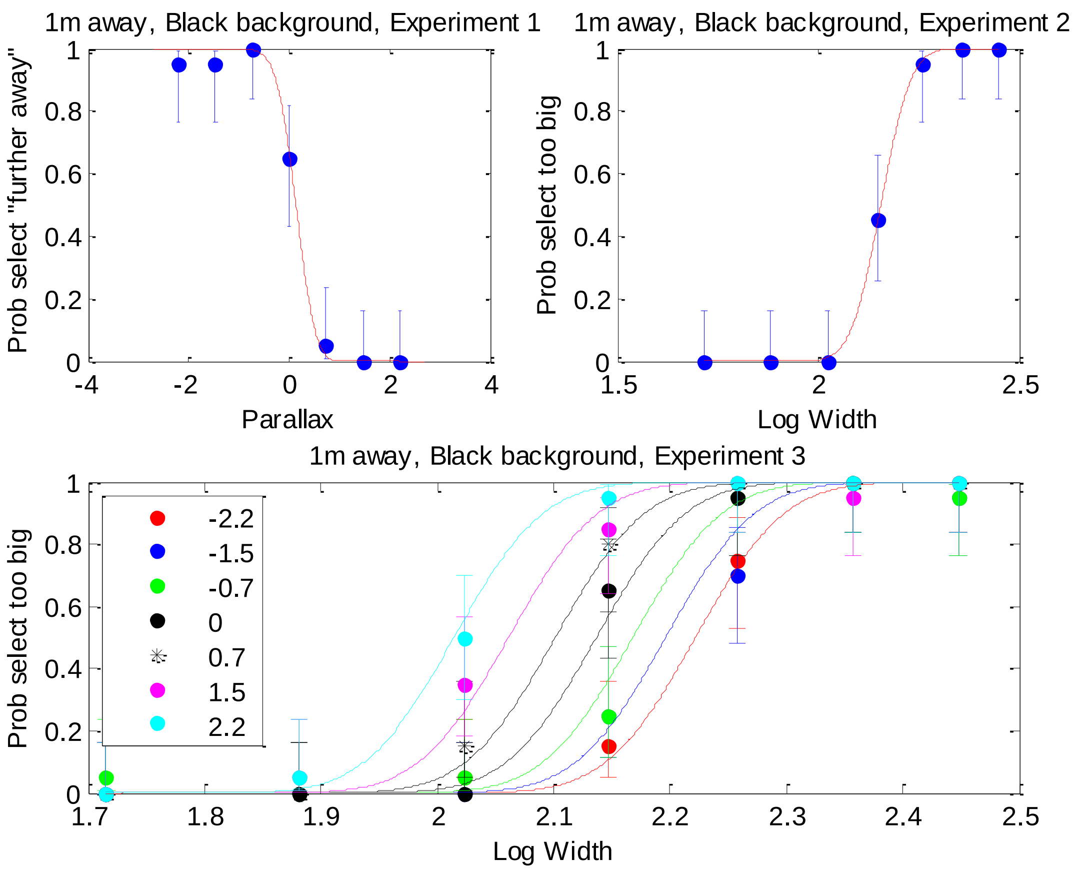


Fig 4.7. *Data and model fits for all three experiments for one participant at a viewing distance of 1m with absolute disparity (black background). Each different colour represents a different parallax value. The x axis represents the parallax values (experiment 1) or the log of the width of the virtual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

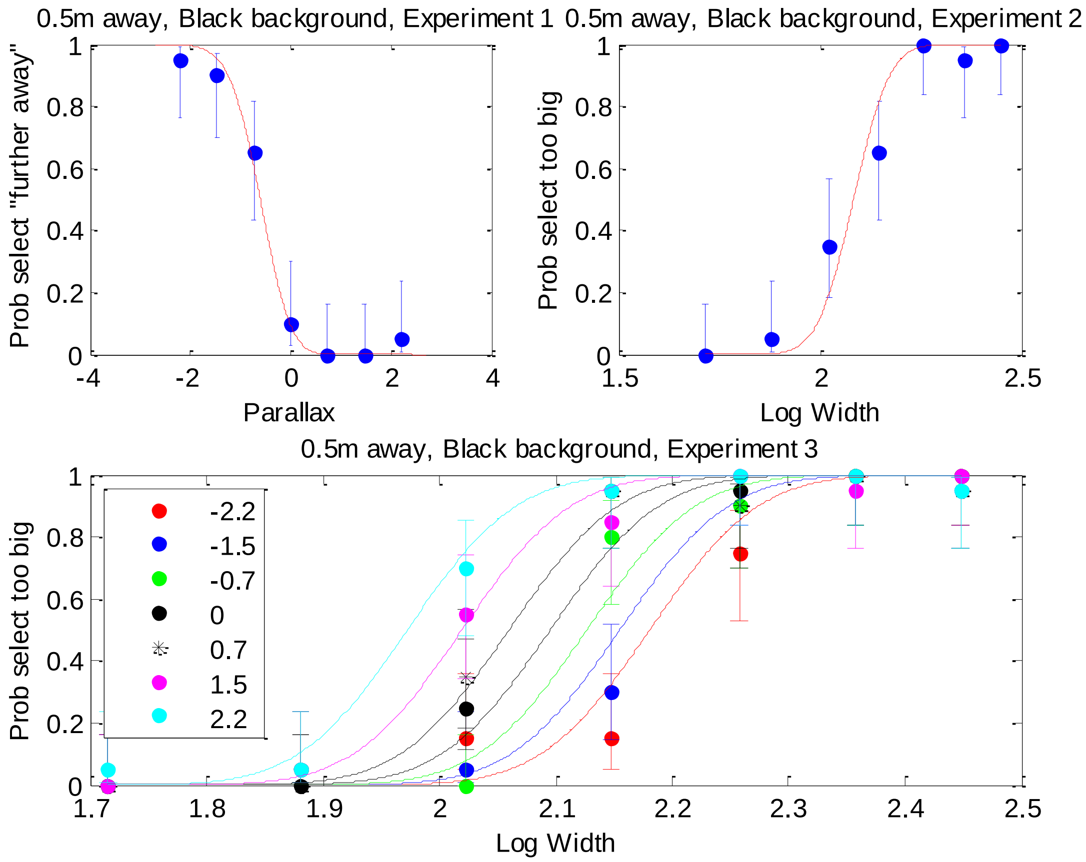


Fig 4.8. *Data and model fits for all three experiments for one participant at a viewing distance of 0.5m with absolute disparity (black background). All other information as in Fig. 4.7.*

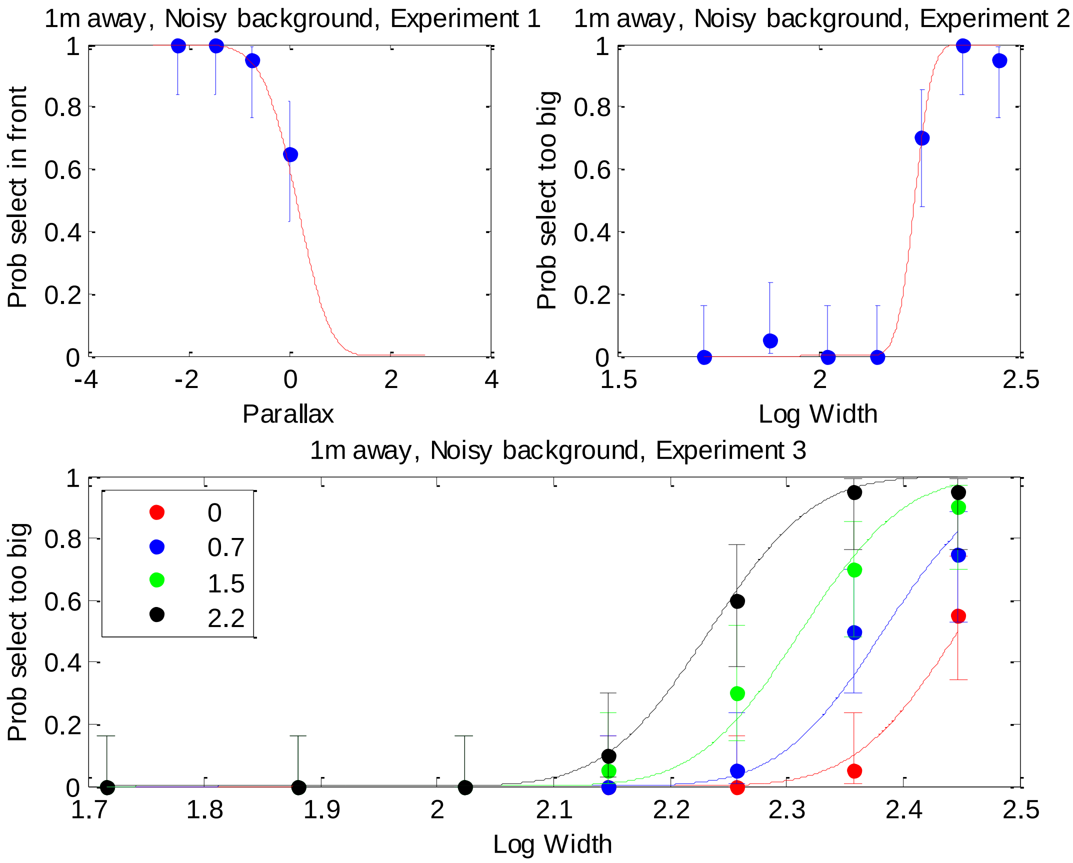


Fig 4.9. *Data and model fits for all three experiments for one participant at a viewing distance of 1m for relative disparity (noisy background). All other information as in Fig. 4.7.*

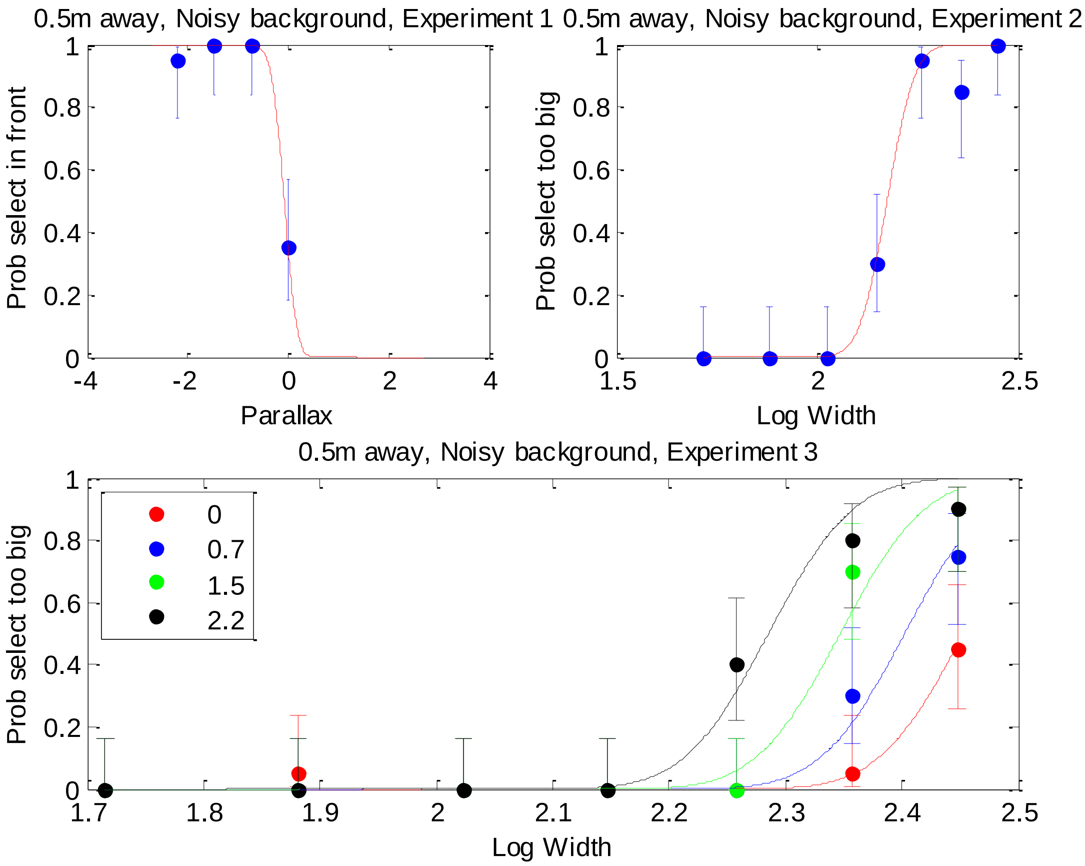


Fig 4.10. *Data and model fits for all three experiments for one participant at a viewing distance of 0.5m for relative disparity (noisy background). All other information as in Fig. 4.7.*

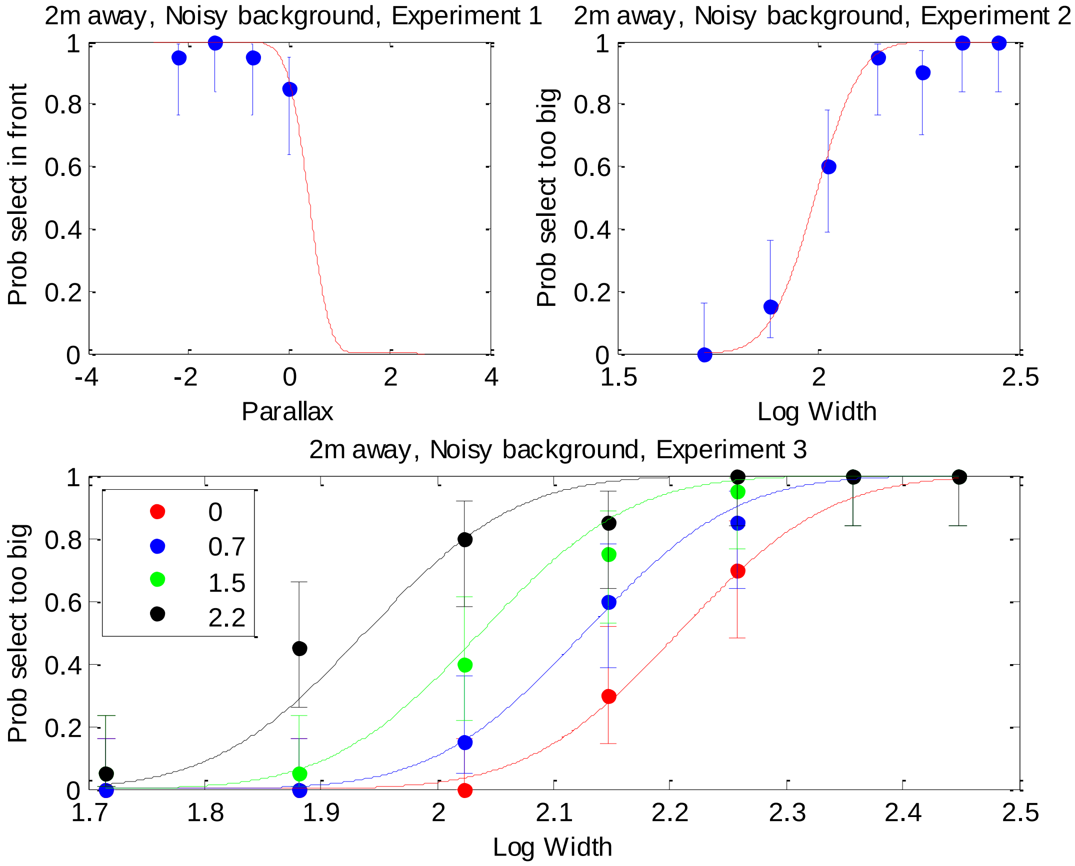


Fig 4.11. *Data and model fits for all three experiments for one participant at a viewing distance of 2m for relative disparity (noisy background). All other information as in Fig. 4.7.*

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

Eq. 4.12.

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

This suggests that the weights and the noises are inversely proportional. This would be the optimal way to integrate the two different depth cues, as seen in previous studies (Ernst & Banks, 2002). To consider if this was an applicable constraint I first considered the relationship between the noise ratio and the weight ratio for black and noisy backgrounds respectively for the size/vergence interaction experiment, in log axes (Fig. 4.12. below). The ratio considers the weight (or noise) of the size signal in comparison to the weight (or noise) of the parallax signal.

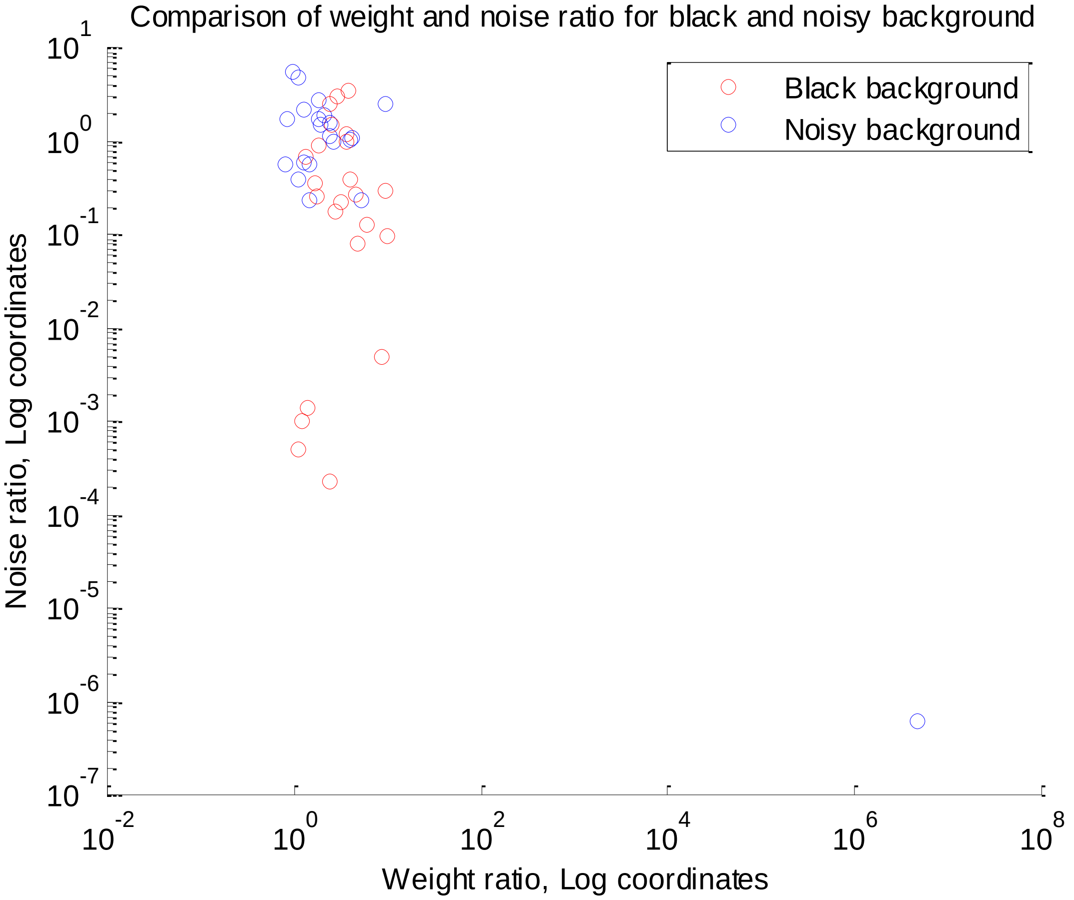


Fig. 4.12. *Comparison of the weight and noise ratios for black background (red circles) and noisy background (blue circles) respectively. As can be seen there are some extreme outliers, but the relationship appears to be negatively correlated, as would fit with previous studies.*

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

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

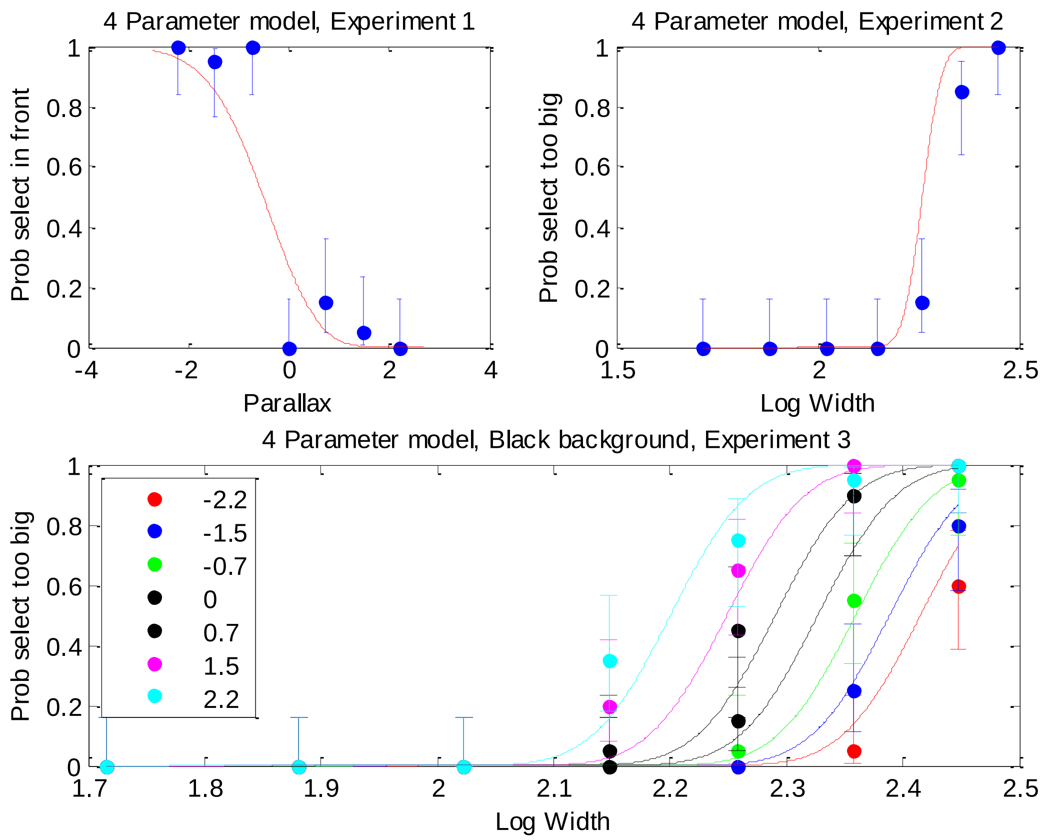


Fig. 4.13. *Data and 4 parameter model fits for all three experiments for one participant at a viewing distance of 2m for absolute disparity (black background). In this model the weights assigned to the two signals are calculated using Eq. 4.12 above. All other information as in Fig.4.7.*

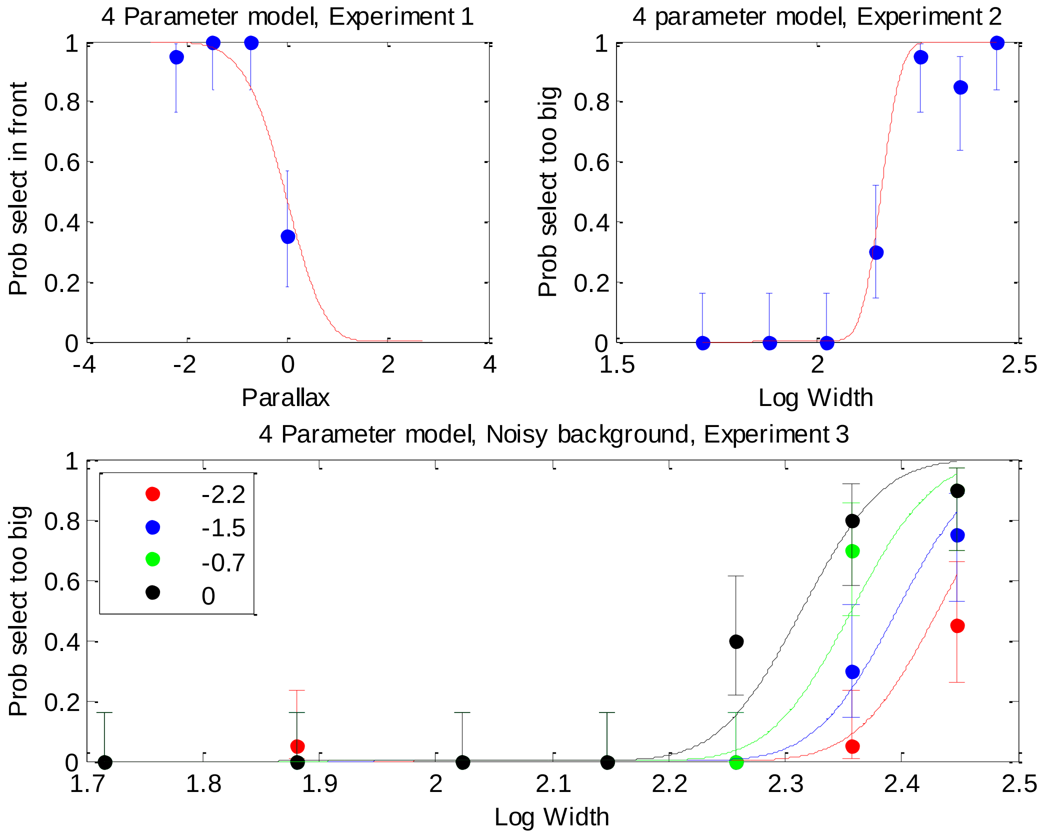


Fig. 4.14 *Data and 4 parameter model fits for all three experiments for one participant at a viewing distance of 0.5m for relative disparity (noisy background). In this model the weights assigned to the two signals are calculated using Eq. 4.12 above. All other information as in Fig. 4.13.*

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

Participants reported finding relative depth judgements (i.e. the noisy background case) easy to consider the depth. This is not new information but backs up what I already know about absolute and relative disparity information in that humans are better at relative depth judgements than absolute. This is also backed up by my model results, where the weights were more equally considered. This suggests the noise from the vergence signal was more reliable.

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

4.8 Discussion

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

This study found that the familiar size cue to depth was more heavily weighted (and hence had less noise associated with its signal). In the case of 2D content, it could be that exposure to images with things the wrong size (i.e. movie screens, television shows, pictures in books and on walls) forces the human visual system to become very adept at discerning depth based on the relative size of the familiar object. Indeed from a young age humans are exposed to many pictures in books, in frames and on devices with screens such as televisions, tablets and more recently smart phones which take up a disproportionate amount of space on the retina for the known familiar size of that object. Humans conversely may develop this ability to understand that the retinal space taken up is directly linked to the depth of the object, and not the size of it, due to many exposures of this type. However it is much rarer for humans (particularly in their developmental years) to experience a similar level of exposure to S3D technology. It could be this imbalance that causes humans to depend more heavily (and make more reliable judgements) on the familiar size cue to depth than the stereoscopic cue that comes from vergence information.

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

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

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

The conflicts between two depth cues can become large enough to become robust, which can lead to one cue to depth and size dominating over the other one, although not to a point of complete ignorance (Hillis et al., 2004). This can be seen in the case of football players and the Subbuteo effect, where players can be seen as being very small on an enormous pitch in S3D displays. In this case, the discrepancy between the size cue to depth and the disparity cue to depth are in such conflict that the image warps. When the cues don’t conflict quite as robustly the effect is not observed.

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

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

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

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

From my initial experiment, which asked a question based on depth in the image, I find that observers reported depth based on disparity where reliable information was available. These experimental results reinforce the conclusion that the human visual system is much more adept at judging depth from relative rather than absolute disparity. However, my results also indicate that other cues may be considered when the disparity information is unreliable, even when these cues – such as expected size – do not strictly relate to depth.

The implications from this suggest that miniaturization, as an industry-known problem, comes from the fact that the disparity information is given in a relative setting, so the viewer can easily distinguish where in depth the disparity of the image suggests objects are, and only then considers whether they appear the correct size or not. In the case of absolute disparity the viewer appears to first use familiar size to determine where the object is, and then either confirm with disparity information, or else discount the disparity information and use the familiar size depth cue alone. This appears to be a similar principle for viewing 2D footage, in that the perceived size is the familiar size, and the perceived depth is considered second. The issue of miniaturisation is still a small one, as people still report enjoying watching S3D content with larger interaxial values and toed in configurations (such as those used in football match filming).

4.9 Conclusion

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

This study considered whether a familiar size cue to depth and depth information from vergence movements were combined in an optimal fashion, and if so which cue to depth was considered the more reliable. Using signal detection theory and mathematical modelling I found that despite individual variance there was a clear trend for participants to rely more heavily on the familiar size cue as an indicator of the displayed depth, and hence this had a more significant impact on judgements when depth information was conflicting between the familiar size and the vergence based depth information. A model was constructed to quantify this relationship and it reflects the findings that cues are combined in an optimal fashion, as has been found in the past with other conflicting cue information (Brumm et al., 2010). In this study more reliability is associated with relative depth judgements than absolute.

5. An eyetracker study on shifting the screen plane

5.1 Introduction

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

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

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

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

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

There is a wealth of evidence that weakening the cues to the existence of the screen plane results in a stronger impression of depth. This goes back at least to Tscherning, cited in Ames (Ames Jr, 1925). Binocular disparity is a powerful cue to the flatness of the screen plane, so weakening disparity cues is an immediate way of reducing the salience of the screen plane. Tscherning discusses the depth illusion produced when 2D pictures are viewed through Javal's iconoscope, an optical device which presents the same image to both eyes. Simply viewing a picture from a greater distance produces a similar effect, but the iconoscope also disrupts the relationship between convergence and viewing distance, a manipulation which itself increases the depth illusion (cited in Ames Jr, 1925). The zograscope (Koenderink, Wijntjes, & van Doorn, 2013) worked in a similar way. Claparède (Claparède, 1904) discussed the "paradox of monocular stereopsis": the stronger depth illusion created when 2D pictures are viewed monocularly, again because this removes the binocular cues to flatness. Ames reports blurring the image in one eye also strengthens the depth illusion, especially if a cylindrical lens is used to blur vertical lines while leaving horizontal ones sharp. Again, this presumably disrupts disparity cues to flatness, leaving pictorial cues free to dominate. Binocular cues are not the only ones indicating the screen plane. Accommodation is a monocular cue to flatness, at least at near viewing distances, so removing this cue (by viewing through a small hole) or disrupting it (by viewing through positive or negative lenses) also strengthens the illusion of depth. Ames reports that viewing a flat image through a mirror produces the same effect. This is presumably by introducing uncertainty as to the position of the picture in space: the frame removes the continuity between the observer and the picture via the surrounding objects and surfaces, while the mirror's surface presents a competing candidate for picture plane. Perhaps most interestingly for the present paper, Ames also discusses "Changing the convergence of the eyes from that normally required by the distance from which the picture is viewed", by placing prisms in front of the eyes (Duane, 1900; Verhoeff, 1935). This is directly equivalent to the “shifted 2D” exploited by current S3D producers.

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

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

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

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

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

The attention that is paid to approaching and moving objects is something creators of television have tried to utilise to draw the viewer's attention to certain objects and key moments in scenes. It is regularly used in advertising and also in production of movies and television, as well as in games consoles (Schild & Masuch, 2011). In 2D the objects that producers wish to have the attention of the viewer at a certain time can be altered using movement, colour, focus, and a change in size and shape; however, due to technological limitations, binocular viewing cannot be used to generate a sense of looming. With the growing popularity of S3D technology in the home, and it becoming more accessible because of falling prices (Darlin, 2005), there is now a potential to utilise the S3D technology to attempt to draw viewers' attention to a non-focus part of the scene, using binocular vision and disparity to increase attention to an object, while also utilising other cues such as size, shape, colour and focus (Welchman et al., 2005). Typically people pay attention to things closer to them (Franconeri & Simons, 2003), and so S3D could be utilised to draw the attention of the viewer. Approaching objects could now be depicted with both looming and stereoscopic depth cues to draw the attention of the viewer to the desired object with even greater strength. This would suggest that there is potential for S3D content to attract attention in different locations to 2D content.

Eyetracking technology has been utilised in many different studies in measuring vergence and then from the vergence measurements calculating where in depth the participant is looking. Studies such as that conducted by Duchowski et al. attempted to use an eyetracker with a customised Wheatstone stereoscope to measure and model vergence movements from participants while viewing images with virtual depth (Duchowski, Pelfrey, House, & Wang, 2011; Wheatstone, 1838). They reported that not only did the vergence movements correspond to the depth displayed on the scene, but when considering the jitter in the function of depth perception (caused primarily by saccades) they could smoothen the function of depth perception with a quadratic filtering method. To aid in surgical procedures, Mylonas et al. considered both binocular eyetracking in robots and using the link between horizontal disparities and viewing distance (Mylonas, Darzi, & Yang, 2004). Developing an integrated stereo viewer and eyetracker they successfully considered vergence changes in their study. Hillaire et al. used eyetracking to consider where the participant was attending to (in the paper they refer to this position as the focus point) and suggest algorithms to 'decimate' some of the periphery details to allow for faster rendering (Hillaire, Lécuyer, Cozot, & Casiez, 2008). This is an interesting use of eyetracking technology in speeding up processing time, which could become even more important if a study utilises S3D technology, and hence needs to process two separate images for each retina.

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

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

5.2 Material and methods

5.2.1 Equipment

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

5.2.2 Stimuli

The stimuli were 13 separate 30 second clips from the BSkyB production 'Micro Monsters with Sir David Attenborough', which was filmed in S3D. Clips were chosen from 2 episodes that were made available by BSkyB for the study, and were chosen so that the 30 second timespan started and ended at a sensible place, avoiding starting or stopping the clip midsentence, and also to be sure the clips were engaging. Both the left and right eye of each clip was made available in an AVI file for the study. The software program 'Stereo Movie Maker' (available at http://stereo.jpn.org/eng/stvmkr/index.html) was used to modify the clips. The subsequent modified clips were each displayed in four different ways:

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

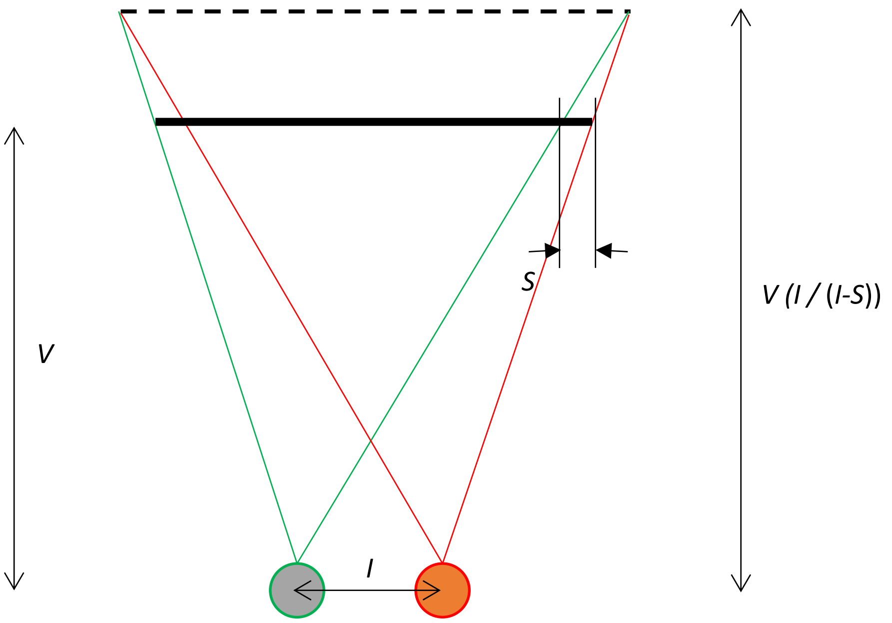


Fig. 5.1. *Geometry of my experiment. The interocular distance I is for calculations assumed to be 6.3cm. By shifting the resized images 30 pixels in each eye, the virtual image is moved behind the screen by a factor of 1.28. So when the viewing distance V is 100cm, the plane of the image should appear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

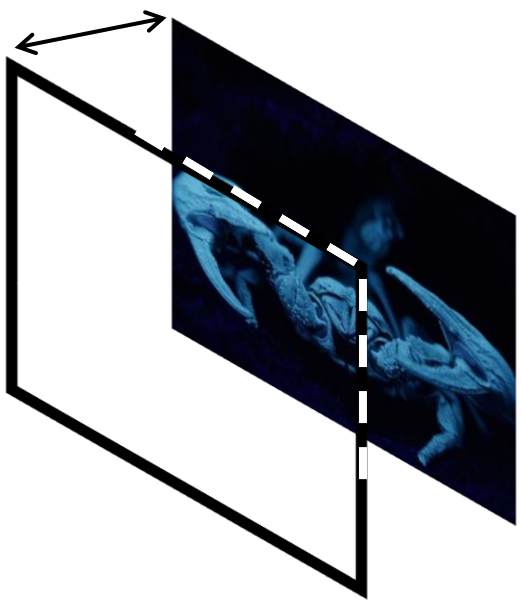


Fig. 5.2. *Diagram to explain the concept of shifted 2D clips. The frame of the television is shown, and the clip is moved in depth behind it, using disparity. My aim is to determine if this*

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

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

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

5.2.3 Procedure

Participants were asked to sit in the chair comfortably, wearing the glasses and resting their chins on the chinrest. They were then given a brief explanation of the eyetracker technology and a very basic explanation of why it was being used in the experiment (i.e. to record where participants are looking during the clip presentation). The participant, before beginning the experimental trials, went through a calibration and validation process using the eyetracker, to be sure the recording was correctly measuring where the participant was looking on the screen. The validation process gave an average deviation from target and a maximum deviation value in degrees. If the average deviation was more than 0.5° away or the maximum was more than 1° away from the validation target on the screen a recalibration took place. This was the limit for 'good' validations according to the eyetracking software. Each individual trial then showed an initial timing clip, to ensure that recording was accurately started when the clip began, and then the 30 second clip of the content was shown. After each clip the participants were asked to assess the perceived depth in the image, stressing that the actual content (i.e. how interesting it was) wasn't important to the study, and to give a score on a 7 point Likert scale, with 1 being "terrible" or "not noticeable 3D" to 7 being "fantastic, immersive 3D". Once the participant reported their score on the impression of the depth in the clip the trial was finished and the next trial would begin. Between each trial another validation test was completed to ensure the participants head hadn't moved away from a position acceptable for the eyetracker (0.5° average, 1° maximum) and, if the validation failed, a recalibration took place immediately before the next experimental trial. The consequence of the procedure being structured as such resulted in different experiments taking different lengths of time. Some participants only needed very few calibrations, while others needed a recalibration on roughly a 3 trial basis. Participants were allowed to move away from the headrest whenever they wanted to, but, if they did, a recalibration took place before the experiment was allowed to continue. Both left and right eye measurements were made, with the intention of being able to calculate the position in depth that the eye was verging to, by considering the parallax between the two eyes. Participants were given a £10 shopping voucher for their participation.

5.2.4 Participants

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

5.3 Results

5.3.1 Subjective ratings

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

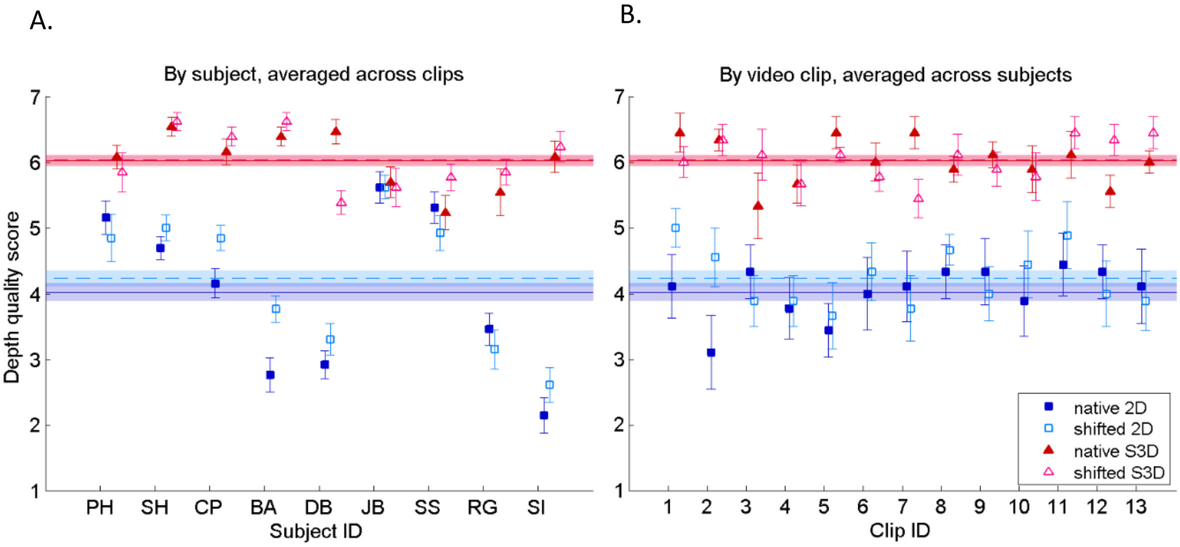
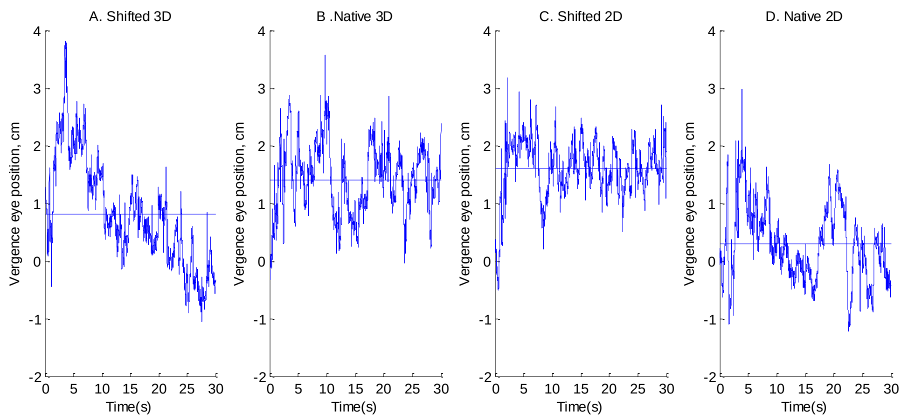


Fig. 5.3. *Depth quality scores for each of the four different viewing conditions, (A) for the 9 different subjects, averaged across the 13 video clips, and (B) for the 13 different clips, averaged across all 9 subjects. Blue squares show results for 2D, red triangles for S3D; filled symbols / solid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

5.3.2 Vergence eye measurements

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

The first analysis conducted on the eyetracking data was to consider the measured parallax between the right and left eye gaze positions in cm, as recorded by the eyetracker when both eyes were being detected successfully at the same time. This was to consider if the visual system reacts differently to shifted 2D and native S3D content, in terms of vergence. I estimate the vergence at each time point by considering the difference in location horizontally between the left and right eyes gaze position. This is given as a measurement in pixels. Hence by using a simple pixel to cm conversion for the monitor in question, the parallax between the eyes position (and hence an estimation of the vergence) for each time point was calculated. An example for a participant for one trial is shown below to illustrate in Fig. 5.4. Sub figures A, B, C and D show the different format the trials were shown in and the recorded parallax between the gaze positions of right and left eyes. The blue solid line indicates the average depth measurement for that trial. Sub figure E shows the average value for both the horizontal and vertical parallax values in cm for each configuration, in blue and red respectively, with 95% confidence intervals displayed. Even with rigorous calibration, eyetrackers can be subject to a considerable amount of error in calculating where the eyes are fixating while recording data, and this is reflected in my vertical parallax measurements which my results suggest were different from zero. However the horizontal parallax measurements are larger than the vertical parallax. These are subject to a high level of noise with large standard deviation values, which can be attributed to the noise of recording.



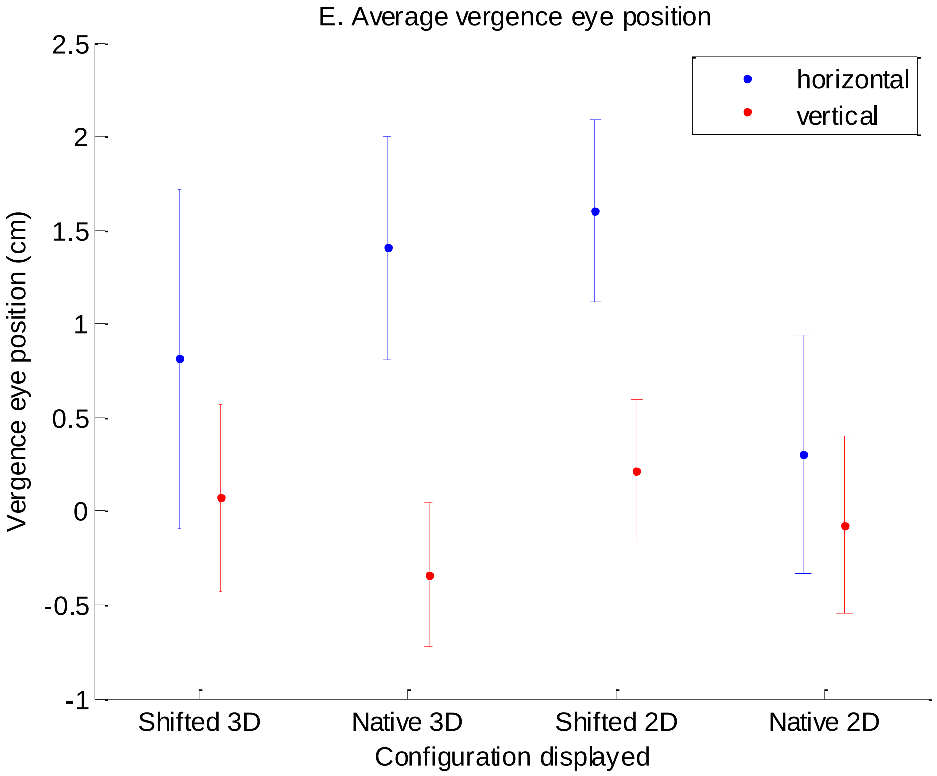


Fig. 5.4. *Participant CPs vergence eye measurements for clip number 6 using left and right eye gaze positions for different configurations (A – D). X axis is the time of the 30 second clip, y axis is the measured depth calculated from the difference between the gaze positions of the left and right eyes.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

This was very typical of all measurements made on all participants, shown in Fig 5.5: mean values that suggest a change in vergence depending on the configuration. Across participants, the average depth measured using vergence movements for the shifted S3D content was 125.6cm, for the shifted 2D content 127.3cm, for the native S3D 102.0cm and for the native 2D 101.5cm, calculated from the parallax in a similar procedure as used above to calculate the depth of the screen plane. Thus the shifted and native depth measurements were, in each case, near to the depth plane that the content was displayed on (100cm, physical screen plane for native, and 128cm, virtual plane through a window for shifted; dashed lines in Fig 5.5). Of course, the S3D content varies in depth, but the depth in S3D clips is typically centred on the screen plane, with some content in front of it and some behind it. Thus it is not surprising that in Fig 5.5. the native and shifted S3D content is shown to have a similar mean parallax value as the 2D content. For both native and shifted content, people on average verge around the depth plane of the content. This was despite the fact that the content plane was purely virtual in the shifted conditions, whereas in the native conditions, the mean depth of the content plane was on the physical screen, so the binocular depth cues were also supported by the accommodation cues and other cues to the physical screen plane.

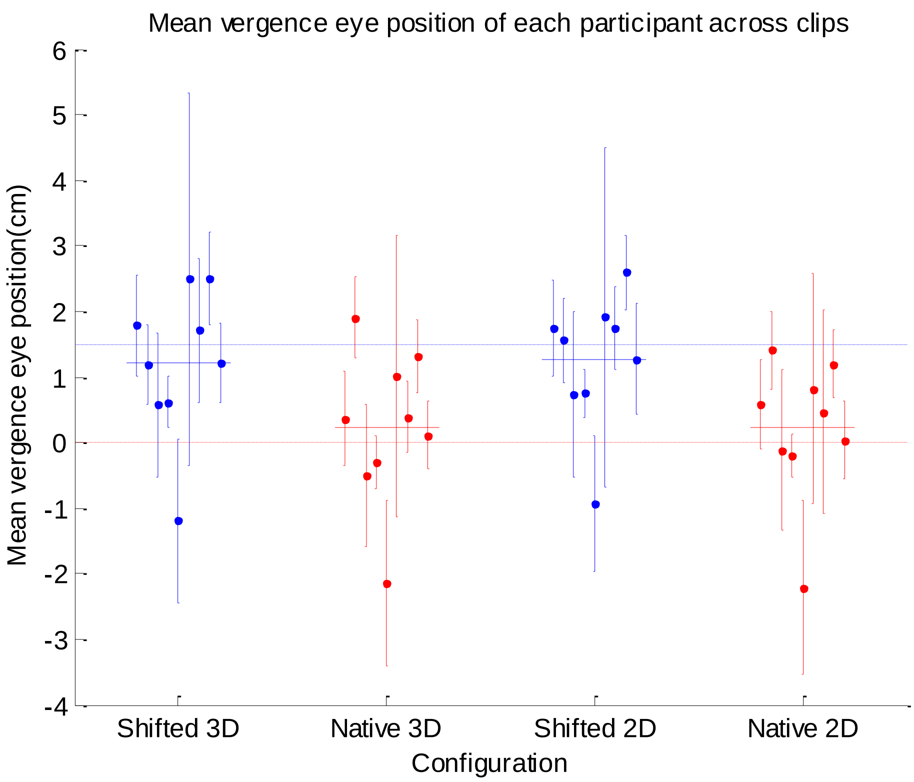


Fig. 5.5. *Mean vergence eye position of each participant across clips. The vergence eye position was averaged across each clip initially and then across all clips for each participant, shown by configuration. Errorbars show ±1 standard deviation. As can be seen, the value of the native vergence eye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

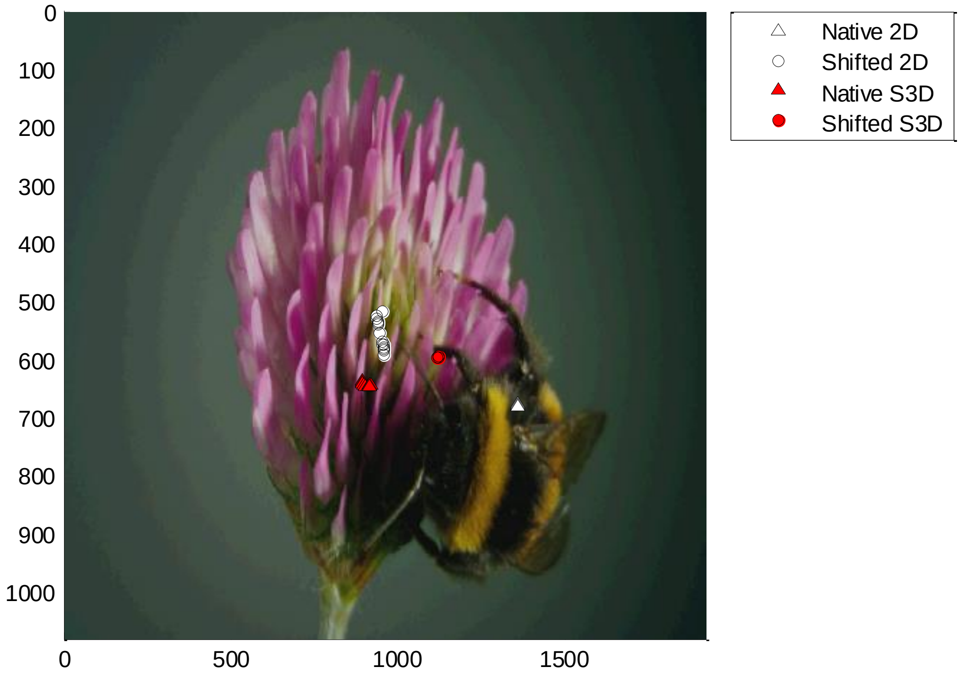
configurations. Mean across participants is shown by solid bars. A blue dotted line indicates zero movement, and a red dotted line indicates 1.5cm (the amount I introduced when shifting the clips).

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

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

5.3.3 Gaze position during the clips

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


0
100
200
300
400
500
600
700
800
900
1000
0
500
1000
1500
Native 2D
Shifted 2D
Native S3D
Shifted S3D


0
100
200
300
400
500
600
700
800
900
1000
0
500
1000
1500
Native 2D
Shifted 2D
Native S3D
Shifted S3D

2746

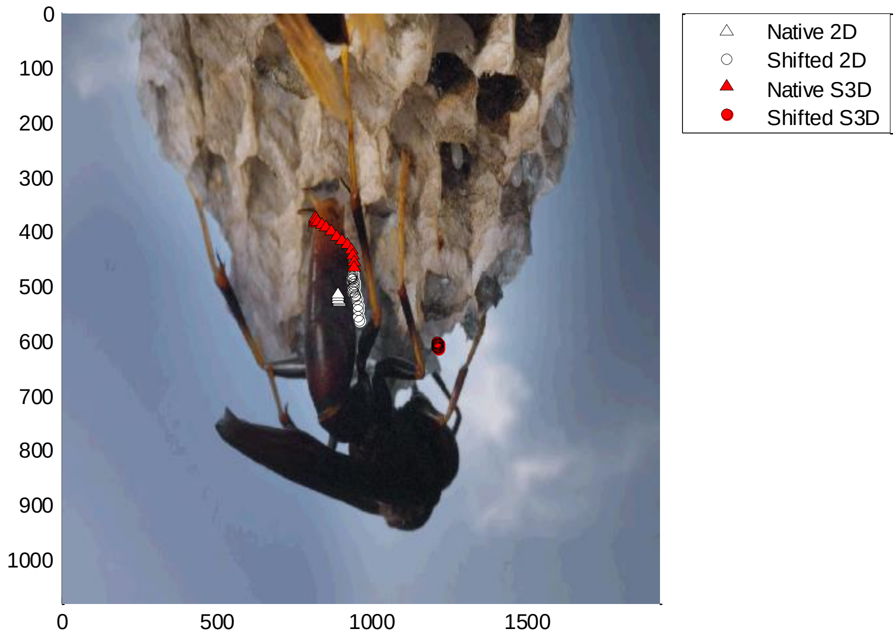


***Fig. 5.6.** 3 Eye gaze positions from example frames of three different clips. As can be seen, the attention in that frame appears to be drawn mostly to the focus object in the scene. The discrepancies between the different gaze positions for different configurations are non-significant when the entire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

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

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

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

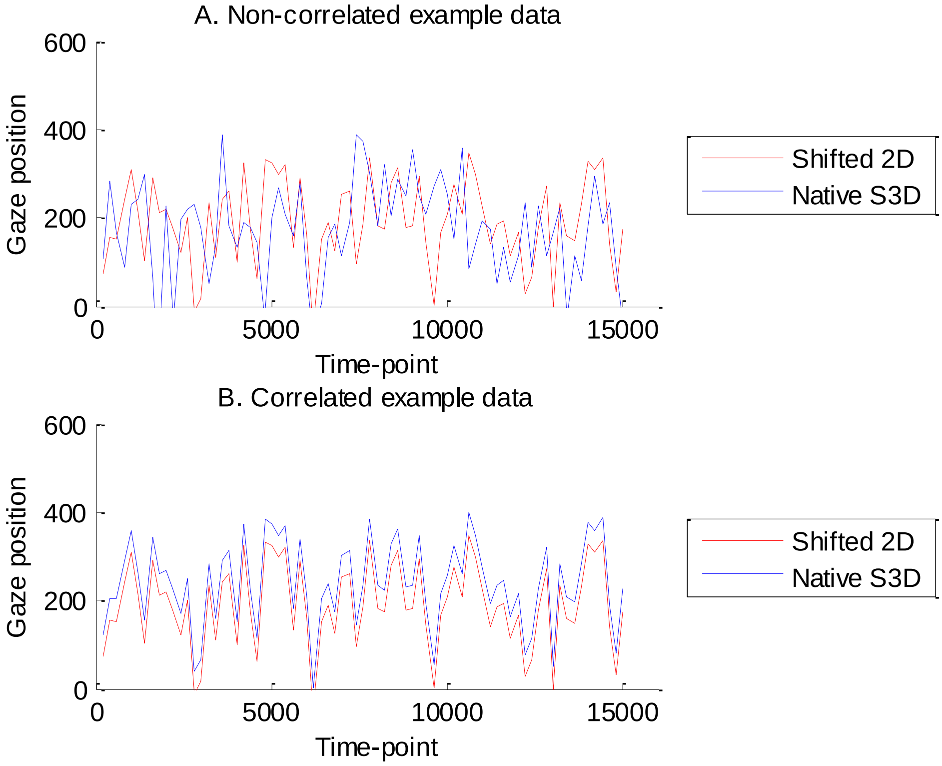
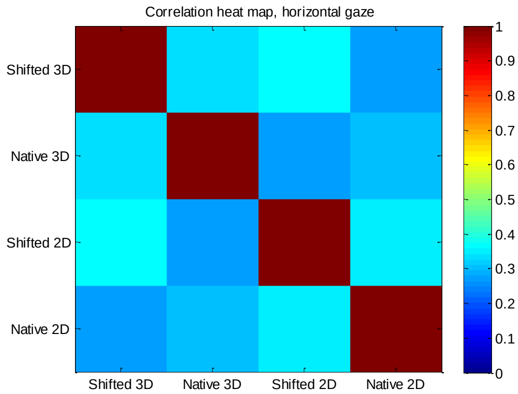


Fig. 5.7. *Example data to show that correlations are an important consideration. A) Two random datasets generated from the same normal distribution, with a very low correlation (0.0023). B) The random dataset for shifted 2D simulated data has been kept, but the native S3D data has now been directly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

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

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



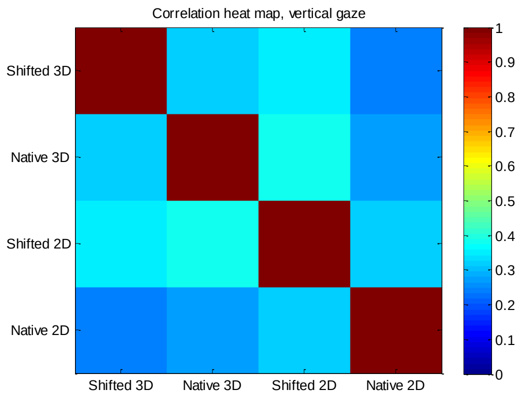


Fig. 5.8. *Example correlation heat maps for participant one in the horizontal and vertical gaze positions. Red squares indicate a high correlation, blue squares a low correlation.*

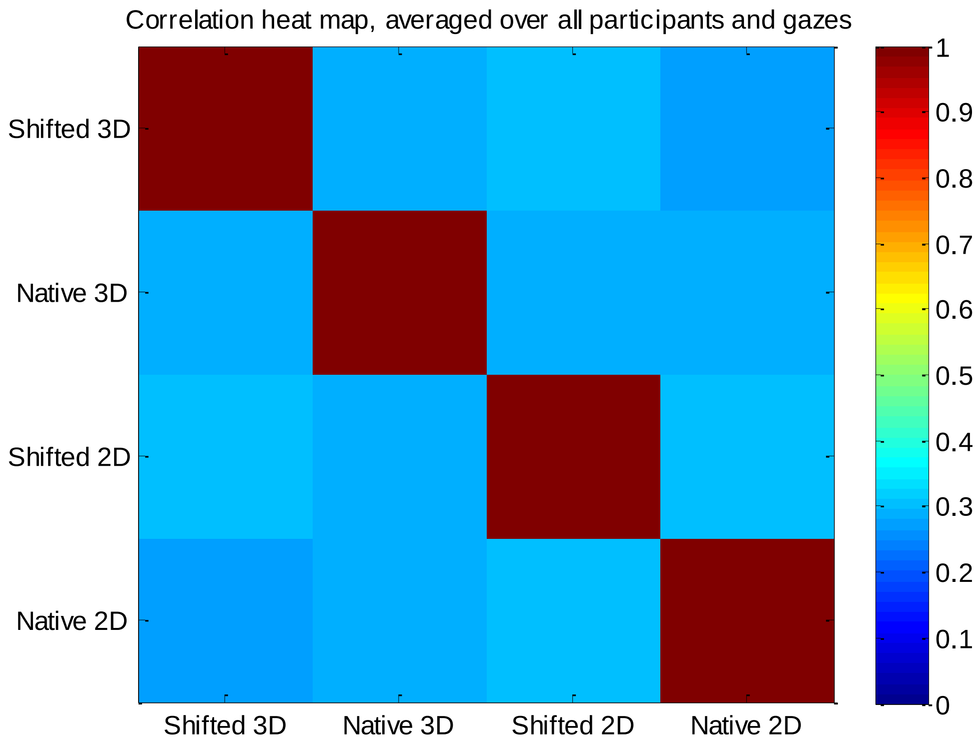


Fig. 5.9. *Heat map for correlation values averaged across all participants and gaze positions.*

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

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

	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

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

5.3.4 Fixation analysis

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

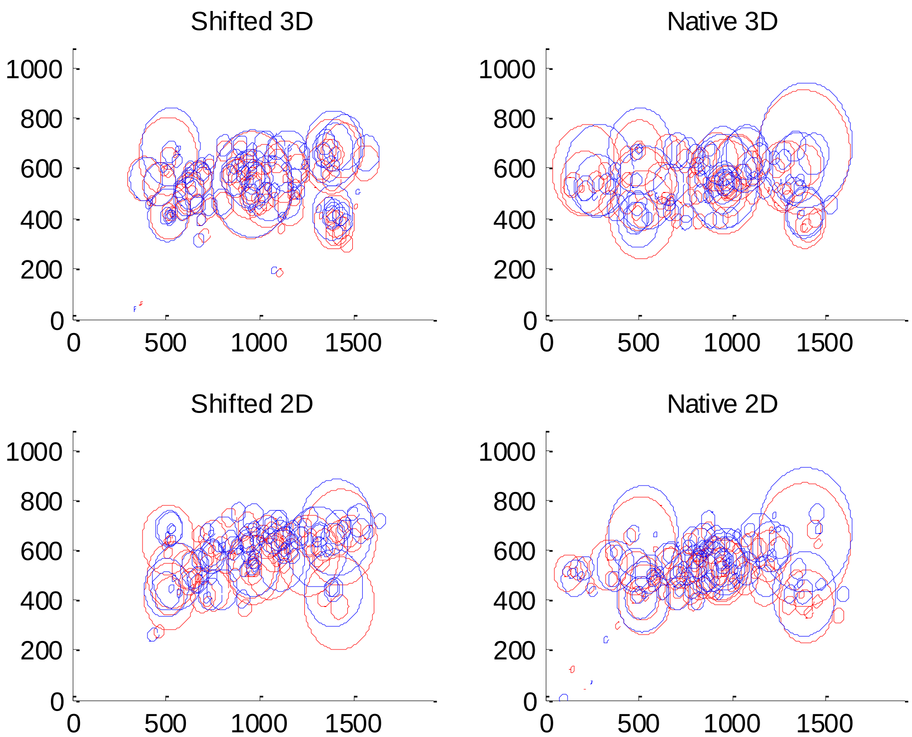


Fig 5.10. *Eyetracking fixations for an example participant in one clip. Red circles are centred on where the left eye fixated and have a radius relative to the length of that respective fixation. The same applies for the right eye with the blue circles.*

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

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

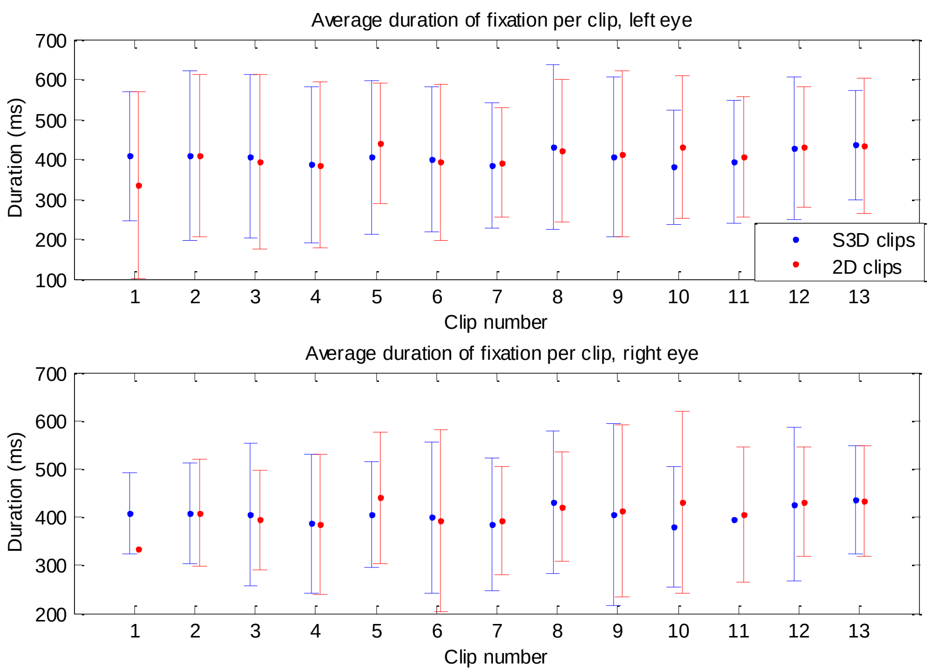


Fig. 5.11. *Mean fixation duration for each clip, pooled across participants and repetitions. Blue data is for S3D and red data is for 2D clips. Errorbars show standard deviations of the mean clip durations. Each point represents 16 datapoints*

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

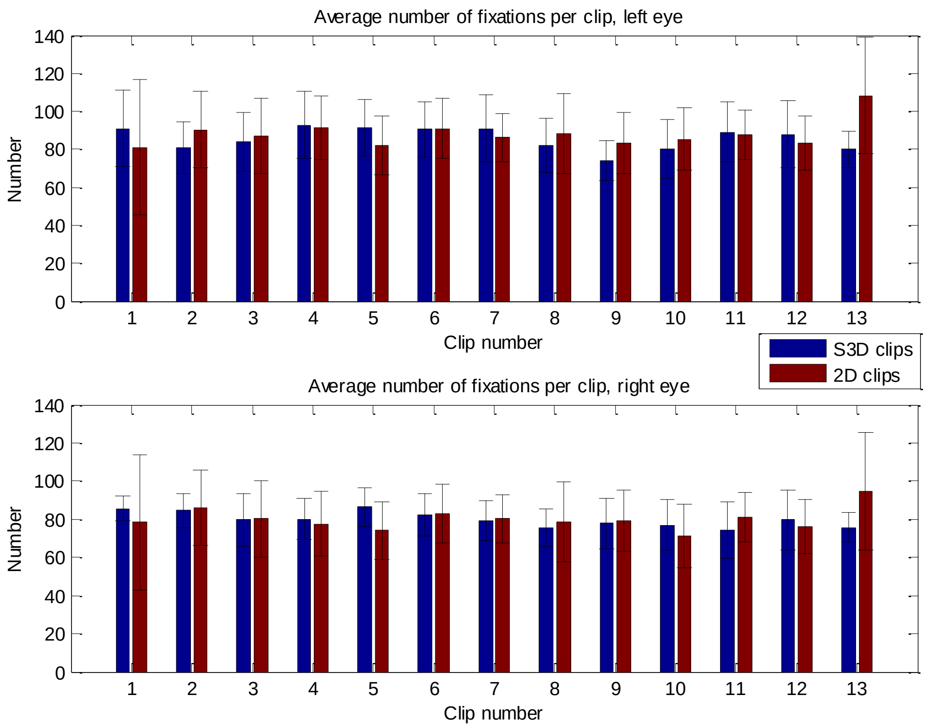


Fig. 5.12. *Mean number of fixations per trial, averaged across participants and repetitions, for left and right eye. Blue bars represent S3D clips, red bars represent 2D clips. Each bar represents 16 datapoints.*

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

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

5.4 Discussion

5.4.1 Subjective ratings

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

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

This conclusion is necessarily limited to the particular clips I used. These were all taken from the same S3D programme 'Micro Monsters', they were all similar in nature (wildlife documentary), and they did not differ in the strength of the depth illusion they created. The logic of the shift manipulation is that weakening cues to the screen plane enables monocular depth cues to dominate perception (Ames Jr, 1925). This would predict that the effect of the shift should be stronger for content with more powerful monocular depth cues. As Koenderink et al write, "A photograph of a brick wall in frontoparallel attitude is not going to reveal any 'zograscopic effect'". More subtly, a 2D photograph of several frontoparallel surfaces at different distances may also display very little zograscopic effect, simply because the monocular depth cues are weak, even though the binocular disparity cues make the surfaces appear clearly separated in depth when viewed in S3D. Could this explain my results? The highly significant increase in depth ratings when the content was displayed in S3D proves that, unlike the brick wall example, my scenes did depict a wide range of depth, but I have not assessed their monocular depth cues objectively. The literature on 'monocular stereopsis' to date contains only cursory discussion of how the nature of the content might affect the strength of the effect (Koenderink et al., 2013). I cannot identify any particular reason why the content I used should be particularly ineffective at producing a 'zograscopic effect'. The literature reviewed in the Introduction used static images, which cannot contain depth cues such as structure-from-motion and looming, whereas my 'Micro Monsters' clips regularly contained these cues. Additionally, the clips were typically of insects filmed in extreme macro, meaning that they contained depth-of-field (blur) cues to three-dimensional structure. Rather than consisting of sets of frontoparallel surfaces with little depth structure within each surface, the clips typically depicted undergrowth, bark and so on extending in depth. Thus, while I cannot rule out that the shift manipulation would have produced a more compelling depth impression with other content, my chosen examples seem likely to have had monocular depth cues at least as strong as other commercial S3D content.

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

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

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

5.4.2 Eyetracking discussion

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

My results are necessarily limited to the content used in my study. In general, disparity-defined depth certainly can influence salience (Lang, Nguyen, Kattie, Yadati, Kankanhalli & Yan, 2012). For example Jansen et al. use natural disparity to create white and pink noise artificial stimuli with different depths, and found that participants fixated closer locations earlier than locations at a greater distance (Jansen, Onat & König, 2009). However, my study wished to focus on the 'real life' scenario and I used commercial content for this reason.

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

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

A potential limitation of the experiment is that despite a rigorous calibration and validation procedure measurements taken with an eyetracker can be subject to a large amount of noise. This is somewhat reflected in my results (as an example, one participant appeared to be verging to 10,000m away based on the parallax between the left and right eye positions, both horizontally and vertically). The vertical difference in left and right eye position was my first indication that the measurements were not going to be 100% accurate. However the Eyelink recommended levels were matched and the data did give us significance and a good indication of gaze, once enough data was collected. Once the technology is more refined, it would be interesting to revisit this study to see if the results hold with less noise.

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

5.5 Conclusion

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

6. Conclusion

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

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

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

There are many results from this thesis that could be relevant and useful in the industry of S3D content production, and will also help to alleviate some of the concerns that are raised when considering S3D production. The eyetracking results suggest that introducing disparity to content does not in fact alter where the viewer fixates, and hence producers don’t need to have special consideration when creating S3D content. It stands to reason that producers of content can use the same considerations that they use during creation of 2D material when creating scenes to attract attention. The subjective results from the same study suggest that it is a waste of producers’ resources to replicate the 2D image in both eyes with a uniform disparity, as the artificial depth created does not generate a more immersive sense of depth than 2D content. If some scenes are displayed in 2D in an overall S3D production it would have as immersive an experience without altering the 2D images, and hence would be more economic. However the viewing distance that the content is shown at is a contributing factor to the viewers impression of size, confirming that the puppet theatre effect is an issue that producers of content need to consider when creating S3D content. A positive result from this study is that when considering the oblique angles chapter. Contrary to what may have been thought, the viewing angle of S3D content reveals no distortions for viewing angles of up 20° from normal. And angles further than that have very similar distortions as that for 2D content, so the same considerations as applied to 2D content can be used, echoing the eyetracking study.

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

It is worth noting that S3D content in the cinema is still very popular, with most mainline cinema releases having an S3D release and a 2D release. The same cannot be said for television, and with the only S3D channels (SKY 3D, and before that, BBC 3D) being removed from streaming content directly, it could be that the home television medium for viewing S3D is not yet a viable option. It may improve with advancing technology and the creation and viewing of S3D content decreasing in price. The conclusions drawn from my thesis may provide some details to help improve the creation, production and viewing of S3D content.

7. References

3D, S. Bskyb technical guidelines for plano stereoscopic (3d) programme content: BSkyB.

Ames Jr, A. (1925). The illusion of depth from single pictures. *J. Opt. Soc. Am, 10*, 137-148.

Ardito, M. (1994). Studies of the influence of display size and picture brightness on the preferred viewing distance for HDTV programs. *SMPTE journal, 103*(8), 517-522.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. IEEE computer graphics and applications, 21(6), 34-47.

Banks, M. S., Held, R. T., & Girshick, A. R. (2009). Perception of 3-D Layout in Stereo Displays. *Inf Disp (1975), 25*(1), 12-16.

Banks, M. S., Read, J. C., Allison, R. S., & Watt, S. J. (2012). Stereoscopy and the Human Visual System. *SMPTE Motion Imaging J, 121*(4), 24-43. doi: 10.5594/j18173

Bereby-Meyer, Y., Leiser, D., & Meyer, J. (1999). Perception of artificial stereoscopic stimuli from an incorrect viewing point. *Perception & Psychophysics, 61*(8), 1555-1563.

Bishop, P. & Pettigrew, J. D. (1986). Neural mechanisms of binocular vision. *Vision Research*, 26 (9): 1587–1600. doi:10.1016/0042-6989(86)90177-x

Brady, M. P. (2009). Picture-Perfect HDTV. *Forbes*.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433-436.

Brenner, E., & van Damme, W. J. M. (1999). Perceived distance, shape and size. *Vision Research, 39*(5), 975-986. doi: http://dx.doi.org/10.1016/S0042-6989(98)00162-X

Brumm, K., Walenski, M., Haist, F., Robbins, S. L., Granet, D. B., & Love, T. (2010). Functional magnetic resonance imaging of a child with Alice in Wonderland syndrome during an episode of micropsia. *Journal of American Association for Pediatric Ophthalmology and Strabismus, 14*(4), 317-322. doi: http://dx.doi.org/10.1016/j.jaapos.2010.03.007

Burks, R., Harper, C., & Bartha, M. C. (2014). Examining 3-D technologies in laptop displays. *Ergonomics in Design: The Quarterly of Human Factors Applications, 22*(3), 17-22.

Cavanagh, P. (1987). Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity, and shape. *Computer Vision, Graphics, and Image Processing, 37*(2), 171-195. doi: http://dx.doi.org/10.1016/S0734-189X(87)80001-4

Claparède, E. (1904). *Stéréoscopie monoculaire paradoxale*.

Cooper, E. A., Piazza, E. A., & Banks, M. S. (2012). The perceptual basis of common photographic practice. *J Vis, 12*(5), 8.

Cullen, B., Galperin, D., Collins, K., Kapralos, B., & Hogue, A. (2012). *The effects of audio on depth perception in S3D games.* Paper presented at the proceedings of the 7th audio mostly conference: a conference on interaction with sound.

Cutting, J. E. (1987). Rigidity in cinema seen from the front row, side aisle. *J Exp Psychol Hum Percept Perform, 13*(3), 323-334.

Da Vinci, L. (2012). *A treatise on painting*: Courier Dover Publications.

Darlin, D. (2005). Falling Costs of Big-Screen TVs to Keep Falling, *NY Times*.

De La Gournerie, J. (1859). *Traité de perspective linéaire contenant les tracés pour les tableaux, plans & courbes, les bas reliefs & les décorations théatrales, avec une théorie des effets de perspective*: Dalmont et Dunod.

Deregowski, J. B. (1969). A pictorial perception paradox. *Acta Psychol (Amst), 31*(4), 365-374.

Devernay, F., & Beardsley, P. (2010). Stereoscopic cinema *Image and Geometry processing for 3-D cinematography* (pp. 11-51): Springer.

Di Luca, M. Domini & F. Caudek, C. (2006). Depth cues do not specify a unique Affine or Euclidean shape representation [Abstract]. *Journal of Vision*, 6(6):340, 340a

Dodgson, N. A. (2004). *Variation and extrema of human inerpupillary distance.* Paper presented at the Stereoscopic Displays and Virtual Reality Systems XI, San Jose, CA.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hillaire, S., Lécuyer, A., Cozot, R., & Casiez, G. (2008). *Using an eye-tracking system to improve camera motions and depth-of-field blur effects in virtual environments.* Paper presented at the Virtual Reality Conference, 2008. VR'08. IEEE.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. Journal of vision, 4(12), 1-1.

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *J Vis, 8*(3), 33.31-30. doi: 10.1167/8.3.33

Hopf, K. (2000). An autostereoscopic display providing comfortable viewing conditions and a high degree of telepresence. *Circuits and Systems for Video Technology, IEEE Transactions on, 10*(3), 359-365. doi: 10.1109/76.836280

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*: Oxford University Press, USA.

Howarth, P. A. (2011). Potential hazards of viewing 3D stereoscopic television and cinema: a review. *Opthalmic & Physiological Optics, 31*, 111-122.

Hudson, W. (1967). The study of the problem of pictorial perception among unacculturated groups. *International Journal of Psychology, 2*(2), 89-107.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on pattern analysis and machine intelligence, 20(11), 1254-1259.

Jahoda, G., & McGurk, H. (1974a). Development of pictorial depth perception: cross-cultural replications. *Child Dev, 45*(4), 1042-1047.

Jahoda, G., & McGurk, H. (1974b). Pictorial depth perception: a developmental study. *Br J Psychol, 65*(1), 141-149.

Jansen, L., Onat, S., & König, P. (2009). Influence of disparity on fixation and saccades in free viewing of natural scenes. Journal of Vision, 9(1), 29-29.

Julesz, B. (1971). Foundations of cyclopean perception.

Julesz, B. (1986). Stereoscopic vision. *Vision Research, 26*(9), 1601-1612.

Karajeh, H., Maqableh, M., & Masa’deh, R. e. (2014). A Review on Stereoscopic 3D: Home Entertainment for the Twenty First Century. *3D Research, 5*(4), 1-9.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What’s new in Psychtoolbox-3. *Perception, 36*(14), 1.

Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. Journal of vision, 7(7), 5-5.

Koenderink, J., Wijntjes, M., & van Doorn, A. (2013). Zograscopic viewing. *i-Perception, 4*(3), 192-206.

Koenderink, J. J., van Doom, A. J., Kappers, A. M. L., & Todd, J. T. (2004). Pointing out of the picture. *Perception, 33*, 513-530. doi: 10.1068/p3454

Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. L. (1994). On so-called paradoxical monocular stereoscopy. *Perception, 23*(5), 583-594.

Kubovy, M. (1988). *The Psychology of Perspective and Renaissance Art*: Cambridge University Press.

Kuo, Y.-T., Chiu, N.-C., Shen, E.-Y., Ho, C.-S., & Wu, M.-C. (1998). Cerebral perfusion in children with Alice in Wonderland syndrome. *Pediatric Neurology, 19*(2), 105-108. doi: http://dx.doi.org/10.1016/S0887-8994(98)00037-X

Kurtz, H. F. (1937). Orthostereoscopy. *Journal of the Optical Society of America, 27*(10), 323-339.

Lambooij, M., Ijsselsteijn, W., Fortuin, M., & Heynderickx, I. (2009). Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaing Science and Technology, 53*(3), 1-14.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Saunders, J. A., & Backus, B. T. (2007). Both parallelism and orthogonality are used to perceive 3D slant of rectangles from 2D images. *J Vis, 7*(6). doi: 10.1167/7.6.7

Scarfe, P., & Hibbard, P. B. (2011). Statistically optimal integration of biased sensory estimates. *J Vis, 11*(7). doi: 10.1167/11.7.12

Schild, J., & Masuch, M. (2011). Fundamentals of stereoscopic 3d game design *Entertainment Computing–ICEC 2011* (pp. 155-160): Springer.

Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. *J Vis, 11*(8), 1-29.

Smith, M. D., & Malia, J. (2015). Controlling Miniaturization in Stereoscopic 3D Imagery. *SMPTE Motion Imaging J, 124*(3), 50-58.

Snowden, R., Snowden, R. J., Thompson, P., & Troscianko, T. (2012). *Basic vision: an introduction to visual perception*: Oxford University Press.

SurveyMonkey. Link to survey. from https://www.surveymonkey.com/r/RZMNFHH

Tanton, N. (2004). Results of a survey on television viewing distance. BBC White paper.

Thouless, R. H. (1931). Phenomenal Regression to the Real Object. *Nature, 131*, 261-263.

THX recommended viewing distances. from http://www.thx.com/test-bench-blog/what-size-tv-should-i-buy-tech-tips-from-thx/

Todd, J. (1955). Alice in Wonderland" syndrome. *Can Med Assoc J, 73*, 701-704.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. Artificial intelligence, 78(1), 507-545.

Ullman, S. (1979). The interpretation of structure from motion. *Proc R Soc Lond B Biol Sci, 203*(1153), 405-426.

Verhoeff, F. H. (1935). A new theory of binocular vision. *Archives of Ophthalmology, 13*(2), 151-175.

Vinci, L. d. (2013). *A treatise on painting*: Read Books Ltd.

Vishwanath, D., Girshick, A. R., & Banks, M. S. (2005). Why pictures look right when viewed from the wrong place. *Nature Neuroscience, 8*(10), 1401-1410.

Vishwanath, D., & Hibbard, P. B. (2013). Seeing in 3-D with just one eye stereopsis without binocular vision. *Psychological science*, 0956797613477867.

Wallach, H., Frey, K. J., & Bode, K. A. (1972). The nature of adaptation in distance perception based on oculomotor cues. *Perception & Psychophysics, 11*(1), 110-116.

Walsh, V., & Kulikowski, J. (1998). *Perceptual constancy: Why things look as they do*: Cambridge University Press.

Watt, S. J., Akeley, K., Girshick, A. R., & Banks, M. S. (2005). Achieving near-correct focus cues in a 3d display using multiple image planes. In Electronic Imaging 2005 (pp. 393-401). International Society for Optics and Photonics.

Welchman, A. E., Deubelius, A., Conrad, V., Bülthoff, H. H., & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in human visual cortex. *Nat Neurosci, 8*(6), 820-827.

Wheatstone, C. (1838). Contributions to the Physiology of Vision. Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision. *Philosophical Transactions of the Royal Society of London, 128*, 371-394. doi: 10.1098/rstl.1838.0019

Wilks, S. S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. 60-62. doi: 10.1214/aoms/1177732360

Woodgate, G. J., & Ezra, D. (1995). Autostereoscopic display device: Google Patents.

Woods, A., Docherty, T., & Koch, R. (1993). Image Distortions in Stereoscopic Video Systems. *Stereoscopic Displays and Applications Iv, 1915*, 36-48. doi: Doi 10.1117/12.157041

Woods, A. J. (1993). *Image distortions in stereoscopic video systems.* Paper presented at the SPIE 1915, Stereoscopic Displays and Applications IV, 36 (September 23, 1993).

Woods, A. J. (2011). *How are crosstalk and ghosting defined in the stereoscopic literature?* Paper presented at the Stereoscopic Displays and Applications XXII, San Francisco Airport, California, USA.

www.mathworks.com.

Yamanoue, H., Okui, M., & Okano, F. (2006). Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images. *Circuits and Systems for Video Technology, IEEE Transactions on, 16*(6), 744-752. doi: 10.1109/TCSVT.2006.875213

Zone, R. (2007). *Stereoscopic cinema & the origins of 3D film*: University Press of Kentucky.

Zorin, D., & Barr, A. H. (1995). *Correction of geometric perceptual distortions in pictures.* Paper presented at the Proceedings of the 22nd annual conference on Computer graphics and interactive techniques.