Human use of horizontal disparity for perception and visuomotor control

Peter Scarfe

Submitted for the degree of Doctor of Philosophy

Friday 16th March 2007

I, Peter Scarfe hereby certify that this thesis, which is approximately 66,840 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date .... Signature of candidate ....

I was admitted as a research student in October 2003 and as a candidate for the degree of Doctor of Philosophy in March 2007; the higher study for which this is a record was carried out in the University of St Andrews between 2003 and 2007.

Date .... Signature of candidate ....

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date .... Signature of supervisor ....

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration.

Date .... Signature of candidate ....

Acknowledgements

First and foremost I am hugely indebted to my supervisor Dr. Paul Hibbard. He has been a dedicated and supportive supervisor, not to mention an excellent scientist, who has always made time to help me in any way possible. I have had an excellent time working with him and have benefited immeasurably. I have also had help and support from all members of the Vision Lab here in St Andrews. In particular, Dr. Samira Bouzit (Sam) helped me hugely with learning programming and provided many a helpful and relaxing chat over coffee. I have also had many discussions with Dr. Harold Nefs and Dr. Vit Drga, on a diverse range of topics, from which I have benefited greatly. I also owe much thanks Dr. Simon Watt who I had the pleasure of working with in Bangor, Wales, during one summer, he taught what careful science is all about. I benefited greatly from working with him and had an excellent time in the process.

The most important thing I have learned whilst doing the PhD is that family and friends are the most important things above all else. Without their help and support I simply would not have got through this. I owe everything to Mum, Neil, and Chris, Dad and Peggy who have been there through everything. I couldn’t possibly say enough.

My time in St Andrews would have been nothing without all the great people I have met, I’m sure this is by no means comprehensive, but in no particular order… Anja (Dr. Farovik, thanks for the massive music enthusiasm, great times and many late nights thesis writing), Doog (without doubt always there to aid the escape from work, since way back when), April (social convener and all round amazing friend), Ash (I will definitely miss the chat over whiskies, and for some reason sunrises on the beach), Paul (you are so much more than a pseudo-Ash! thanks), Neal (music guru), Scott (what can I say), Jenny, Shelly, Erin, Colin (Dr. Snodgrass), Al (Dr. McWilliam), Martin (Dr. O’Neil), David (Dr. Wilson), Ian, Lynn, Jamie, Lydia, Graham, Laura (a mountain of support, I don’t know what I’d do without you), Rolf, David (Chippy), Alex, Marietta (I’m glad I finally unplugged from my headphones), Erica, Blerina, Simon and Doug. I owe you all so much thanks in so many ways. I hope I can repay you all some day. I have also had a great flat sharing times at Boase and elsewhere with Anja, Marietta, Blerina, James and Jordi and Al.

My thanks also go to Andrew Burnley and Pete Wilcox who provided expert help and assistance in manufacturing experimental equipment for much of the research presented in this thesis. Peter Foldiak, was a very helpful second supervisor. Thanks also go to Dr. Mike Oram and Dr. Eric Bowman who have both given me statistical advice for some of the chapters (any mistakes remaining are all my own).

The experiment reported in Chapter five was carried out at the University of Bangor, Wales with Dr. Simon Watt. This research trip would not have been possible without generous funding from a Richard Eagle Award from the Applied Vision Association. Prof. Steve Tipper generously allowed use of his lab and equipment. My thanks also go to Chie Takahashi, Dr. Alison Wiggett, Richard Piech, Dr. Andrew Bayelis and Annukka Lindell, who provided much help with carrying out the research and the organisation of the trip. Everyone made me really welcome.

Well, I think that’s everything, writing this page has seemed so far away at times.

My apologies to anyone I have missed out.

To Dad

Abstract

Our eyes are horizontally separated in the head by approximately 6.5cm. As a result of this separation there are subtle differences in the position of corresponding image points within the two eyes. The horizontal component of this binocular positional difference is termed horizontal disparity. Horizontal disparity is an important visual cue as once scaled with an estimate of the viewing distance, it can theoretically provide full metric information about the structure of the world. This thesis will address the issue of how binocular visual cues are used by the human visual system for the estimation of three-dimensional (3-D) shape for perception and visuomotor control. The research presented is particularly focused on understanding why biases in the perception of 3-D shape from binocular cues are found, their importance for perception and visuomotor control and how these biases may be overcome by combining binocular cues with other sources of visual information.

Table of contents

Chapter One

Introduction

1 Introduction

1.1 Obtaining information about the world

1.1.1 The nature of visual cues

Perception is a process of inferring the structure of the world from current and past sensory information. Information about the world is systematically transformed by the way in which it is sensed. This means that the visual system must take account of this transformation in order to recover useful information about the world. Many diverse sensory systems have evolved, these range from the more familiar senses of touch, vision, olfaction and audition, to those such as echolocation (Moss & Sinha, 2003), electrolocation (Assad, Rasnow & Stoddard, 1999) and the 'magnetic sense' (Walker, Dennis & Kirschvink, 2002). Organisms are adapted to a specific environmental niche, so will require different types of information in order to survive and reproduce. Sensory systems, although diverse in their nature, have each evolved to provide an organism with sufficient information to control its behaviour adaptively. For primates such as humans, vision is one of the most important means by which we obtain information about the world.

Sensory information is statistically structured by the regularities present in the world, the brain can learn to exploit these statistical regularities to infer the structure of the world (Haijiang, Saunders, Stone & Backus, 2006). Visual information is generally broken down into various visual 'cues'. There is no fixed definition of what constitutes a cue, visual or otherwise. Ernst and Bülthoff (2004) use the term to mean "any sensory information that gives rise to a sensory estimate" (p. 163), this will be the usage that is adopted here. Identifying cues in the information available from a sensory organ suggests that there are independent statistical regularities in that organ's information, which reliably signal specific aspects of the environment. For example, the projective foreshortening of elements in a homogeneous isotropic texture carries information about surface shape and orientation (Knill, 1998).

The statistical regularities that constitute visual cues fall along a continuum in how widely applicable their use is. Some cues, such as binocular disparity, can theoretically be used in a broad range of circumstances to estimate many different environmental attributes, such as the 3-D shape and position of objects in the environment. Binocular disparity is a direct consequence of the geometric relationship between the position of an observer's eyes and the scene. Other cues, such as texture, can be used widely to estimate different environmental attributes, but require certain statistical assumptions to be made in order that estimates, such as the estimated slant of a surface, can be made from that cue's information (Knill, 1998). The assumptions required for the use of textural information are discussed at greater length below.

Other cues occur under very specific viewing circumstances and/or provide an estimate of a single environmental property (Todd & Norman, 2003). These cues have been labelled as 'non-generic' and 'context specific' or as 'heuristics' to signify that they are only available under restricted viewing circumstances. However, beyond their range of use there is no fundamental difference between these and more standard textbook visual cues, all are regularities in the sensory input that allow useful information about the world to be obtained. The identification of visual and non-visual cues is dependent on the observer's sensory organs, but at the same time independent of the observer, as the identification of a cue does not necessarily mean that it is, or can be, exploited. In contrast the definition of a cue given by Ernst and Bülthoff (2004) suggests that a cue needs to give rise to a sensory estimate in order to be labelled as such.

The identification of cues is also dependent on the level of analysis at which this is investigated. For example visual texture can be treated both as a single unitary cue (Rosas, Wagemans, Ernst & Wichmann, 2005), or in terms of the component sources of information that make up a texture (Knill, 1998). This highlights that fact that the definition and constitution of visual cues is dependent on understanding how the brain combines and uses different sources of information. Visual cues are typically broken down into retinal and extra-retinal cues. Retinal cues are those that are available to us directly from the retina, and include cues such as texture, colour, occlusion and shadow. Extra-retinal cues are those that we gain from the eyes but not directly from the retina, for example the physical orientation of the eyes in their orbits (Howard & Rogers, 2002).

1.1.2 Binocular and monocular visual information

Humans have two forward facing eyes, laterally separated in the head by around 6.5cm (Howard & Rogers, 2002); this means that each eye receives a slightly different view of the world. For a given static field of view each eye receives information both from regions seen by one eye alone and regions seen by both eyes (Gillam & Borsting, 1988). Within the region where both eyes share a view of the same portion on the scene binocular differences in the eyes' images are present, these are termed binocular disparities (Figure 1.1). The binocular disparities between the two eyes' images include horizontal and vertical differences in the location of corresponding points across the two eyes images; these are termed horizontal and vertical disparity respectively. The geometry of horizontal disparity is demonstrated in later Figure 1.5 and the geometry of vertical disparity later in Figure 1.7.

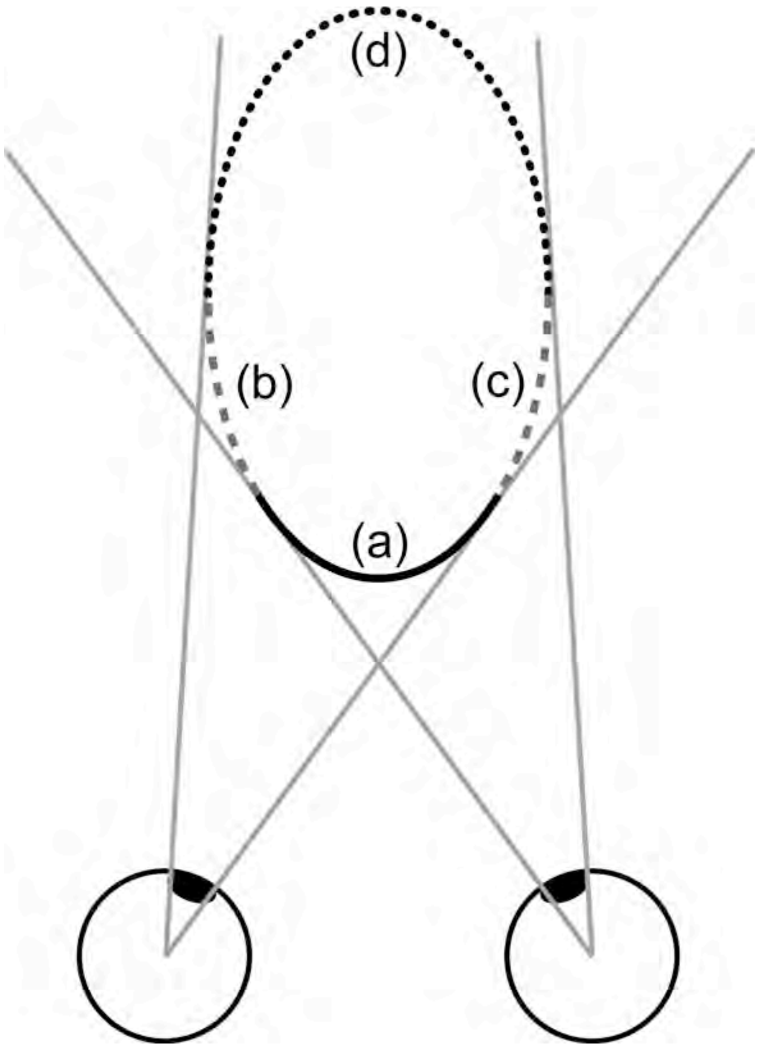

Figure 1.1: A cylinder is viewed on the median plane at a fixed distance from the eyes. The grey lines coming from each eye show the regions of the cylinder's surface viewable from each eye. The solid black region of the cylinder's surface labelled (a) can be seen by both eyes, the grey dashed region labelled (b) can be seen by only the left eye, and the grey dashed region labelled (c) by only the right eye. Neither eye can see the black dotted region labelled (d). Binocular disparities between the two eyes' images are only present in the areas of the scene that can be viewed by both eyes, here surface region of the cylinder labelled (a).

The shift in corresponding points in the two eyes' images can be seen in Figure 1.2, which shows a pair of binocular images of a natural scene taken with two cameras positioned 6.5 cm apart, and the two images superimposed on top of one another. By considering a single eye's image in isolation (Figure 1.2a and 1.2b) it is clear that this is an image of a duck with its head extended forward in depth relative to its body. The duck is sitting in front of some leafy foliage on a mossy surface sloping toward the viewer. There are three rocks to the duck's left; two of these rocks are at approximately the same distance as the duck from the viewer, whereas one is to the front of the duck in the foreground. Some of the leaves in the background foliage have smooth relatively flat surfaces, whereas others are thin, branched and rough in texture. From this brief description it is clear that there are many retinal cues available to the visual system from which judgements about the position, distance, texture and slope of objects in the scene can be made without the need for binocular cues.

Many monocular sources of information are however inherently ambiguous with regard to the metric structure of the scene (Hershenson, 1999, Koenderink, 2001). In some instances metric structure can be estimated from monocular information, but certain assumptions must be made in order to do this. As an example, consider estimating the slant of a textured planar surface. Because an observer is unaware of the precise statistical structure of the surface texture in the world they are unable to estimate its slant definitively. Knowledge of the texture's structure is needed to undo the perspective projection that maps the texture of the physical surface to the texture present in the eye's image of that surface. However, if the texture is assumed to be homogenous and isotropic, the observer can make a statistically optimal guess as to the slant of the surface in the world (Knill, 1998).

Given the ambiguity of many monocular cues, binocular disparity provides a potentially important source of information for guiding behaviour. The superimposed images in Figure 1.2b demonstrate the small differences that exist between the two eyes' views of the world; these provide potentially valuable information about the structure of the scene. The horizontal and vertical shifts of corresponding points in each eye can be seen by the non-registration of features in the superimposed images. This is particularly clear when looking at objects in the foreground of the image, such as the nearest rock, as the magnitude of the disparity is greater. The fact that the eyes are predominantly horizontally, rather than vertically, separated means that horizontal disparities tend to be much larger than vertical disparities.

Figure 1.2: Images (a) and (c) show a pair of binocular images taken with a pair of digital cameras positioned 6.5cm apart (this is the typical interocular separation of the human eyes). Image (a) is that taken from the position of the left eye and image (c) from the position of the right eye. From these images it is clear that there are many sources of information, or cues, available to the observer from which to make estimates about the shape, size, distance, orientation and layout of objects and surfaces in the scene. Image (b) shows the left and right eyes' images superimposed on top of one another. The non-registration of points in the overlaid images demonstrates the differences that exist between the two eyes' views of the world, that is, the horizontal and vertical disparity that exists between the two eye's images. (Images are courtesy of Dr. Samira Bouzit.)

Horizontal binocular disparity[1] is a particularly useful source of information as once suitable scaled it can geometrically specify the full three-dimensional structure of the scene and is sufficient alone for the perception of depth (Howard & Rogers, 2002). This is strongly demonstrated by the perception of depth from random dot stereograms (Julesz, 1971). A random dot stereogram and its component images are shown in Figure 1.3, the left and right eyes' images are shown in (a) and (b) respectively and a red green anaglyph containing both images is shown in (c). The left

[1] Throughout this thesis horizontal binocular disparity will be referred to interchangeably as 'horizontal disparity' and 'disparity'. Where particular reference is being made to the different types of disparity present between the eyes' images, such as vertical disparity, this will be referred to specifically so as to avoid confusion.

and right eyes' images are composed entirely of random dots, but with an identical square of random dots shifted an equal and opposite amount horizontally in the two images. This horizontal shift introduces a horizontal image disparity between the two images, which is perceived as depth when the images are fused when presented appropriately to each eye in the anaglyph. This is because the disparity present between the images is consistent with a square of dots in a different depth plane to the rest of the image. This demonstrates that disparity alone is sufficient for the perception of depth. Even in a natural image where there are multiple redundant sources of information about depth and shape, the addition of binocular disparity strongly enhances our subjective impressions of perceived depth (Figure 1.4).

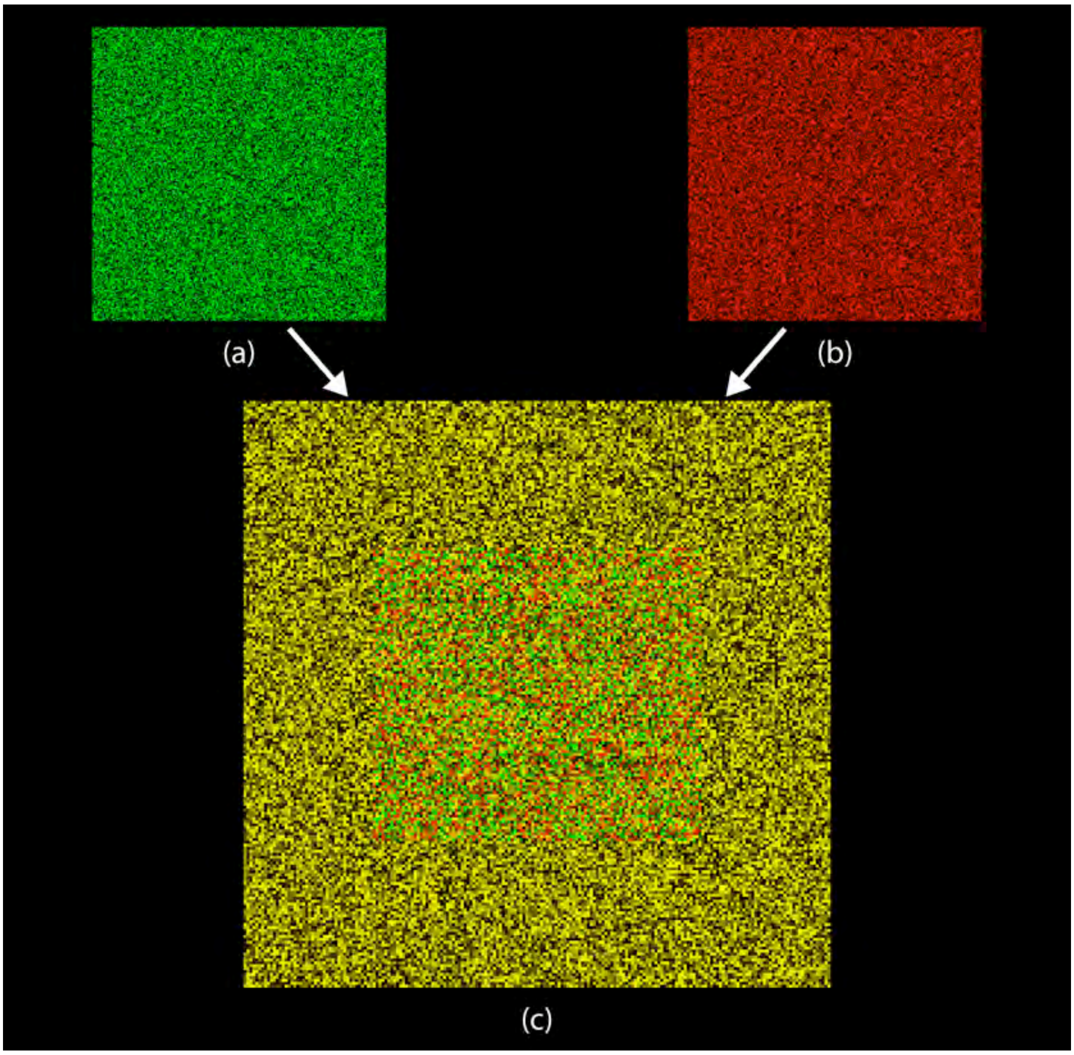

Figure 1.3: Shows a random dot stereogram as a red-green anaglyph and its component images. The left and right eyes' images labeled (a) and (b) respectively, are each composed of randomly positioned dots. The left eye's image is coloured green and the right eye's image is coloured red. Within each eye's image an identical square of random dots has been placed in the image and shifted horizontally by an equal and opposite amount (leftward in the right eyes' image and rightward in the left eye' image). This shift introduces a binocular disparity between the two images, which is perceived as depth when the images are viewed appropriately in the anaglyph (green lens over the left eye and red lens over the right eye). A square of random dots should be perceived in a depth plane nearer to the observer than the rest of the image. This is because the disparities present in the images forming the anaglyph are consistent with a square of dots in the centre of the image in different depth plane to the rest of the image. The perception of depth in random dot stereograms demonstrates that disparity alone is sufficient for the perception of depth.

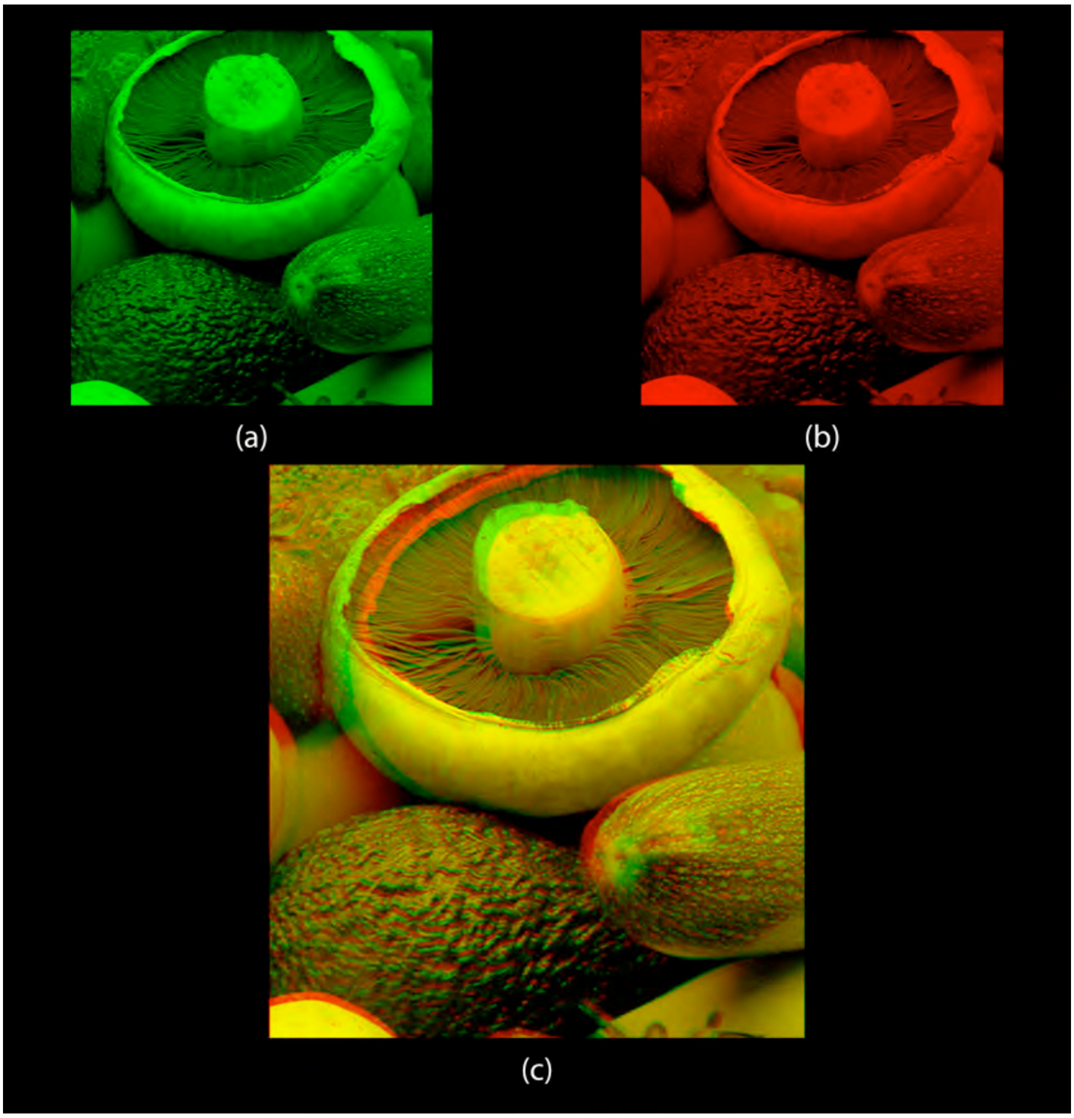

Figure 1.4: Images (a) and (b) show a set of binocular images taken with a pair of digital cameras positioned 6.5 cm apart (as in Figure 1.2). These images are presented in the same form as the random dot stereogram shown in Figure 1.3. The left eye's image (a) is coloured green and the right eye's image (b) is coloured red and a red-green anaglyph of the two images is shown in (c). From each eye's image it is clear that there is sufficient information available to make confident judgements about the structure of the scene. However, the addition of binocular information when viewing the anaglyph (green lens over the left eye and red lens over the right eye) strongly enhances the subjective impression of 3D structure. Some part of this subjective difference may result from cues to the surface slant flatness of each eye's image lessening the perceived depth from monocular cues. (Images courtesy of Dr. Samira Bouzit and Dr. Harold Nefs)

1.1.3 The geometry of horizontal disparity

The basic geometry of horizontal disparity is shown in Figure 1.5. When fixating point F at a distance D_F corresponding points within the two eyes' retinas are stimulated and the lines of sight from the two eyes define the vergence angle θ_F. Similarly, when fixating point A at a distance D_A the vergence angle θ_A is defined and when fixating point B at a distance D_B the vergence angle θ_B is defined. The interocular separation is labeled I. The depth between the fixated point F and the non-fixated point B, denoted by d, is equal to $D_F - D_B$. The points F and B produce an angular separation of α_L in the left eye and α_R in the right eye. The difference between these two angles defines the *absolute* horizontal disparity between the points F and B, $\eta_{ab} = \alpha_L - \alpha_R$. This is equivalent to the difference between the vergence angles produced by the two points when fixated, $\eta_{ab} = \alpha_L - \alpha_R = \theta_B - \theta_F$.

The *absolute* disparity of a point is defined relative to fixation, so as the vergence angle changes to fixate a different point in depth the absolute disparities of points in the scene will also change. However, the *relative* disparity between points does not change, this is because the relative disparity between points is not defined with regard to the point of fixation. Relative disparity can be demonstrated by considering the depth between the points A and B, both of which are non-fixated. The depth between A and B is denoted by d', is equal to $D_A - D_B$. With fixation on point F the points A and B produce an angular separation of ρ_L in the left eye and ρ_R in the right eye. The difference between these two angles defines the *relative* horizontal disparity between the points, $\eta_{rel} = \rho_L - \rho_R$. This is equivalent to the difference between the vergence angles when fixating points A and B, $\eta_{rel} = \rho_L - \rho_R = \theta_B - \theta_A$.

It is easy to see the difference between absolute and relative disparity from Figure 1.5. If point F moved forward or backward in depth its visual direction on the retina would change, this would alter the angles α_L in the left eye and α_R in the right eye, the dark grey line in each eye would alter its position relative to the light grey line in each eye. This would cause the *absolute* disparity of point B to change. However, the relative positions of the light grey and black lines in the left and right eye would not change, this would result in the angles ρ_L in the left eye and ρ_R in the right eye remaining the same, the *relative* disparity between point A and B would therefore not change. This highlights how the absolute disparity of a point in the scene is dependent on fixation whereas the relative disparity of a point in the scene is independent of fixation.

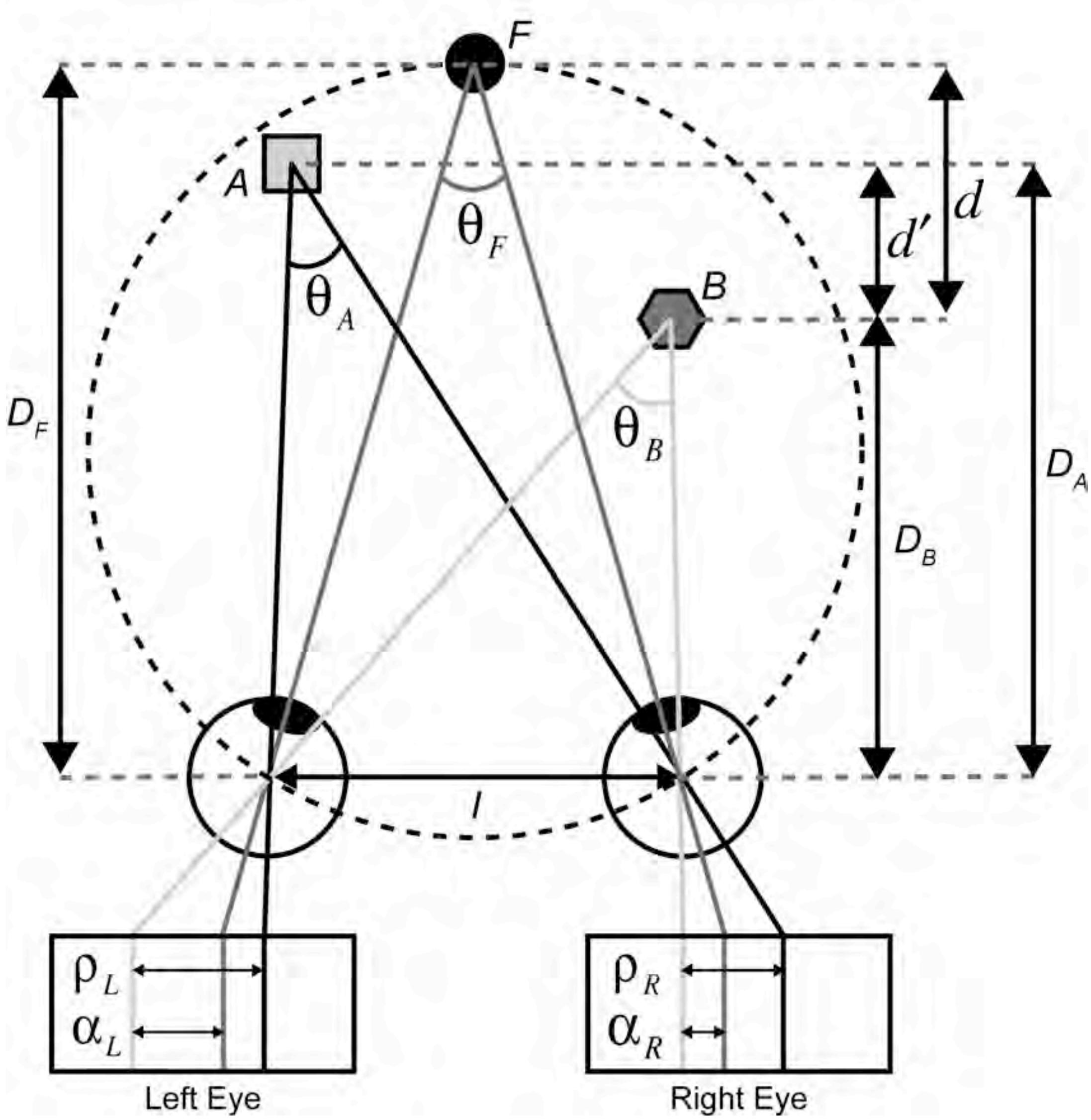

Figure 1.5: A schematic diagram of the basic geometry of horizontal disparity. A scene of three objects (F, A and B) positioned within the plane of the eyes is shown as viewed from above. The projected angular position of these points in the retina of each eye is shown in the to rectangles positioned below each eye. For a detailed description see the accompanying text.

1.1.4 *Scaling horizontal disparity*

Assuming that a depth interval d (such as the depth of an object) is much smaller than the viewing distance D and given the small angle approximation ($\theta \approx \tan\theta$), the projected retinal disparity of this depth can be expressed as,

$$\eta \approx \frac{Id}{D^2} \tag{1.1}$$

From equation 1.1 (for a derivation see Appendix 1) it can be seen that the projected retinal disparity, η, is dependent on the depth of the object in the environment, d, the interocular separation, I, and the viewing distance, D. More specifically the projected retinal disparity scales inversely with the square of distance, $\eta \propto 1/D^2$. For a given depth the projected retinal disparity decreases by a factor of four if the viewing distance is doubled. In order for the viewer to obtain an estimate of absolute depth from the projected retinal disparity, an estimate of the viewing distance is needed as well as knowledge of the interocular separation. This is more easily seen by rearranging equation 1.1 to make d the subject,

$$\hat{d} \approx \frac{\hat{\eta}\hat{D}^2}{\hat{I}} \tag{1.2}$$

In equation 1.2, $\hat{d}$ now represents the visual system's estimate of object depth, and $\hat{D}$ its estimate of object distance. It is generally assumed that the visual system's knowledge of interocular separation and measured retinal disparity are unbiased i.e. $\hat{I} = I$ and $\hat{\eta} = \eta$ respectively. If the distance to the object is estimated correctly ($\hat{D} = D$) the depth of the object will also be estimated correctly ($\hat{d} = d$). However, if the distance to the object is misestimated ($\hat{D} \neq D$) the depth of the object will also be misestimated ($\hat{d} \neq d$). Now consider the same object of a fixed depth d presented at a range of viewing distances, $D_0 \ldots D_n$. If the visual system is able to estimate each distance correctly such that $\hat{D}_0 \ldots \hat{D}_n = D_0 \ldots D_n$, the estimated depth of the object $\hat{d}$ will be invariant with viewing distance (Figure 1.6), this is known as depth constancy

(Howard & Rogers, 2002). However, if the visual system is unable to correctly estimate each distance the perceived depth of an object might change as a function of its distance.

As an example consider the case where all viewing distances within a range are under- or over- estimated by a fixed amount (Figure 1.6a), in this case depth will also be under- or over- estimated, but by an amount that depends on the actual viewing distance (Figure 1.6b). Similarly, in the case where estimated distance is a linear function of actual distance but near distances are overestimated and far distances are underestimated (Figure 1.6a), depth will be overestimated at near distances and underestimated at far distances, again by an amount that depends on the actual viewing distance of the object. In each of the above cases depth constancy would no longer hold, the perceived depth of a physically constant object would depend on the distance at which it was viewed. This highlights the fact that a given retinal disparity is consistent with different depths at different distances (viewing the anaglyphs in Figures 1.3 and 1.4 from different distances demonstrates the dependence of depth on distance). Specifically, perceived depth is proportional to the square of perceived distance, $\hat{d} \propto \hat{D}^2$. As will be discussed below, a large body of research has focused on understanding the extent to which disparity is appropriately scaled in the estimation of 3-D shape.

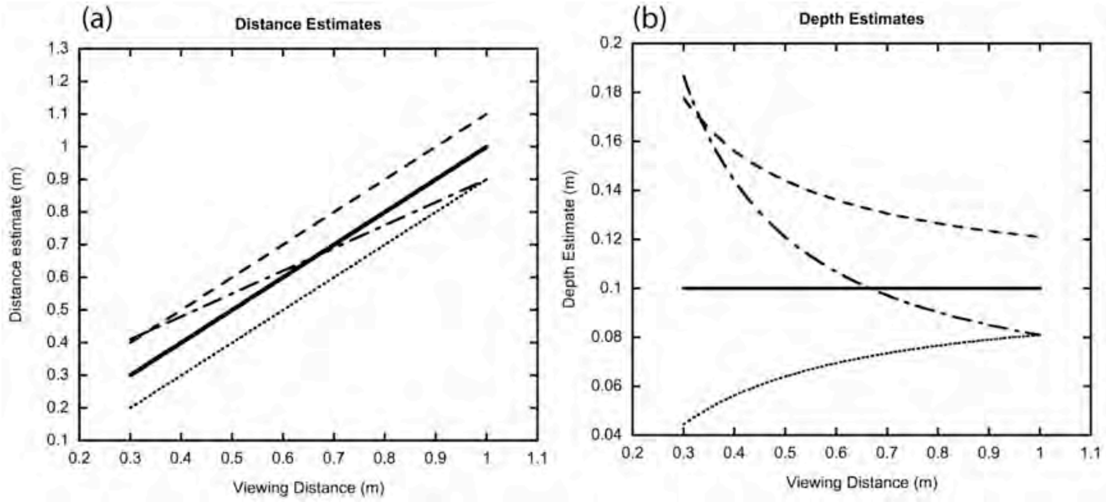

Figure 1.6: Demonstrates the relationship between estimated distance and estimated depth. (a) Shows four distance functions that plot estimated distance against physical distance. The solid black line shows the case where estimated distance is a linear function of physical distance with a slope of one and intercept of zero i.e. veridical estimation of distance. The dashed line shows the case where estimated distance is a linear function of perceived distance with a slope of one, but with a fixed constant error of 0.1m i.e. all distances are overestimated by a fixed amount. Similarly, the dotted line shows the case where all distances are underestimated by the same fixed amount. Finally, the dot-dash line shows the case where perceived distance is a linear function of physical distance, but near distances are overestimated and far distances are underestimated. (b) shows the estimated depth of a 0.1m deep object as a function of each of the perceived distance functions shown in (a). See accompanying text for discussion.

1.1.5 Visual cues for disparity scaling

Horizontal disparity alone can only provide relative depth information, absolute depth information can only be gained once an estimate of the viewing distance is obtained, and used to scale disparity, or when multiple cues are combined. If the visual system were able to accurately scale horizontal disparity it could recover geometrically correct information about the structure of the environment, this could provide potentially valuable information for the control of behaviour. There are numerous retinal and extra-retinal cues available that may be used to scale horizontal disparity. A number of these cues will be introduced here and discussed more fully in subsequent sections of this and other chapters. The position of the eyes in their sockets provides extra-retinal information in the form of the vergence angle (Bradshaw, Glennerster & Rogers, 1996, Rogers & Bradshaw, 1993) and height in the visual field (Howard & Rogers, 2002).

The role of the vergence angle in scaling disparity is discussed throughout particularly in section 1.4 of this chapter and in detail throughout chapter 3, so discussion will be brief here. The vergence angle, illustrated in Figure 1.5, is related to distance from the observer by equation 1.3. Here D_F denotes the distance to a point of fixation F along the median plane, I is the interocular distance and θ_F the vergence angle produced by fixating point F. If the observer is able to correctly estimate the vergence angle of the eyes and has knowledge of the interocular distance they would be able to accurately estimate absolute distance by using the relationship shown in equation 1.3. This distance information could be used for the scaling of horizontal disparity in the estimation of 3-D shape.

$$D_F = \frac{(I/2)}{\tan(\theta_F/2)} \tag{1.3}$$

Height in the field refers to the angle of gaze relative to a line of sight parallel to a ground plane or supporting surface, which is typically orthogonal to gravity. Height in the field can also be used to estimate absolute distance, if the observer is provided with information about the support surface and makes some plausible assumptions about the viewing situation. Detailed discussion of height in the field as a cue to distance and its role in scaling disparity is provided in Chapter 4. Retinal motion caused by either motion of an object or the observer could also provide an estimate of object shape and therefore object distance, which could be used to scale retinal disparity (Richards, 1985). Object and observer motion as a cue to 3-D shape and distance are discussed further in section 1.4.4 of this chapter and in more detail in Chapter 4. The vertical differences in the position of corresponding points on the two retinas, termed vertical disparity, a component of differential perspective, can also provide information about object distance which could be used for the scaling of horizontal disparity (Bradshaw et al., 1996, Mayhew & Longuet-Higgins, 1982).

Vertical disparity is a consequence of points that are not on the median plane of the eyes being closer to one eye than the other (Bradshaw et al., 1996), this is illustrated in Figure 1.7. Consider an observer fixating on the mid-point F of a vertical line, defined by two points P_A and P_B, which is positioned to the left of the median plane of the eyes. The line will subtend a larger vertical visual angle in the nearer left eye (ϕ_L) compared to the further right eye (ϕ_R), due to the different perspective view each eye has of the line. In terms of the end points of the line, point P_A will project lower relative to fixation in the left compared to right eye and point P_B will project higher relative to fixation in the left compared to right eye. This shift in the relative position of each point in the two eyes' images geometrically determines that point's vertical disparity. The combination of horizontal and vertical disparity can be used to estimate geometrically correct depth (Bradshaw et al., 1996, Rogers & Bradshaw, 1993). Much work has focused on how depth and shape are recovered from horizontal disparity alone or in conjunction with other visual cues.

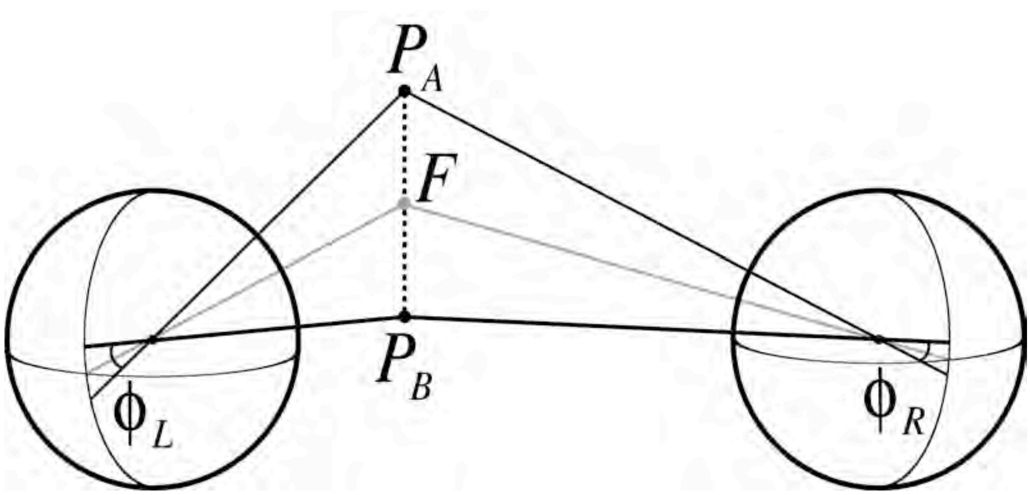

Figure 1.7: Schematic diagram showing the basic geometry of vertical disparity. An observer fixates point F, the mid-point on a vertically orientated line defined by points P_A and P_B. This line is positioned to the left of the median plane of the eyes and subtends an angle ϕ_L in the left eye and ϕ_R in the right eye. This causes the projected images of points P_A and P_B in one eye to be shifted relative to their projected images in the other eye; this shift in position between the eyes defines the point's vertical disparity.

1.2 Cue combination

1.2.1 Models of sensory fusion

The brain receives multiple sources of information about the world, which under most circumstances are fused to create a single unified percept. Models of how the brain fuses information fall along a continuum from weak to strong fusion (Clark & Yuille, 1990). In terms of depth cue combination, weak fusion posits that multiple independent sources of depth information are available from separate cues such as texture, disparity and motion parallax. In the most basic form of weak fusion these estimates are then averaged to give a single depth estimate to act on. The fundamental problem with weak fusion in its extreme form is that different depth cues provide information in non-common units and co-ordinates, which cannot be meaningfully averaged with one another (Landy, Maloney, Johnston & Young, 1995). For example, object depth can be specified by both disparity and occlusion, but whereas disparity can geometrically provide absolute depth information, occlusion can only specify depth order. This means that information from these cues cannot be meaningfully averaged with one another to give a single estimate of object depth (Burge, Peterson & Palmer, 2005).

A further problem with weak fusion in its simplest form is that all sources of information are given equal influence on the final percept. This is problematic as a cue's reliability will vary with both the viewing geometry (Gepshtein & Banks, 2003, Jacobs, 2002) and over time. It would make sense to weight cues by their reliability, in this way more reliable information has a greater influence on the final percept (Clark & Yuille, 1990, Ernst, 2006, Landy et al., 1995). A number of experimental studies have shown that cue reliability is indeed taken into account when cues are combined and recalibrated (Atkins, Fiser & Jacobs, 2001, Atkins, Jacobs & Knill, 2003, Jacobs, 2002). Moving along the continuum toward strong fusion results in progressively more interactions being allowed between cues before a final estimate of depth is produced. In its extreme form strong fusion allows multiple nonlinear interactions to occur between cues in producing a depth estimate, such that the notion of separable depth cues loses meaning (Clark & Yuille, 1990, Landy et al., 1995).

The fundamental problem with strong fusion is that it places no limits on the interactions allowed between cues, information is treated as a whole and is not partitioned into separate cues. Although the visual system could combine information in this way, strong fusion makes it difficult to investigate cue combination experimentally, as it is essentially unfalsifiable (Landy et al., 1995). With unlimited interactions allowed between different sources of information, understanding how the visual system makes perceptual estimates becomes an intractable problem. Landy et al. (1995) have suggested modified weak fusion (MWF) as a model of the depth cue combination. MWF is based on weak fusion, but allows additional interactions between cues before they are averaged. Cues are allowed to 'promote' one another and can use information from 'ancillary cues' in order that they provide depth estimates in common depth units that can be meaningfully averaged.

Promotion refers to interactions between individual cues that allow the specification of missing parameters needed for cues to specify depth in common units e.g. the interaction of stereo and motion cues to specify 3D shape (Richards, 1985). Information from ancillary cues is used in a similar way, but in this case information is gained from cues that do not themselves provide a depth estimate for integration (Landy et al., 1995). Clearly promotion and the use of information from ancillary information may be hard to disambiguate. For example, distance information from the vergence state of the eyes could be used to scale disparity in order to estimate depth, which would be termed the use of ancillary information. However, if perceived 3-D shape were reducible to a series of distance estimates across an object's surface vergence could provide direct information about 3-D shape for combination (Bingham, 2005).

The final stage in MWF is a linear combination of cue information where each cue is weighted in relation to its reliability (Landy et al., 1995). For example, consider disparity and texture as cues to the depth of an object, each provides an independent estimate of the depth of the object, $\hat{D}_d$ from disparity and $\hat{D}_t$ from texture, which are combined through weighted averaging (Equation 1.4). Assume that each of these cues is an unbiased estimator of object depth and is subject to independent Gaussian noise with a variance of σ_d and σ_t respectively. The depth estimate from disparity receives a weight of w_d and the depth estimate from texture a weight of w_t, these weights are based on the relative reliability of the cues (Equations 1.5 and 1.6) and together sum to one, $w_d + w_t = 1$. The reliability of each cue is given by the inverse of its variance (Equations 1.7 and 1.8), the more variable the cue the lower its reliability and the less weight it receives in producing the final depth estimate (Hillis, Watt, Landy & Banks, 2004, Landy et al., 1995, Oruc, Maloney & Landy, 2003).

$$\hat{D} = w_d \hat{D}_d + w_t \hat{D}_t \tag{1.4}$$

$$w_d = \frac{r_d}{r_d + r_t} \tag{1.5}$$

$$w_t = \frac{r_t}{r_d + r_t} \tag{1.6}$$

$$r_d = 1/\sigma_t^2 \tag{1.7}$$

$$r_t = 1/\sigma_t^2 \tag{1.8}$$

Combining cue through weighted averaging provides the most reliable and unbiased estimate of an environmental property from multiple cues, assuming that each cue is subject to statistically independent Gaussian noise and that the estimates from each cue are not too discrepant from one another. Central to MWF is the idea that the brain seeks to integrate information from cues in common units of depth through weighted averaging (Landy et al., 1995). Cues information is combined through promotion and the use of ancillary cues in order that cues provide estimates in common depth units for integration. Weighted averaging has been shown to provide a good model of sensory fusion in a number of circumstances (e.g. Ernst & Banks, 2002, Hillis, Ernst, Banks & Landy, 2002, Hillis et al., 2004), but it remains an open question as to whether cue information is processed in order to provide information in common units of *depth* for integration. Many sources of information cannot in themselves provide metric depth information (Koenderink, 1998) and others require certain statistical assumptions to be made before they can be used to estimate metric depth. The MWF combination scheme is a specific example of the more general Bayesian framework to sensory fusion (Knill & Richards, 1996, Landy et al., 1995), this is discussed in the next section.

1.2.2 Perception as Bayesian inference

Estimating the three dimensional structure of the world from the two dimensional information available on the retina, is coming to be seen as a process of Bayesian inference (Knill & Richards, 1996). The general form of this framework is shown in equation 1.9. Consider as an example a single cue that provides information about the slant, S, of a surface in the world. The sensory information available to the observer from this cue is represented by I. The likelihood function, $p(I|S)$, gives the probability with which the current sensory information could have been produced by each possible value of world slant, this embodies a model of how information from the world is transformed by the sensory apparatus to give an estimate of surface slant. As no sensory information is free of noise, there will be a distribution of possible slants consistent with any given cue. This is typically modelled as a Gaussian distribution with its peak at the true value of world slant (Figure 1.8), i.e. the cue is unbiased but corrupted by Gaussian noise.

$$p(S|I) \propto p(I|S)p(S) \tag{1.9}$$

The prior probability function $p(S)$ gives the probability of a given slant S occurring in the scene independent of the current sensory data, and is similarly modelled as a

Gaussian distribution centred on the most likely slant to occur in the world (Figure 1.8). The posterior distribution, $p(S|I)$, gives the probability of each possible value of world slant, S, given the current sensory information available to the observer, I. This is what the visual system seeks to estimate, the most likely state of the world given the current sensory information. The posterior probability distribution does not provide a single estimate of the measurement of interest, but a probability distribution of the likelihood of all possible values. Given this probability distribution the visual system needs to make some decision as to the current state of the world (Mamassian, Landy & Maloney, 2002).

A common rule used when modelling perception is to choose the peak of the posterior distribution, this is known as the Maximum a Posteriori (MAP) rule (Mamassian et al., 2002). The peak of the posterior distribution represents the most likely state of the world given the current sensory input and past sensory experience. The relative variances of the likelihood function and prior determine the position of the posterior distributions peak (Figure 1.8). Figure 1.8a shows the combination of the likelihood and the prior to produce the posterior probability distribution. If the variability of the likelihood is increased, as in Figure 1.8b, the peak of the posterior distribution, $\hat{S}_P$, is shifted toward that of the prior, $\hat{S}_{PR}$. The prior now has an increased weight in determining the shape of the posterior distribution, the system's inference about the world is shifted away from that defined by current sensory information toward that which is from experience most likely to occur in the world.

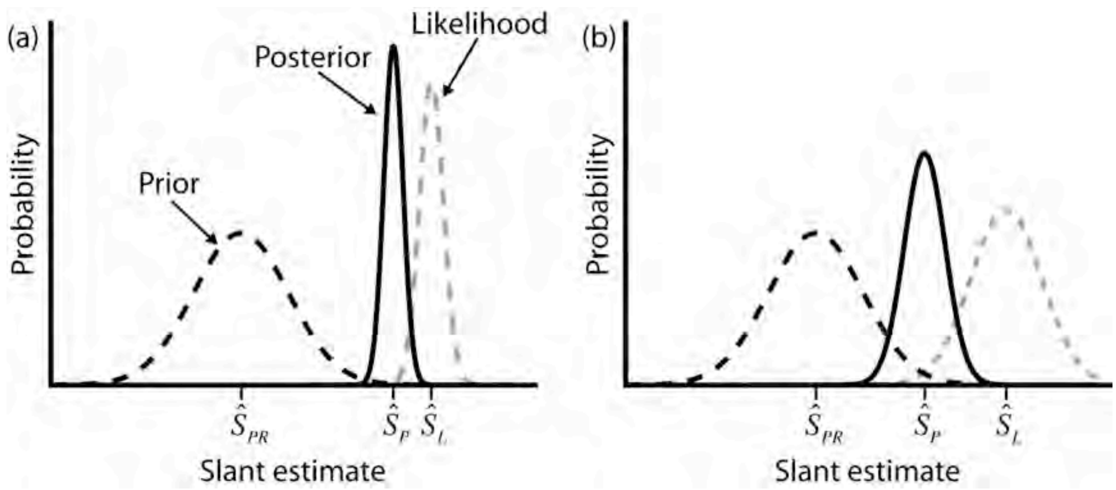

Figure 1.8: A schematic diagram showing how the prior probability distribution (black dashed line) and likelihood function (grey dashed line) combine to produce the posterior probability distribution (solid black line). (a) The likelihood, prior and posterior probability distribution each have an associated peak slant estimate, $\hat{S}_L$, $\hat{S}_{PR}$ and $\hat{S}_P$ respectively. As can be seen, the peak of the posterior probability distribution lies between that of the prior and likelihood. (b) The effect of increasing the variance of the likelihood function, whilst holding $\hat{S}_L$, $\hat{S}_{PR}$ and the variance of the prior constant, is to shift the peak of the posterior toward that of the prior. This demonstrates how the relative variances (i.e. reliabilities) of the likelihood and the prior determine the position of peak of the posterior distribution.

Conversely, if the variability of the likelihood is decreased it has more weight in determining the shape of the posterior distribution, and the peak of the posterior shifts toward that of the likelihood, $\hat{S}_L$. Similar, but opposite, effects occur if the variance of the prior is changed. It can therefore be seen that the relative variances of the likelihood and prior determine the variance of the posterior distribution and the position of its peak. When the prior is uniform i.e. no states of the world are *a priori* more likely than others, or when the prior has a much higher variance compared to that of the likelihood function, the influence of the prior is removed/greatly diminished and equation 1.9 can be simplified to equation 1.10. Now using the MAP rule on the posterior distribution becomes equivalent to choosing the peak of the likelihood function, this is termed maximum likelihood estimation (MLE) (Ernst & Banks, 2002, Mamassian et al., 2002).

$$p(S|I) \propto p(I|S)$$

(1.10)

1.2.3 *Bayesian cue combination*

The Bayesian framework can also be applied to the process of cue integration (Hillis et al., 2004, Knill & Richards, 1996), where multiple cues provide information about an environmental property, this is shown by equation 1.11. Consider again the task of estimating the slant of a surface in the world, but now with two cues providing information about surface slant (Hillis et al., 2004), I_A represents the sensory information available to the observer from cue A, and I_B the sensory information available to the observer from cue B. Again, the estimated slant from cues A and B have an associated variability and are represented by the Gaussian likelihood functions $p(I_A|S)$ and $p(I_B|S)$ respectively. Each likelihood function gives the probability with which the current sensory information from that cue was produced by each possible value of world slant. The visual system can obtain an estimate of slant using information from both these cues.

$$p(S|I_A, I_B) \propto p(I_A|S)p(I_B|S)p(S)$$

(1.11)

The prior probability function $p(S)$ is the same as in equation 1.9 for a single cue, and gives the probability of a given world slant, S, occurring in the scene independent of sensory data available from both cues (Hillis et al., 2004, Mamassian et al., 2002). The posterior distribution now becomes $p(S|I_A, I_B)$ and gives the probability of each possible value of world slant, S, given the sensory information available to the observer from cues A and B. As before, when no values of slant are *a priori* more likely than any others, or when the variance of the prior is much greater compared to that of the likelihood functions, equation 1.11 simplifies to equation 1.12. Using the MAP rule again becomes equivalent to MLE, but now through choosing the peak of the combined likelihood functions.

$$p(S|I_A, I_B) \propto p(I_A|S)p(I_B|S)$$

(1.12)

Combining information in this way is statistically optimal in that it provides the least variable estimate when information is integrated from multiple visual cues (Ernst & Banks, 2002). It also means that the system is more robust to cue degradation or dropout (Ernst, 2006, Ernst & Banks, 2002). As was discussed above, the likelihood, prior and posterior distributions are typically modelled as Gaussian probability distributions. When these are statistically independent and assuming a flat prior, integrating information through MLE is equivalent to averaging cues weighted by the inverse of their variance i.e. MWF (Ernst & Banks, 2002, Hillis et al., 2004, Landy et al., 1995). A number of studies have shown sensory signals to be combined in a way approaching statistical optimality, both within a single sensory modality (Hillis et al., 2002, Hillis et al., 2004) and between different sensory modalities (Alais & Burr, 2004, Ernst & Banks, 2002, Gepshtein & Banks, 2003).

Choosing the peak of the posterior distribution through MAP or MLE is an example of applying a decision rule to give a single estimate of the measurement of interest from the posterior distribution. A decision rule details the costs and benefits associated with making a particular estimate from the posterior. The MAP and MLE rules are examples of a broader class of rules called delta gain functions. With a delta gain function the greater benefit, or smaller loss, is gained when the sensory estimate made from the posterior falls within a window around the correct world value (Mamassian et al., 2002). The Bayes decision rule determines the estimate that maximises the expected gain from the posterior distribution given the particular costs and benefits related to a decision. Both the MAP and MLE rules are the Bayes decision rule for delta gain functions when the window around the gain is infinitely small (Mamassian et al., 2002). Evidence has shown that the visuomotor system is able to take account of the cost benefit landscape and motor variability associated with making rapid goal directed movements in order to maximise expected gain (Trommershauser, Maloney & Landy, 2003).

1.2.4 Weighted linear cue combination

Bayesian inference is neither an inherently weak or strong form of sensory fusion (Clark & Yuille, 1990, Landy et al., 1995), it makes no assumptions about how information in the world or sensory system is structured, such as whether or not information is partitioned into independent sensory cues. However, the majority of studies of Bayesian cue combination have applied Bayesian inference in a more modular fashion akin to MWF (e.g. Hillis et al., 2004). In this situation each cue is modelled as providing a statistically independent estimate of an environmental property in the form of a Gaussian likelihood function with its peak centred on the true world value. Using Gaussian distributions simplifies things as a Gaussian is fully characterised by its mean and variance, and allows the modelling of cue combination to be reduced to weighted averaging (Landy et al., 1995, Oruc et al., 2003). It is also empirically difficult to determine the true shape of the likelihood and prior (Clark & Yuille, 1990), although progress is being made in this area (Yang & Purves, 2003).

It might be reasonable to assume that cue estimates are statistically independent when the cues are from different sensory modalities, but when the cues are from the same sensory modality this assumption may no longer be reasonable. Similarly, cue estimates might not be well represented by Gaussian probability distributions in all circumstances. Oruc, Maloney and Landy (2003) investigated the extent to which deviations from these assumptions affected weighted linear cue combination. Using a slant estimation task they demonstrated that in a situation where two cues (texture and perspective) are likely to be correlated, observers could benefit from combining (non-Gaussian) cue estimates through weighted linear averaging. However, not all of their observers' data were consistent with optimal linear combination, which together with other research (Rosas et al., 2005) highlights that fact that weighted linear cue combination is one of a range of possible ways in which sensory information could be combined (Clark & Yuille, 1990, Howard & Rogers, 2002).

Demonstrations of optimal linear cue combination provide evidence in favour of largely independent estimates of a property, such as slant, being obtained from different visual cues. However, as Ernst and Bülthoff (2004) highlighted, this does not necessarily provide evidence in favour of brain modularity (see Zeki, 1993). Oruc et al. (2003) point out that for optimal linear combination to provide a good model, cue estimates must be largely statistically independent, but this does not mean the processing underlying each cues estimate could be localisable to a separable locus in the brain. Processing could be organised in the brain so that cues to object properties such as surface slant are de-correlated even though there are correlations present in their input signals (Barlow & Foldiak, 1989, Oruc et al., 2003). However, it is an interesting empirical question as to how cue combination will map onto processing in the brain (Welchman, Deubelius, Conrad, Bulthoff & Kourtzi, 2005).

In equations 1.9 and 1.11 a prior probability distribution was applied at a stage of cue processing where information was represented in the form of slant estimates. Although this is a reasonable stage at which to model the influence of prior information, it is also possible that prior information is used at different stages of sensory processing. For example, it has been suggested that the visual system could make use of prior information about the most likely distance of objects in the world during disparity scaling (Yang & Purves, 2003). Using the example of slant estimation from disparity, prior distance information would influence the estimation of slant from disparity before cue estimates are combined as shown in equations 1.9 and 1.11. It is also possible that the distribution of retinal disparity itself provides prior information about the viewing distance of objects or surfaces in the world (Glennerster, Rogers & Bradshaw, 1998). Prior information could therefore act at multiple levels for processing before the stage at which cues to a perceptual estimates of an attribute, such as slant, are combined. These issues are integral to understanding cue combination, the applicability of the MWF framework and the nature of visual cues themselves.

1.2.5 Bayesian inference and the veridicality of perception

Bayesian models of perception highlight the fact that perception is a process of statistical estimation about the most probable state of the world given current sensory information and past sensory experience. An interesting consequence of this is that visual processing need not result in geometrically correct estimates being made about the world. This is commonly termed veridicality. A large body of literature has sought to address the extent to which the 2-D information available at the eyes' retinas is sufficient to make veridical estimates of object attributes such as 3-D shape (for a review see Todd & Norman, 2003). However, this literature has to some extent made the implicit assumption that optimal use of information would result in perceptual veridicality. Recent evidence is suggesting that even with optimal use of information perceptual veridicality may not be achievable.

Weiss, Simoncelli and Adelson (2002) have demonstrated that a number motion illusions may be a direct consequence of combining noisy sensory data, which increases in variability at low stimulus contrasts, with a prior probability for slow motion speeds. Similarly, Fermuller and Malm (2004) have shown that by assuming Gaussian noise in local image intensity and position measurements, the extrapolation of local edges (edgels) from image intensity, the intersections of lines from edgels and local 2D image motion from image motion perpendicular to local edges, is subject to bias. This results in distortions in the perceived geometry of edgel-extrapolated lines, such that straight lines appear tilted, curved or displaced. This type of analysis can predict the distorted geometry of checkerboard type illusions such as the Café wall illusion and distortions in the perceived intersections of lines such as in the Poggendorff and Zöllner illusions.

A Bayesian model of 3-D motion perception in which noisy sensory data is combined with prior information has been shown to predict distortions in the perceived trajectory of an approaching object (Lages, 2006). Observers have been shown to overestimate the approach angle of a small disparity-defined target over a range of trajectories with the exception of when the target approached the observer along the mid-sagittal plane (Harris & Dean, 2003). This generalises to larger computer rendered wire frame spheres that provided additional looming information and small real world targets (a red LED) which controlled for conflicting accommodative and blur cues present with computer rendered objects (Welchman, Tuck & Harris, 2004). The model proposes that motion trajectory is estimated through the combination of noisy sensory information from binocular disparity detectors and with a prior for slow motion speeds (Lages, 2006). Overall, the above discussion indicates that instances of strikingly "incorrect" perception may be the consequence of largely optimal processing with noisy sensory data and/or prior information.

1.3 Perceiving three-dimensional shape from binocular visual cues

1.3.1 Distortions in perceived 3-D shape

Perception of 3-D shape from binocular visual cues in isolation has received much attention, as binocular cues offer the possibility of obtaining metric shape information. In an influential study Johnston (1991) had participants judge whether the shape of a disparity defined hemi-cylinder was circular i.e. that its depth was equal to half its height. This apparently circular cylinder (ACC) task was performed over a range of viewing distances between 53.5 and 214cm. With a task such as this the vergence state of the eyes provides one of the only means by which to estimate object distance in order to scale disparity for the estimation of 3-D shape. The results of the ACC task show that at near distances an apparently circular cylinder needed to be physically squashed in depth extent to be perceived as circular (depth less than half height), whereas at far distances an apparently circular cylinder needed to be physically stretched in depth extent to be perceived as circular (depth greater than half height).

From observers' responses it is possible to calculate the distance at which the disparities present in an apparently circular cylinder would geometrically specify a physically circular cylinder. The results of the ACC task are consistent with the overestimation of near distances and the underestimation of far distances from the vergence state of the eyes. For example, consider a physically circular disparity-defined cylinder located at a far distance, if its distance is underestimated and this incorrect distance estimate is used to scale disparity, the depth of the cylinder will also be underestimated. Therefore, in order for this cylinder to appear circular in shape its disparity-defined depth must be increased. Geometrically, this apparently circular cylinder now specifies a cylinder stretched in depth extent, but if interpreted at the incorrect nearer viewing distance this disparity specifies a circular cylinder.

This is a direct consequence of a given retinal disparity being consistent with different depths at different distances.

The interpretation of ACC results generally assumes that angular size is scaled correctly, which is highly unlikely. It could therefore be argued that the ACC task requires two distance estimates, one to scale disparity and one to scale angular size and that these may not necessarily be concordant with one another. Evidence does however suggest that in these types of task the same estimate of distance may be used to scale disparity and angular size (vanDamme & Brenner, 1997). In this case matters are simplified. Because depth from disparity d scales with the square of distance ($d \propto D^2$) whereas angular size α scales linearly with distance ($\alpha \propto D$), misestimating distance by a set amount will have a greater effect on depth estimates compared to size estimates. This will result in the same pattern of results as if angular size was scaled correctly, but the magnitude of the effect of distance on perceived shape will be less.

1.3.2 Vergence and a cue to object distance

It has been known since the time of Descartes that the vergence state of the eyes might be used as a cue to distance (Howard & Rogers, 2002), but there is still ongoing debate as to its utility for the visual system (Mon-Williams & Tresilian, 1999). It has been shown that when the occulomotor state of the eyes provides the only available information about distance, far distances tend to be underestimated and near distances tend to be overestimated, while isolated objects tend to be located at some intermediate 'abathic' distance (Foley, 1980). This has been termed the specific distance tendency (SDT) (Gogel, 1969). The SDT is evident during both monocular and binocular viewing of sparse reduced cue objects in the dark (Gogel, 1969, Gogel & Tietz, 1973) and has been suggested to underlie the poor disparity scaling seen in shape judgement tasks such as the ACC. Many studies have found qualitatively similar results to the SDT using both distance and shape estimation tasks (for reviews see Foley, 1980, and Todd, Tittle & Norman, 1995), but the inferred abathic distance between studies tends to vary and both over-estimation and under-estimation of distance are not always found (Viguier, Clement & Trotter, 2001) even within the original studies demonstrating the SDT (Gogel & Tietz, 1973). This is coupled with uncertainty as to the cause of the SDT.

It has been suggested that the SDT could be related to the resting vergence state the eyes adopt in the dark (Owens & Liebowitz, 1976). In the experimental setting, normally a dark reduced cue environment, the eyes may to some extent adopt this resting vergence state, which could act to bias estimates of distance from vergence. It is also possible that in the face of uncertainty in the reduced cue experimental setting the visual system's estimates of distance could be biased toward an intermediate value within an experimental distance range. This strategy would act to minimise average error, but would result in a contraction bias toward intermediate values consistent with the SDT (Mon-Williams & Tresilian, 1999, Poulton, 1989). Alternatively the visual system may embody prior information about the natural distribution of distances present in the environment (Yang & Purves, 2003) or information about the most likely distance to have produced a measured distribution of retinal disparities (Glennerster et al., 1998, Harris, 2004). Both could act to bias estimates of distance, which would act to bias the perceived 3-D shape of objects.

1.3.3 Natural scene statistics: disparity and the specific distance tendency

Typically psychophysical investigation has sought to understand visual processing by reducing the visual information available to the observer in order to gain precise experimental control over the visual cues being studied. This is a powerful technique, but the extent to which findings such as poor shape constancy (e.g. Johnston, 1991) are representative of processing with natural visual information is an open question. As a result work has begun to focus on understanding the statistical structure of natural scenes and how this could influence neural processing. Yang and Purves (2003) measured the distribution of distances in 74 scenes (23 fully natural scenes, and 51 scenes containing natural and constructed objects) using a laser range finding technique. From these data they were able to calculate measures of the distribution of distances and relative distances, within the scenes and subsets thereof.

They suggest that the obtained distance distributions can explain a large range of findings relating to the misperception of distance, including the specific distance tendency (Gogel, 1969, Gogel, 1977, Gogel & Tietz, 1973), the equidistance tendency (Gogel, 1965), distortions in localisation of objects on the ground plane (Ooi, Wu & He, 2001) and the effects of terrain on perceived distance (Sinai, Ooi & He, 1998). Applying a Bayesian model to distance perception, they propose that with impoverished visual information typical of many psychophysical experiments, the likelihood function(s) for distance will be flat due to the relative absence of sensory data. Because of this, distance judgements will be based primarily on the prior information about the natural distribution of distances present in the environment i.e. the distance distributions obtained from the laser range finding data. In the simplifying case of a single cue to distance, this means equation 1.9 discussed above reduces to equation 1.13. Distance estimates are based on the posterior distribution for distance, $p(S|I)$, but because of the relative absence of sensory data the influence of the likelihood is removed, or much reduced, this means that the shape of the posterior distribution is determined primarily by the distance prior, $p(S)$.

$$p(S|I) \propto p(S) \tag{1.13}$$

In terms of the specific distance tendency, Yang and Purves (2003) found that the probability distribution of radial distance from the scanner averaged across all scenes peaked at around 3m, this corresponds with the distance at which isolated objects tend to be localised in some demonstrations of the specific distance tendency. They suggest that observers have internalised these environmental statistics in the form of a prior for distance and base their distance judgments primarily on this information when faced with decreased or uncertain sensory information. They use similar analyses to account for the miss-localisation of objects on the ground plane, but with the introduction of a likelihood function for distance as these effects have been demonstrated in an open field setting where full cue visual information is available (Ooi et al., 2001).

A distance prior measurable from the environment would be an elegant way to explain the specific distance tendency and related findings, but there are a number of issues with the way in which its influence has been modelled. Firstly, the scanner was positioned at a minimum distance of 3m from objects within the scenes to prevent the majority of the camera's view of the scene being blocked. However, the value of 3m was also the peak in the probability distribution of radial distance from the scanner that was used as a probabilistic explanation for the specific distance tendency. The peak could therefore be a somewhat arbitrary consequence of methodologically defining this as the minimum measurable distance.

It is plausible to assume that during our natural interaction in the world we rarely position ourselves very close (< 20cm) to objects, but at the same time the majority of our visuomotor interaction and object fixation takes place at close distances (<1m). The experimentally inferred value of $\approx$3m is also too large to explain the results of tasks such as the ACC which require a prior to contract distance estimates around a value of $\approx$80cm (Johnston, 1991). This is also true of most measures of the SDT, which qualitatively but not quantitatively predict the results of shape judgement tasks such as the ACC.

In modelling the influence of the distance prior for the SDT it was also assumed that the likelihood function for distance would be flat. Whilst it is true that most cues to distance are absent in experimental demonstrations of the SDT, due to these experiments being carried out in the dark, vergence will be available as a cue to distance. Geometrically the utility of vergence as a distance cue will drop off as the distance of fixation increases. This is because as fixation distance increases the vergence angle decreases, meaning that any noise in its measurement will have a progressively deleterious effect on estimated distance. It is therefore reasonable to assume that the likelihood function for distance in experimental demonstrations of the SDT will be relatively flat at distances beyond $\approx$2m. However, at near distances vergence is a more reliable cue, so the assumption of a flat likelihood for distance will no longer hold. This suggests that any explanation of effects such as the ACC (Johnston, 1991) and the misestimation of distance (Viguier et al., 2001) will have to include some influence of sensory information from vergence.

1.3.4 *Combining disparity and motion cues to three-dimensional shape*

Demonstrations of poor perception of 3-D shape from binocular visual cues in isolation has prompted research to focus on how other cues can be used with disparity to better estimate shape (Harris, 2004). Theoretically the conjunction of binocular disparity and motion information in an object can be used to specify its geometrically correct 3-D shape and therefore its distance (Richards, 1985). This is possible as depth from disparity and depth from motion scale differently with distance. Both depth from motion and the angular size of an object scale linearly with the inverse of distance, this means that shape from motion is independent of distance up to a scaling for size. Depth from disparity on the other hand scales with the square of the inverse of distance so shape from disparity is dependent on distance (Landy & Brenner, 2001). Un-scaled motion and disparity information are therefore each consistent with a family of possible objects, but in conjunction there is only one distance at which both indicate the same object shape, this is the true object distance (Landy & Brenner, 2001, Richards, 1985). If the visual system could combine stereo and motion in this way, geometrically correct 3-D shape and distance could be recovered.

A large number of studies have investigated human perception of 3-D shape from stereo and motion cues. The evidence is equivocal, numerous studies have shown that observers are unable to determine geometrically correct 3-D shape from these cues alone (Norman, Todd, Perotti & Tittle, 1996, Tittle, Todd, Perotti & Norman, 1995, Todd & Norman, 2003, Todd et al., 1995). However, in some instances the addition of motion is seen to improve shape from stereo, although in most cases this is a partial improvement and does not result in geometrically correct perception (Bradshaw, Parton & Eagle, 1998, Brenner & Landy, 1999, Durgin, Proffitt, Olson & Reinke, 1995, Econopouly, 1996, Johnston, Cumming & Landy, 1994, Rogers & Collett, 1989, Tittle & Braunstein, 1993). Given that studies showing stereo motion combination are in the minority, it is possible that these instances represent the use of non-generic, context-specific strategies for recovering 3-D shape when stereo and motion cues are available (Todd, 2004, Todd & Norman, 2003).

1.3.5 Task dependent processing strategies

The brain is an evolved biological organ just like any other, but historically much work in perception has implicitly made the assumption that the visual system's attainable goal is to provide us with a geometrically correct real-time representation of the world. The stochastic nature of sensory information and success of Bayesian models of perception suggest that even with statistically optimal estimation this may not be obtainable. The visual system need only provide us with *sufficient* information to adaptively control behaviour, a geometrically correct real-time representation might be useful if it is attainable, but unless each adaptation toward that goal is beneficial for the evolutionary fitness of the organism it will not be selected for (Dawkins, 1986). An organism may also benefit from acting on incomplete or imprecise information quickly rather than waiting for complete precise information and acting too late (Brenner & Smeets, 2001).

Many task specific processing strategies could be exploited by the visual system to provide sufficient information to control behaviour adaptively whilst avoiding the need to recover precise geometrically correct information unless needed (Ramachandran, 1985). This information would allow for accurate responses to particular tasks without the need to have a single geometrically correct representation of 3-D shape (Garding, Porrill, Mayhew & Frisby, 1995, Glennerster, Rogers & Bradshaw, 1996, Todd, 2004). One such strategy exists for matching the shape of two disparity-defined objects at different distances. Evidence has shown that with disparity information alone observers are poor at tasks requiring geometrically correct 3-D shape information (Johnston, 1991). However, when matching the 3-D shape to two disparity-defined objects at different distances, geometrically correct performance can be attained if the ratio of the distances to the two objects is known. Shape constancy in this type of task is close to 100%, which suggests that observers are able to exploit this alternative processing strategy and avoid having to scale disparity to obtain metric shape information for completion of the task (Bradshaw, Parton & Glennerster, 2000, Glennerster et al., 1996).

Similarly, in order to judge that a disparity-defined surface is frontoparallel with respect to the observer the visual system need only gauge that the horizontal size ratio of the surface is equal to the square of its vertical size ratio, scaling disparity with an estimate of the viewing distance is not again required (Rogers & Bradshaw, 1995). The visual system might employ numerous task dependent strategies for the completion of specific tasks without the need to scale disparity. It is evolutionary more plausible that an organism act on sufficient information as and when it becomes available (Brenner & Smeets, 2001, Ramachandran, 1985), than construct a geometrically correct representation or model of the world (Marr, 1982). This might result in conflicts between decisions made with different sources of sensory information, or in different tasks, but it is also entirely plausible that the need for geometrically correct information may rarely occur (Bradshaw & Elliott, 2003, Bradshaw et al., 2000).

Bayesian models of perception and visuomotor control have ample scope to incorporate task dependent processing. Different tasks such as a perceptual judgement and a motor act may utilise different sources of information (Smeets, Brenner, de Grave & Cujpers, 2002) or differentially weight the same sources of information in their completion (Knill, 2005). This could result in different estimates of the same environmental property being inferred from different response tasks. The costs and benefits associated with making a perceptual judgement or motor response may also differ, which would alter the Bayes rule used in making an estimate from the posterior probability distribution (Mamassian et al., 2002). In this case the same information in the posterior distribution could be acted on differently depending on the costs and benefits associated with making different decisions or actions. Caution should therefore be taken in making direct comparisons between perceptual and visuomotor measures of the same environmental property, as it is unclear how information is being processed.

1.3.6 Comparing shape constancy with real and virtual stimuli

The majority of studies investigating how 3-D shape is perceived utilize computer-generated stimuli displayed on a computer monitor or projector. This is advantageous as it allows a high degree of control over the information used to specify an object and scene. However, there are important differences between viewing a real scene and a computer generated 3-D display of the same scene. Specifically, accommodation and retinal blur (focus cues) conflict with the information present in the simulated scene. This is because the light reaching the eyes from the simulated scene comes from a single flat surface, the computer monitor or projection screen, rather than the surface of the depicted objects (Watt, Akeley, Ernst & Banks, 2005a). At near distances screen pixelation may also be visible, which will introduce an additional cue to surface flatness (Watt et al., 2005a). It is therefore possible that the lack of depth constancy demonstrated in studies such as Johnston (1991) might be fully, or in part, due to conflicting screen cues rather than any inherent bias in the visual system.

Watt et al. (2005a) investigated whether conflicting focus cues from a computer monitor affect perceived 3-D shape in two ways; directly, through monitor slant influencing the perceived slant of a simulated surface displayed at the distance of the screen and indirectly, through focus cues from the screen providing a distance estimate that is used in disparity scaling. Focus cues had both direct and indirect effects on perceived 3-D shape. Monitor slant was found to directly influence the perceived slant of a simulated texture-defined surface, but not a disparity-defined surface. Screen cues to distance also indirectly affected disparity scaling in an ACC-like task where monitor position was fixed and objects were simulated at distances in front of and behind the screen distance. Importantly, full shape constancy was not achieved when real stimuli were used in the same ACC-like task and shape constancy with disparity-defined stimuli presented with the monitor screen at the simulated object distance was similar, though not identical, to that when real objects were viewed.

This suggests that the lack of shape constancy seen in tasks such as the ACC is not purely an artefact of the way in which computer generated stimuli are presented, but it raises important issues regarding the role of conflicting visual information. Monitor displays where information is presented on multiple focal planes, which achieve near correct focus cues, demonstrate both the deficiencies inherent in using conventional displays, and offer a possible future solution (Akeley, Watt, Girshick & Banks, 2004, Girshick, Akeley, Watt & Banks, 2004, Watt et al., 2005a, Watt, Akeley, Girshick &

Banks, 2005b). The visual system is robust and likely to exploit all available sources of information when estimating the 3-D structure of the world. It will therefore be difficult to make firm conclusions about how the visual system utilises information when conflicting or uncontrolled cues are present. However, the evidence suggests that presenting disparity-defined objects on a monitor screen positioned at the simulated object distance minimises cue conflict and produces similar, if not identical, results to using real stimuli.

Distortions of perceived distance, shape and size have been found using real-world stimuli and in the open field natural environment (Bradshaw et al., 2000, Cuijpers, Kappers & Koenderink, 2000, Koenderink, van Doorn, Kappers & Todd, 2002, Koenderink, van Doorn & Lappin, 2000, Loomis & Philbeck, 1999, Loomis, Philbeck & Zahorik, 2002, Wagner, 1985), however in most cases it is not possible to make direct comparisons to studies using computer-generated stimuli. In addition to the role of focus cues, studies in the natural environment differ in a number of respects from those in the laboratory setting. Observers in the natural environment are generally much less constrained and view stimuli presented at much larger distances, this means that both the availability and reliability of cues will differ. Depth intervals are underestimated relative to width intervals in both settings, at large and short viewing distances, but this is likely to be a product of using information differently in each case.

Bradshaw et al. (2000) demonstrated a comparable lack of shape constancy using sparse real-world stimuli over a similar range of distances to that used with monitor presented stimuli, but others have found stark differences when comparing real and simulated objects, such that perceived shape is closer to being geometrically correct when viewing real objects (Buckley & Frisby, 1993, Durgin et al., 1995, Frisby, Buckley & Duke, 1996, Frisby, Buckley & Horsman, 1995, van Ee, Banks & Backus, 1999). Both uncontrolled cues from the projection surface (Watt et al., 2005a) and additional cues present when using real objects will have a role to play in these differences by altering the way in which sensory information is combined or by allowing alternative processing strategies to be used. The important questions relate to understanding which sources of information present in the world are utilised in the control of behaviour. As long as the nature of the information available to the observer in the experimental setting is understood, reliable inferences can be made about the way in which information is used in more natural situations.

1.4 The visual control of prehension

1.4.1 Binocular disparity and prehension

Binocular disparity has been seen as especially important for the planning and control of prehension (Marotta & Goodale, 1998, Servos, Goodale & Jakobson, 1992), this is largely due to the fact that once suitably scaled binocular disparity can theoretically specify the geometrically correct structure of the environment. Geometrically correct information has been considered to be particularly valuable for the control of prehension, as errors in guiding the hand could result in missing or colliding with objects. Binocular information may be used in both the pre-movement phase of prehension, where the appropriate motor programs are selected for the reach and in the movement-execution phase of prehension, where information about the relative position of the object and hand are available (Bradshaw & Elliott, 2003). A combination of both feed-forward and feed-back processing would be a useful strategy for the visuomotor control of prehension.

Kinematic indices of prehension are closely related to object properties. Maximum wrist velocity, an early marker related to the ‘reach component’ of prehension, increases linearly with the distance of the object from the hand, whereas peak grip aperture, a later marker characterising the ‘grasp component’ of prehension, increases linearly with the grasped dimension of the object e.g. depth or width (Jeannerod, 1988). These indices have been used as surrogate measures of perceived distance and size respectively. Additional indices such as time in the deceleration phase of the movement, when the hand closes in on the object, and total movement duration, are taken as measures of the visuomotor system’s use of online feedback to adjust and correct the initial programming of the movement. For example, if the distance of an object were initially underestimated, the observer would need to use feedback to reprogram the movement in order to reach further to come into contact with the object. This would extend the length of the reach in its later stages and may introduce an additional peak in the wrists velocity profile. Timing indices such as these, time to peak grip aperture, and time to peak wrist velocity, are highly correlated with one another which reflects the temporal coupling of the reach and grasp components of the reach (Jeannerod, 1988).

Observers have been found to integrate multiple sources of information about the position of the hand continuously over the course of the movement (Saunders & Knill, 2003, Saunders & Knill, 2004, Saunders & Knill, 2005). Servos and Goodale (1994) investigated the contribution of binocular cues to the planning and ‘online’ control of prehension in natural full cue conditions. When the reach was carried out entirely monocularly or when an initial binocular view was replaced by monocular feedback, increased time was spent in the deceleration phase of the movement as the hand closed in on the object. Bradshaw and Elliott (2003), using self illuminated objects and hand markers, manipulated the proportion of the reach (0-100%) during which binocular feedback was available after an initial monocular view. Early kinematic indices such as maximum wrist velocity were unaffected by the addition of binocular feedback, whereas later indices such as peak gip aperture and time in the slow phase of the movement were affected. Peak grip aperture was smaller and less time was spent in the slow phase of the movement, when binocular feedback was available before 50% of the reach had been completed.

1.4.2 What is the importance of disparity for prehension?

These results are in accordance with binocular information being particularly important for the control of the grasp component of the reach (Watt & Bradshaw, 2000, Watt & Bradshaw, 2003) and the value of feedback from the hand relative to the object during the time after peak deceleration (Churchill, Hopkins, Ronnqvist & Vogt, 2000). The hand could be successfully guided to the object by ‘nulling’ the relative disparity between the hand and object (Bingham, Bradley, Bailey & Vinner, 2001, Bradshaw & Elliott, 2003, Hibbard & Bradshaw, 2003, Morgan, 1989), this would remove the need to scale disparity and recover geometrically correct 3D shape. Recent evidence has suggested that a disparity nulling strategy may only be useful during slow movements requiring high precision (Brenner & Smeets, 2006). This is characteristic of later stages in the reach where the velocity of the hand is slowed and the aperture of the digits adjusted, in order to grasp the object. Observers are particularly adept at using relative disparity information and are poor at scaling disparity to recover metric 3-D shape (Bradshaw et al., 2000, Glennerster et al., 1996, Johnston, 1991). Relative disparity thresholds are also much smaller compared to absolute disparity thresholds (Howard & Rogers, 2002). Disparity nulling could therefore be an advantageous strategy for the visuomotor system to use.

In addition to information from the hand and object, valuable cues to distance and shape are available from other objects and surfaces in the environment throughout the prehensile movement. These may also be differentially important for distinct components of the reach. During binocular viewing observers are seen to spend more time between peak deceleration and object contact and have larger grip apertures, when environmental cues are removed by turning the room lighting off but providing feedback of hand and object position with luminous material. These effects are further increased when feedback from the hand is also absent. In contrast, peak wrist velocity decreases with the absence of environmental cues, but is unaffected by the availability of feedback from the hand (Churchill et al., 2000). Others have found that during fully illuminated binocular viewing, occlusion of the table surface and feedback from the moving limb increased time spent in both the acceleration and deceleration phases of the movement, but left maximum wrist velocity and maximum grip aperture unaffected (Connolly & Goodale, 1999).

There is significant variability within the prehension literature as to the effects of removing binocular information from the hand, object or environment, in different stages of the prehensile movement (see Melmoth & Grant, 2006 for a discussion). This is coupled with subtle differences in how timing indices such as time in the deceleration phase are defined and contrasting methodologies for the provision of initial visual information and feedback. Broadly speaking, the general ‘advantage’ of binocular vision seems to be reflected in decreased grip apertures and decreased time in later portions of the movement when the hand comes into close proximity with the object to be grasped. This is consistent with a disparity nulling strategy (Bradshaw & Elliott, 2003) and conservative performance, whereby in the absence of binocular information grip apertures are increased and occur earlier in the reach, and more time is spent in later portions of the reach when the hand closes in on the object to be grasped (Watt & Bradshaw, 2000). This conservative strategy would provide an increased margin of error and more time in which to use online feedback to correct initial programming errors.

1.4.3 Summary: prehension and disparity

Some studies have found effects of the removal of binocular vision on earlier indices such as peak wrist velocity (Jackson, Jones, Newport & Pritchard, 1997, Servos, 2000, Servos et al., 1992) and time in the acceleration phase of the movement (Connolly & Goodale, 1999). Servos et al. (1992) found that monocular reaches exhibited slower peak wrist velocities and smaller peak grip apertures compared to binocular reaches, which was interpreted as subjects underestimating object distance and therefore object size (see also Servos, 2000). Jackson et al. (1997) also found lower peak wrist velocities under monocular viewing, but in contrast to Servos et al. (1992) *increased* peak grip apertures. In this case the effect of monocular viewing on peak wrist velocity was only present when objects were presented at a single viewing distance. When objects were presented at more than one viewing distance the effect on peak wrist velocity disappeared.

Given the methodological differences between those studies showing no effect of the removal of binocular cues on early indices such as peak wrist velocity and those studies showing an effect, the exact cause of these differences is unclear. However, it remains apparent that binocular cues are particularly important for the control of later stages of the reach when the hand closes in on the object to be grasped (Bradshaw & Elliott, 2003, Bradshaw, Elliott, Watt, Hibbard, Davies & Simpson, 2004, Churchill et al., 2000, Jackson et al., 1997, Servos & Goodale, 1994, Watt & Bradshaw, 2000, Watt & Bradshaw, 2003). The lack of consistent evidence for concurrent effects of the removal of binocular information on peak wrist velocity and peak grip aperture also argues against interpreting current data in terms of an altered distance estimate (indexed by peak wrist velocity) being used to scale object shape (as indexed by peak grip aperture).

It is somewhat surprising that binocular cues are found to be relatively unimportant in controlling the transport component of the reach as vergence can provide valuable distance information within reach space (Viguier et al., 2001) and manipulations of vergence have been shown to affect the transport component of the grasp (Gardner & Mon-Williams, 2001). It may be that other distance cues such as height in the field (Gardner & Mon-Williams, 2001) are more reliable than vergence and are thus weighted more heavily in the control of the transport component of prehension. This would negate any effect that the removal of binocular cues would have. It is therefore possible that the use of vergence information and disparity nulling do not represent useful control strategies for programming and control of the transport component of prehension.

1.5 Overview of subsequent chapters

Subsequent to this introductory chapter are four experimental chapters followed by a concluding chapter in which the research constituting the thesis is discussed as a whole. Chapter two investigates the consequences of incorrect disparity scaling for the shape constancy of objects moving in depth. Specifically, whether the visual system is able to maintain accurate shape constancy for disparity-defined objects moving in depth, by either combining information from disparity and motion, or by using a simpler heuristic strategy. In chapter three a simple model is presented that investigates the effect of noise in vergence, disparity and gaze direction signals in producing biases in the estimation of 3-D shape. The predictions of this model are compared to a comprehensive set of psychophysical data from observers performing the shape judgement task that is modelled.

The fourth chapter details an experiment that investigated how height in the field and a view of an objects upper surface and contour may be used in the estimation of 3-D shape. These cues are typically absent in experimental studies on the estimation of shape from disparity, but are present in natural tasks such as prehension. It is therefore possible that distortion in the perception of 3-D shape from stereo cues that has been demonstrated previously may be a consequence of artificially limiting the availability of visual cues. The final experimental chapter investigates similar issues to the rest of the thesis but in the context of prehension. Specifically, the chapter details an experiment investigating whether stereo and motion information is integrated from multiple objects in the scene for the control of prehensile movements. Chapter six, the concluding chapter, provides discussion of the research presented in this thesis as a whole. Appendices referred to throughout are provided at the end of the thesis.

Chapter Two

The shape constancy of disparity-defined objects moving in depth

2 The shape constancy of disparity-defined objects moving in depth

2.1 Introduction

2.1.1 The shape constancy of disparity defined objects

As was introduced in the in Chapter 1, despite the possibility that scaled horizontal disparity may provide accurate three-dimensional shape information, the perception of shape from binocular cues is often far from perfect. Johnston (1991) demonstrated this lack of shape constancy in an highly influential study. Observers were asked to judge the apparent three-dimensional shape of disparity-defined hemi-cylinders. In performing this task, observers were subject to systematic inaccuracies, which depended on the distance at which the hemi-cylinders were presented. At near viewing distances (less than $\approx$80cm) observers perceived a physically circular cylinder to be stretched in depth extent and at far viewing distances (greater than $\approx$80cm) observers perceived a physically circular cylinder to be squashed in depth extent (Figure 2.1a).

This pattern of results is consistent with observers overestimating near distances and underestimating far distances, and using these incorrect distance estimates to scale disparity. As a result, disparity-defined objects tend not to exhibit three-dimensional shape constancy. That is, the perceived three-dimensional shape of an object when presented at two different distances may differ, even when the physical shape of the object has not changed (Bradshaw et al., 2000, Glennerster et al., 1996, Johnston, 1991). This raises intriguing questions regarding shape constancy for objects that are moving in depth. One simple prediction is that an object of constant physical shape will appear to stretch in depth as it moves towards an observer, and to compress in depth as it moves away (Figure 2.1b). Conversely, in order to maintain perceptual shape constancy, an object would have to squash in depth as it moved towards the observer, and stretch in depth as it retreated (Figure 2.1c).

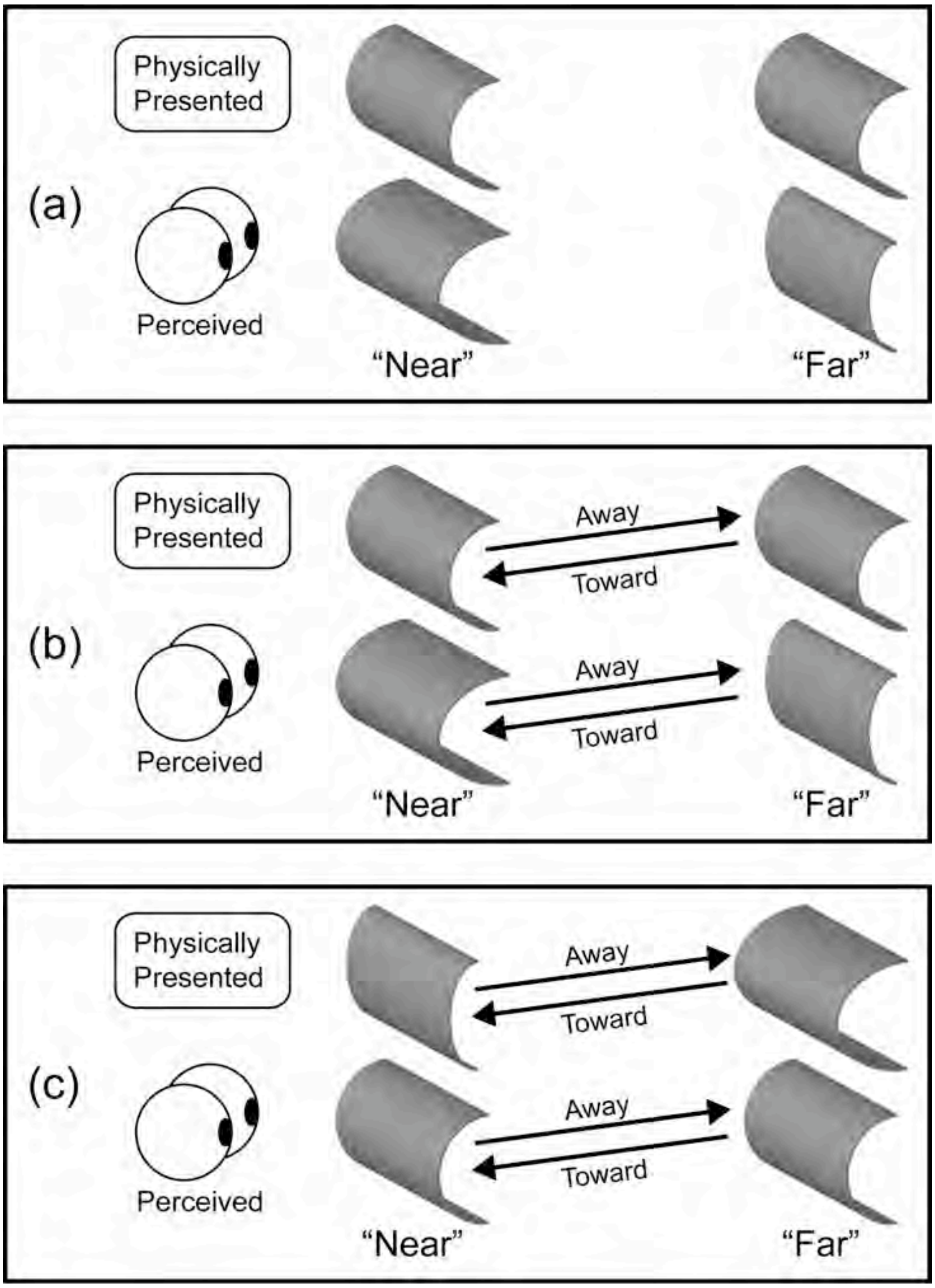

Figure 2.1: Schematic diagram of the ACC effect and the predictions this makes for the perceived shape of disparity-defined objects moving in depth. The standard ACC effect is shown in (a), the depth of a physically circular cylinder tends to be overestimated at near distances and underestimated at far distances relative to its height. If observers remain unable to scale disparity or combine stereo and motion information to determine 3-D shape, a physically constant cylinder would be predicted to 'stretch' in depth extent when moving towards the observer and 'squash' in depth extent when moving away (b). Conversely, for a cylinder to be perceived as constant in shape, it would need to contract in depth extent when moving toward the observer and expand in depth extent when moving away (c). For further details see the text.

While these predictions follow directly from the lack of shape constancy demonstrated by Johnston (1991) and others, there are at least two reasons for suspecting that such a result would not be observed, and that observers would maintain accurate shape constancy for objects moving in depth. Firstly, the retinal motion generated by the movement of an object in depth provides additional information as to its three-dimensional shape. As explained in Chapter one, due to the differential way in which shape-from-disparity and shape-from-motion scale with distance, the conjunction of stereo and motion information theoretically allows a geometrically correct interpretation of three-dimensional shape to be made (Richards, 1985). This is a form of cue promotion whereby interactions between cues allow missing parameters inherent to the cues to be specified in order to recover depth information (Landy et al., 1995). Although most studies on stereo motion promotion have focused on rotational object motion (e.g. Brenner & Landy, 1999, Johnston et al., 1994) or observer movement (e.g. Bradshaw et al., 2000), motion in depth could be used in a similar way to determine geometrically correct shape.

However, current evidence is equivocal as to whether the human visual system can promote stereo and motion cues in this way. In most instances observers are unable to recover geometrically correct three-dimensional shape from stereo and motion cues, and in those instances where improved perception of shape is seen with the addition of motion information, this is typically partial, and does not result in geometrically correct perception (for reviews see Landy & Brenner, 2001, Todd & Norman, 2003). It seems that in most circumstances observers are unable to carry out stereo motion cue promotion, which raises the possibility that those studies that have found improvements in perceived shape represent the use of alternative processing strategies. Alternative processing strategies could allow for both the accurate estimation of three-dimensional shape when stereo and motion cues are available, and for the completion of specific types of task without the need to scale disparity.

For example, when stereo and motion cues are available in a horizontally-orientated cylinder rotating round its vertical axis, the depth of the cylinder can be geometrically specified from two views of the cylinder's right-angle end-cut contour if one of these is fronto-parallel to the line of sight (Todd & Norman, 2003). This is a special case solution for completion of the task as it requires a specific view and rotation of an object for its application, and cannot be generalised to other object views or motions. Similarly, specific tasks, such as judging that a disparity defined surface is fronto-parallel with respect to the viewer (Rogers & Bradshaw, 1995) and matching the shape of two disparity defined objects at different distances (Glennerster et al., 1996), can be completed accurately without the need to scale disparity. Accurate completion of these tasks indicates that observers have a geometrically correct representation of 3-D shape, but that this representation is task dependent and is not generally available for other perceptual estimates (Glennerster et al., 1996).

It is generally assumed that the visual system would profit from scaling disparity, but many tasks can be completed without the need to do this (Garding et al., 1995). The brain may use many alternative processing strategies if these provide sufficient information for the adaptive control of behaviour (Brenner & Smeets, 2001, Ramachandran, 1985). A similar strategy can be identified for determining the rigidity or otherwise of stereoscopically viewed objects that are moving relative to the observer. For a given depth separation, binocular disparity is proportional to the inverse of the square of the viewing distance. Similarly, retinal image size is proportional to the inverse of the viewing distance. For an object moving directly towards an observer along the cyclopean line of sight, it can be shown that the rate of change of disparity, divided by the instantaneous disparity, will be twice the value of the rate of change of retinal image size, divided by the instantaneous retinal image size (Appendix A2). Calculation of this ratio would therefore provide a check as to whether an object maintains a constant shape as it moves through space.

2.1.2 Experimental predictions

The current study investigated the extent to which binocularly defined objects moving in depth exhibit shape constancy. Two possible results can be predicted. Firstly, a perceptually constant cylinder might be one whose shape modulates as it moves through depth, countering the effects of incomplete distance scaling as shown by Johnston (1991) and others (Figure 2.1c). Alternatively, a perceptually constant cylinder might be one that is physically constant, if observers are able to combine disparity and motion information to accurately recover depth, or to use some other processing strategy in order to ascertain the rigidity of the moving object.

The degree of shape constancy for simple disparity-defined objects was determined in three tasks. In the first, a standard apparently circular cylinder (ACC) task, observers were required to adjust the shape of a hemi-cylinder so that it appeared to have a circular depth profile. Appropriate scaling of horizontal disparity to take account of object distance is required in order to accurately complete this task. In the second, a matching task, observers were asked to adjust the shape of an object presented at one distance, so as to match the shape of another object at a different distance. Unlike the ACC task this task does not require the scaling of disparity to take account of viewing distance (Glennerster et al., 1996). Finally, observers were asked to judge whether a cylinder moving in depth was increasing or decreasing in depth extent. Again, accurate performance in this task can be obtained without the need for disparity scaling. Taken together, the results of these three experiments will allow us to determine whether observers are able to achieve shape constancy by circumventing the need for disparity scaling, either by making simple comparisons across two objects, or by taking account of the changing information available in a single moving object.

2.2 Methods

2.2.1 Stimuli

The stimuli were random dot stereograms of elliptical hemi-cylinders; the surfaces of the cylinders were rendered with 100 red Gaussian blobs, 0.5mm in diameter. Red was used as the shutter goggles have the minimum cross talk at longer wavelengths. The blobs size scaled appropriately with distance, which resulted in mean screen luminance increasing with decreasing distance. Blob size and screen luminance could therefore be used as a cue to object distance, but not shape. The stimuli were programmed using an assumed interocular distance of 6.5cm (Howard & Rogers, 2002), meaning that only subjects with this interocular distance would achieve geometrically correct depth perception. In all three tasks we are interested in the pattern of responses across distance, rather than assessing whether judgements were geometrically correct. A deviation from the programmed interocular distance would affect whether judgements were geometrically correct, rather than the pattern of set shape over distance.

Each hemi-cylinder had a fixed height of 3cm, and was presented at eye height with the principle axis aligned horizontally in the fronto-parallel plane. The ends of the cylinders were jittered with respect to the fixed length of 8cm; this was achieved by displacing the horizontal position of the end dots of the cylinder by an amount determined by the sum of two sinusoids that had frequencies of 4 and 6 cycles over the circumference of the cylinder. The amplitude of the two sinusoids was 8mm, and their phases varied randomly across trials. This, in conjunction with the relatively low dot-density used, ensured that the optical deformation of visible right-angle end cuts on the moving cylinders did not provide useful information about 3D shape (Todd & Norman, 2003).

2.2.2 Apparatus

Stimuli were viewed on a 21” monitor through Stereographics LCD shutter goggles synchronized to the 100Hz refresh rate of the screen. The left and right eyes each received a new image every 20ms. The monitor had a resolution of 800x600 pixels, and was viewed in a dark room at a fixed distance of 32.5 cm. Head movements were minimized with the use of a chin and headrest. Stimuli were presented at simulated distances of 40, 50, 60, and 70cm. This resulted in unavoidable cue conflicts with accommodation, which would indicate the true distance of the screen, whereas binocular disparity, motion and vergence were consistent with the simulated distance and shape of the cylinder. Cue conflicts were however reduced by using a restricted range of simulated distances. However given the problems inherent in using simulated versus real objects (Akeley et al., 2004, Buckley & Frisby, 1993, Durgin et al., 1995, Frisby et al., 1996, Frisby et al., 1995, Girshick et al., 2004, Watt et al., 2005a, Watt et al., 2005b) these conflicts must be kept in mind.

2.2.3 Participants

There were seven participants in total, two experienced psychophysical observers (the authors) and five observers who were naïve to the purpose of the experiment, and had limited familiarity with psychophysical research. All had normal or corrected to normal vision, and good stereopsis.

2.2.4 General procedure

Observers were allocated to complete either the ACC or matching task first at random. All however completed the motion task last, since data from the ACC task of each observer were needed to generate some of the stimuli in the motion task. To standardize stimulus presentation across the three tasks cylinders were presented along two diagonal paths, at a distance of 40, 50, 60 or 70 cm. These distances were chosen in relation to the experimental literature (and considerations of cue conflict) to cover a sufficient range to expect an effect of distance on perceived shape (van Damme & Brenner, 1997). At the 40 and 70 cm distances the cylinders were laterally separated from the mid-point of the screen by 8cm (either to the left or to the right), resulting in two diagonal depth profiles with an angle of 28º relative to the line of sight (Figure 2.2).

2.2.5 The apparently circular cylinder (ACC) task

The experimental task was to set the depth of the hemi-cylinder to be equal to half its height, i.e. to set the cylinder so that it appeared circular (Johnston, 1991). Shape adjustments were made on a computer keyboard, with one key increasing and another decreasing the cylinder's depth. The cylinder's depth could be adjusted between 0cm (a flat frontoparallel surface) and 9cm (six times the depth needed for a circular cylinder), this range had proved sufficient in a pilot study, and was reported to be sufficient here. A key press triggered the presentation of the first stimulus which had an initial depth taken at random from the full range. There was no time limit to adjustment. Once satisfied with their settings observers pressed a key once, which removed the cylinder from the screen, and again to trigger the presentation of the next stimulus. Fifteen settings were made at each distance, each presented at random, resulting in a total of sixty settings. The diagonal depth profile (left or right, see Figure 2.2a) on which the cylinder appeared on a given trial was also determined at random.

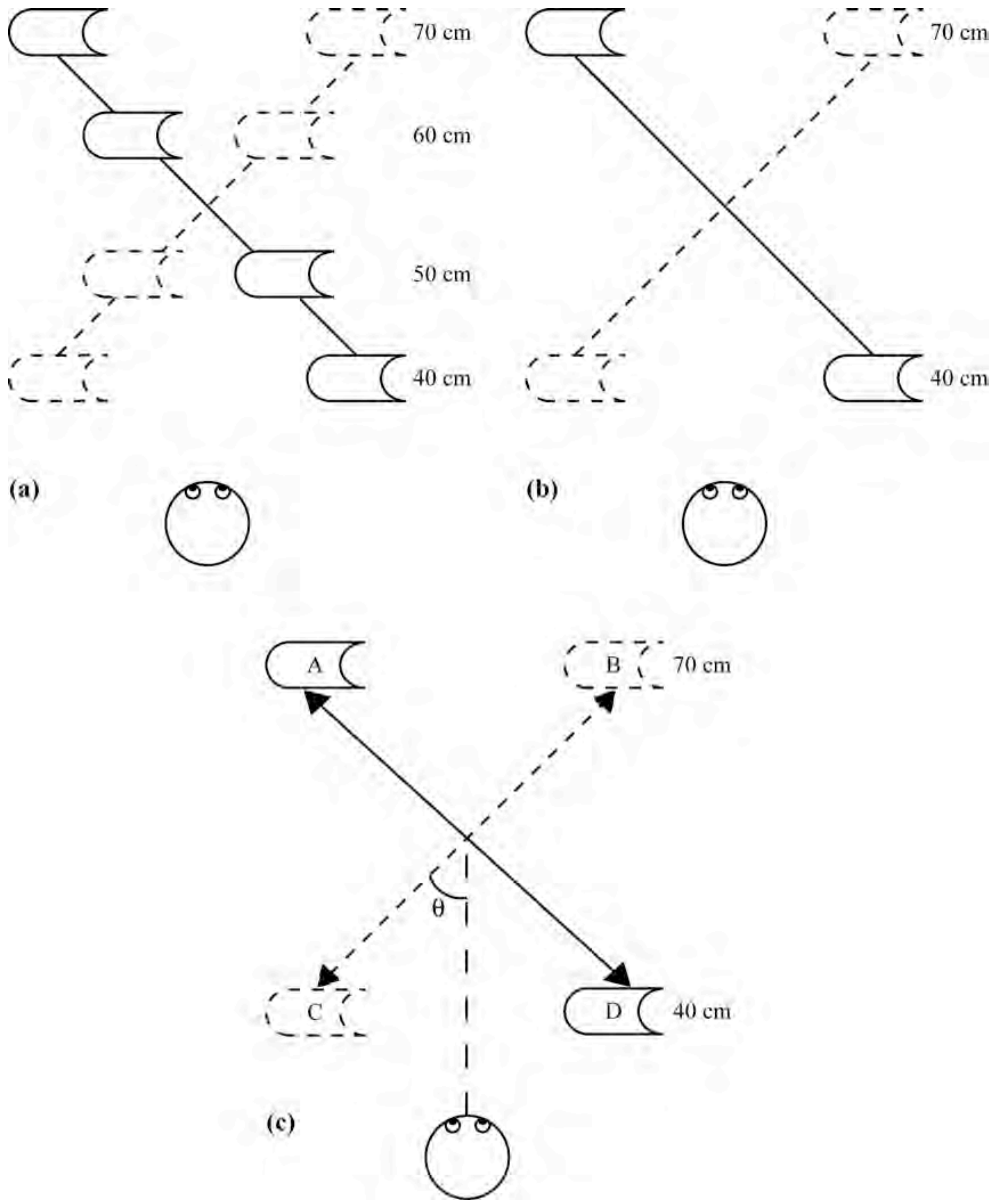

Figure 2.2: Schematic diagram of the stimuli for the three tasks seen from above, (a) the ACC task, (b) the matching task, and (c) the motion task. Stimulus presentation was standardised by presenting the stimuli along two diagonal depth profiles in each task. These

depth profiles (notionally labelled 'left' solid line/arrow, and 'right' dashed line/arrow) cut an angle of θ = 28° with respect to the observers' line of sight. For further details see the text.

2.2.6 The matching task

The matching task required observers to match the shape of two cylinders (the reference and the match) that appeared at the maximum and minimum distances used in the ACC task (70 and 40cm). This task is similar to the ACC task but does not require the recovery of metric shape information (Glennerster et al., 1996). Rather, accurate settings may be achieved if the ratio of the distances to the two objects can be determined (see Chapter 1 for details). An initial key press triggered the simultaneous presentation of two cylinders, each at eye level, laterally separated from the midpoint of the screen by ±8cm. The cylinder to be adjusted (the match) was indicated by a monocularly presented dot positioned 12.5° above it; the dot was monocular so as not to provide any additional depth information. The initial depth of the match cylinder was taken at random from the full adjustment range, which was the same as in the ACC task. The reference cylinder always specified a physically circular cylinder. Adjustments to the match cylinder were made using the computer keyboard. When satisfied with their setting, observers pressed a key once to remove the stimuli, and again to trigger the presentation of the next trial. The match was presented 15 times at each distance in a random order, with the match and reference alternating position (left or right, see Figure 2.2b) randomly over trials.

2.2.7 The motion task

In the motion task observers saw a cylinder translating diagonally toward or away from them along the diagonal depth profiles (Figure 2.2c). Two directions of motion (toward and away) were used for a number of reasons. If observers are unable to combine motion and stereo information, or use an alternative processing strategy to accurately access the 3D shape constancy of objects moving in depth, three direct predictions can be made from the results of Johnston (1991). Firstly, objects moving toward the observer should need to contract in depth extent to be perceived as perceptually constant in shape. Secondly, objects moving away from the observer should need to expand in depth extent to be perceived as perceptually constant in shape. Thirdly, if observers accurately track the moving objects and predominantly base their decisions on scaled object disparity at each point along the motion path the contraction needed with movement toward, and the expansion needed with movement away, should be isotropic.

The direction of motion, toward (start point 70cm) or away (start point 40cm), varied randomly as did the diagonal depth profile along which the cylinders moved (Figure 2.2c). There were therefore four possible combinations of start point and direction of motion, (1) the cylinder started at ‘A’ and moved toward the observer to ‘D’, (2) the cylinder started at ‘D’ and moved away from the observer to ‘A’, (3) the cylinder started at ‘B’ and moved toward the observer to ‘C’, or (4) the cylinder started at ‘C’ and moved away from the observer to ‘B’ (Figure 2.2c). During their movement, cylinders remained either physically circular in shape, expanded or contracted in depth extent along the line of sight, or modulated their depth extent in relation to individual observers’ settings in the ACC task. These latter stimuli were generated by performing a linear regression of set shape onto distance for each observer’s results in the ACC task. Thus at each point along the motion trajectory these cylinders assumed the shape of an apparently circular cylinder for that observer at that distance.

If perceived shape varies linearly with distance, and objects moving in depth exhibit the same degree of shape constancy as static objects, then it is predicted that these stimuli would appear constant in shape, regardless of their direction of motion. There were five levels of expansion or contraction; these were 0.25, 0.5, 1, 2 and 4. These values were chosen to incorporate the range of depth settings seen in a typical ACC task. When the cylinder’s depth extent changed during translation, it did so smoothly over the course of the whole movement, with the mid-motion depth always being equal to a circular cylinder. The exception to this was the ACC modulating cylinders. As the shape these cylinders took at each point in depth was based on a linear regression of each observers ACC settings, the shape these cylinders took did not necessarily include one that was circular, as a consequence the shape of the cylinder at the mid-motion depth did was not in general circular.

There were therefore six conditions in the motion task; four levels of expansion or contraction, the ACC modulated cylinders, and the physically constant cylinders. Each was presented moving toward and away twenty times (i.e. ten times along each depth profile), this resulted in 240 trials overall. The trials were presented in a random order, and split into two blocks of 120 to avoid observer fatigue. Observers triggered the first trial with a key press, a cylinder then appeared randomly at either the far or near distance, either to the left or to the right (Figure 2.2c). This remained stationary for one second before translating in depth at a speed of 17.8cm s^{-1}, either toward or away from the observer. At the end of the movement trajectory the cylinder disappeared. The observer's task was to decide if the cylinder they saw moving toward or away from them was expanding or contracting in depth extent. They registered their decision with a key press, which also triggered presentation of the next stimulus. Prior to starting the task the terms 'expanding' and 'contracting' in depth extent were explained clearly to observers using a cardboard model of the stimuli. This ensured that observers did not confuse the cylinders' changing depth extent with any perceived change in size during the movement.

2.3 Results

2.3.1 The ACC task

The mean set shape (depth to half height ratio) was calculated as a mean of the set shape at each of the two positions (one on each diagonal depth profile) for a given distance (Figure 2.3). The data indicate that across observers there was only a weak effect of distance on the set shape of an ACC, with marginally more depth being set in the ACC at far compared to near distances (Figure 2.3). A Friedman test on the mean set shape of an ACC at each distance across observers showed there to be a significant effect of distance on observer's settings, χ^2=10.71, df=3, p<0.01 (Monte Carlo sig, 99% confidence, 100,000 iterations). As can be seen from Figure 2.3 the magnitude of this effect is moderate. Given the known individual differences found between observers in shape judgment tasks such as these (Champion, Simmons & Mamassian, 2004, Glennerster et al., 1996, Oruc et al., 2003, Todd & Norman, 2003), individual observers' data were investigated. The individual ACC data show there to be large differences in the set shape of an ACC between some observers (Figure 2.4). In particular, only observers D and PS appear to show a clear effect of distance on ACC settings, which is likely therefore to be driving the overall significant result. This is supported by the fact that when these observers are removed the group effect of distance on ACC settings becomes non-significant χ^2=5.88, df=3, p=0.123 (Monte Carlo sig, 99% confidence, 100,000 iterations). This does however need to be tempered by the loss of statistical power due to removing two observers.

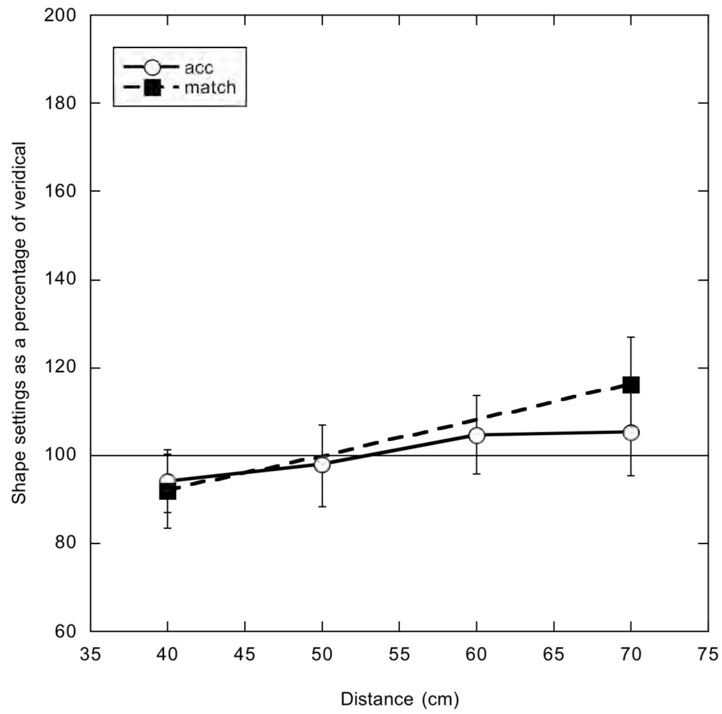

Figure 2.3: The overall ACC and matching results (error bars show 95% confidence intervals). The horizontal line indicates the depth needed for a physically circular cylinder. The scale is the same as Figure’s 2.4 and 2.5 for comparison.

2.3.2 The matching task

The matching task data showed a clearer effect of distance than the ACC task (Figure 2.3). Overall, when matching the shape of a cylinder across distance observers set significantly more depth in far compared to near cylinders (Wilcoxon z = 2.20, p=0.03). This overall result was reflected by individual observers with the exception of observers F and PBH (Figure 2.5). The clearer distance effect in the matching task was not however accompanied by reduced variability in observer's settings, as there was no significant difference between the standard deviation of shape settings in the matching and ACC tasks at 40cm, Wilcoxon z=1.18, p=0.30, or 70cm, Wilcoxon z=0.17, p=0.94 (these being the two shared distances between tasks). This contrasts with the subjective impressions of the majority of observers, who found this task much easier to complete. The matching task data are thus consistent with observers scaling the disparities they set in the match cylinder with an incorrect estimate of viewing distance in a similar manner to the standard ACC task (Johnston, 1991). It also suggests that in this instance observers did not use the alternative processing strategy to set the relative shapes of the two objects, which does not require distance scaling (Bradshaw et al., 2000, Glennerster et al., 1996).

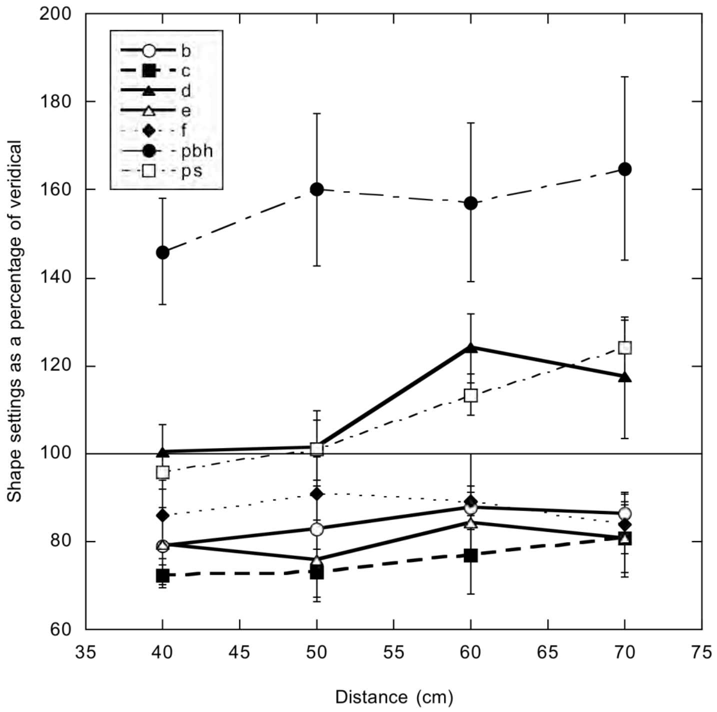

Figure 2.4: Individual observer's ACC data (error bars show 95% confidence intervals). The horizontal line indicates the depth needed for a physically circular cylinder.

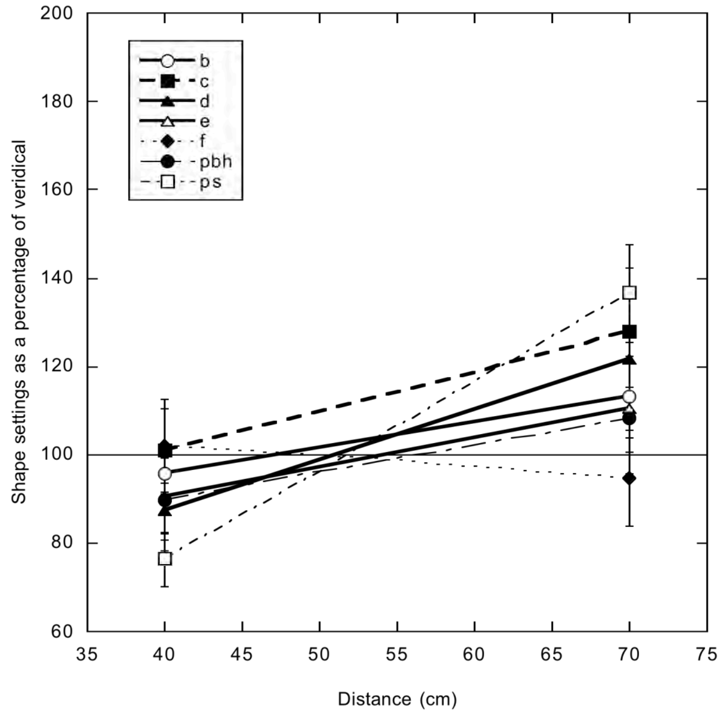

Figure 2.5: Individual observer's matching data (error bars show 95% confidence intervals). The horizontal line on the graph represents the depth needed for a physically circular cylinder. Symbols are the same as the ACC data in Figure 2.4 for comparison.

2.3.3 The motion task

Psychometric functions were fitted to the results using Probit analysis, and 95% confidence limits were calculated using a Bootstrap technique (Foster & Bischof, 1991). The 50% point of this fitted function represents the amount by which a cylinder needed to expand or contract when moving toward or away from the observer in order that the observer was equally likely to judge it to be expanding or contracting, i.e. to be perceived as physically constant in shape with movement in depth. Representative psychometric functions from observer E are shown in Figure 2.6, a clear separation of the toward and away curves can be seen. In order to have the appearance of a physically constant three-dimensional shape, cylinders moving toward the observer needed to contract in depth extent (PSE <1), whereas cylinders moving away from the observer needed to expand in depth extent (PSE >1). The mean PSE with movement toward the observer was 0.89, and away from the observer 1.35 (Figure 2.7), this difference was significant (Wilcoxon z=2.37, p=0.02).

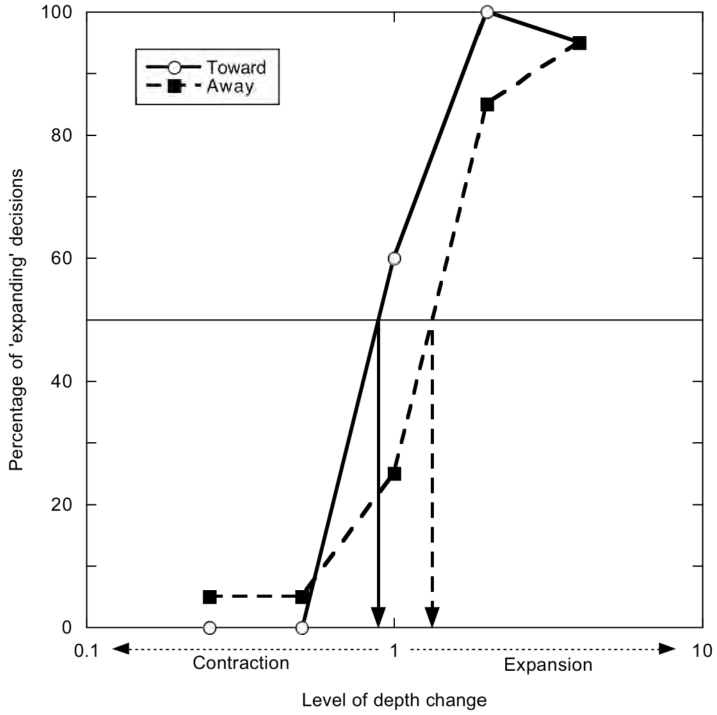

Figure 2.6: Representative psychometric functions from observer E. The point of subjective equality (PSE) in the motion task measures the level of expansion or contraction needed for half the observer's decisions to be 'expanding' and half 'contracting', i.e. the level of expansion or contraction in depth needed for a 'perceptually constant cylinder'. This is represented on the graph where the 'toward' and 'away' functions intersect the horizontal 50% line, and is indicated by the solid 'toward' and dashed 'away' arrows respectively. For this observer cylinders needed to expand when moving away (PSE>1) and contract when moving toward (PSE<1) to be perceived as physically constant in shape.

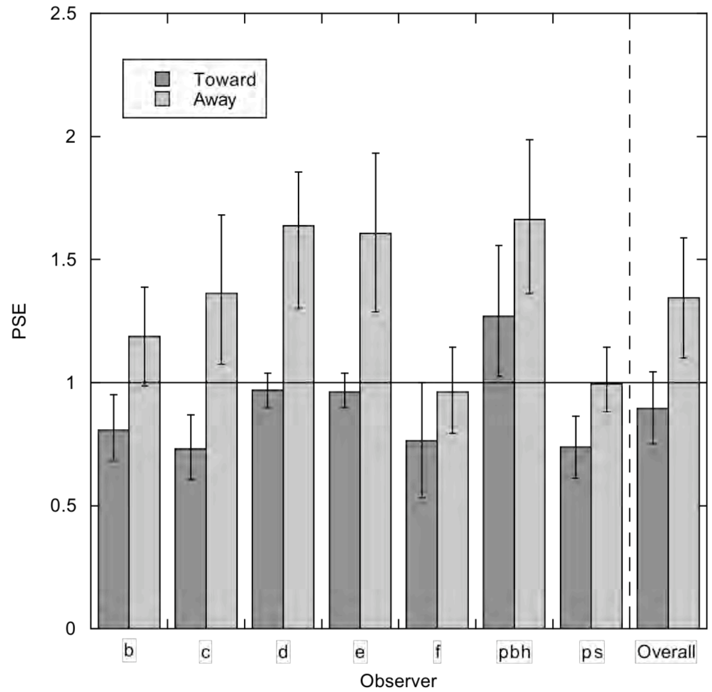

Figure 2.7: Individual observer's PSE's with movement toward and away in depth (error bars show 95% confidence intervals determined from the psychometric function fit). A PSE<1 indicates a cylinder needed to contract in depth extent to be perceived as constant in shape, whereas a PSE>1 indicates that a cylinder needed to expand in depth extent to be perceived as constant in shape.

Figure 2.7 shows that the significant effect of direction of motion on observers' responses was consistent across all observers bar two. Observer F showed no clear effect with both PSE's approximately equal to one, whereas the PSE's of observer PBH were both approximately the same but greater than one indicating that for this observer cylinders needed to expand in depth extent whichever direction they were moving. Looking at the physically constant cylinders in isolation supports the overall PSE result, 66% of physically constant cylinders moving toward the observer were seen as expanding in depth extent, compared to only 35% of cylinders moving away from the observer. This difference was statistically significant, Wilcoxon z=2.38, p=0.02.

Given that only two out of seven observers showed a significant effect of distance in the ACC task, and that this distance effect was weak, it was not meaningful to further explore the extent to which the ACC modulating cylinders were seen as remaining perceptually constant in depth extent. Interestingly, although there were large individual differences in the level of expansion and contraction needed with movement away from and towards the observer (Figure 2.7), there was no corresponding significant difference between PSE's[2], Wilcoxon z=0.68, p=0.58. This suggests that overall the level of expansion needed with movement away and contraction needed with movement toward was isotropic. It also suggests that asymmetries that have been observed between convergence and divergence responses (Hung, Zhu & Ciuffreda, 1997) did not affect observers' perceptions of shape. Hung et al. (1997) found that the initial fast component of convergence responses to a disparity step stimulus exhibited a greater gain to that found with divergence responses. This might have caused observers to think that stimuli moving towards them had moved a greater distance than those moving away from them, therefore necessitating a greater change in the depth of the stimulus when it moved toward the observer.

2.4 Discussion:

2.4.1 Comparing the tasks

The ACC task produced only a weak effect of distance in a limited number of observers. This is surprising as, although the distance range was restricted so as to minimize cue conflicts, a similar distance range has produced clear effects in previous studies (vanDamme & Brenner, 1997) and was sufficient to produce clear distance effects in both the matching and motion tasks presented here. It may be that the use of a handheld reference (vanDamme & Brenner, 1997) acts to reduce the variability of shape settings in ACC type tasks resulting in a clearer effect of distance. Alternatively, the effect of distance on ACC settings may have been less clear here as cylinders were presented along two diagonal depth paths. Along these paths disparity

[2] PSE's for movement towards were compared with 1/PSE's for movement away so as to compare the degree of expansion/compression appropriately.

and vergence thresholds may vary with eccentricity, which might be expected to systematically affect the perception of shape from stereopsis (see Chapter 3). In contrast, the matching task showed a robust distance effect, with observers tending to underestimate the far and overestimate the near distances, consistent with incorrect distance scaling. Observers also failed to use the alternative processing strategy by which to match the shape of the two cylinders (Bradshaw et al., 2000, Glennerster et al., 1996).

The failure of observers to use this strategy may be due to a number of factors. To set the ratio of disparities accurately, so as to match the shapes of the cylinders, some estimate of the ratio of the viewing distances must be obtained. This might come, for example, from the ratio of the heights of the stimuli, if it is assumed that they are the same size (Glennerster et al., 1996). In the Glennerster et al. (1996) study, the match and the reference were viewed in dim lighting conditions on separate monitors placed on a black and white textured linoleum surface. From this information, the ratio of distances could more readily be estimated from the angle subtended by the monitors, or other distance cues such as the texture gradient.

One possibility is that the situation in the matching task presented here was different as the stimuli were viewed in darkness. This may have made it much less appropriate for the visual system to assume size constancy, and estimate the ratio of stimulus distances in order to match the cylinder’s shape. The matching strategy, as with the ACC task itself, will be affected by the available distance information (Glennerster et al., 1998) and in this instance the distance information available may not have been sufficient. Contrary to this, Bradshaw et al., (2000) have demonstrated that observers can accurately match shape when viewing very sparse stimuli (triangles defined by three isolated LED’s) in darkness. The absence of additional cues to size and distance cannot therefore fully explain why observers did not accurately match shape here.

It does however seem that stimuli are not always matched in instances where the strategy could be used (Todd & Norman, 2003), which together with the present data points to the fact that the strategies used may be more context dependent than was originally thought. One possibility is that the pattern of eye movements required to fixate the two objects may affect matching performance. Although evidence suggests that observers can make relatively good distance judgments across changes in vergence (although see Brenner & Landy, 1999, Brenner & van Damme, 1998) this may not be the case for changes in version. The Bradshaw et al., (2000) study required observers to make relatively small (<1°) vertical version movements to fixate the stimuli whereas the present study and others (Todd & Norman, 2003) have required much larger horizontal version movements (≈18° in the present study). This difference may have affected whether the matching strategy was used.

Interestingly, even though observers failed to use the matching strategy this did not result in all observers making similar shape settings in the ACC and matching tasks. The results were mixed; some observers did make similar settings in the two tasks (e.g. Observers D and F), whereas others did not (e.g. Observers C and PBH). If observers perceived the reference correctly during the matching task, accurate performance in both the ACC and the matching task would result in the same settings being made, however this is unlikely to be the case. Observers were clearly unable to use the matching strategy here, so both the match and the reference will have been subject to incorrect distance scaling. This would have acted to magnify the effects of distance mis-scaling in the matching task compared to ACC task, and could explain why a robust distance affect was found in the matching task, but only a weak effect in the ACC task.

Consider the case where observers are presented with a physically circular 'reference' cylinder at a far distance, and have to match its shape in an adjustable 'match' cylinder at a near distance. Firstly, the reference will be perceived as squashed in depth relative to its physically circular shape. The observer's task is then to match this *perceived* shape in the near cylinder. However, any settings made in this near cylinder will also be misperceived, such that depth is overestimated. The set *physical* shape of the match will therefore be further squashed relative to the *perceived* shape of the (*physically circular*) reference. The converse applies when the 'reference' is near and the 'match' is far. Therefore, when the matching strategy cannot be used, the matching task will produce greater measured mis-scaling compared to the ACC task because in doing the matching task observers are mis-scaling twice.

Different observers may also have processed the two tasks differently, or adopted different alternative strategies in their completion. As Todd (2004) has suggested, the strategies that observers adopt to mediate their responses in the face of ambiguous information do not necessarily remain constant over different response tasks. Individual differences are a consistent feature of many shape judgment tasks, and to this extent the present data are no exception. These differences are an important and consistent feature of this area of research, and thus need to be explained in order to fully understand the processing that underlies observers performance (Glennerster et al., 1996, Hibbard, Bradshaw, Langley & Rogers, 2002, Todd, 2004, Todd & Norman, 2003).

2.4.2 Stereo-motion cue promotion

Despite the availability of sufficient information from the combination of disparity and motion cues (Richards, 1985) observers failed to accurately identify objects that exhibited physical shape constancy as they moved in depth. To be perceived as remaining constant in shape, cylinders needed to expand in depth extent when moving away, and contract in depth extent when moving towards the observer. A perceptually constant cylinder was thus one that modulated its shape in a way that countered the effects of scaling disparity with incorrect estimates of viewing distance (Johnston, 1991). Overall the expansion and contraction appeared to be isotropic suggesting that observers accurately tracked the moving objects, and based their decisions on (mis)scaled object disparity at each point in the course of the movement. These results are further evidence that stereo motion promotion is not carried out by the human observer (see Landy & Brenner, 2001, Todd & Norman, 2003 for reviews)

Observers were also unable to compare the changing disparity of a moving cylinder with its changing image size to determine if it was remaining rigid in shape. There have been reports in the literature that the visual system may be poor at utilizing information regarding the rate of change of disparity (Harris & Watamaniuk, 1995). However, in the current study the dot lifetimes were not limited, as in the Harris & Watamaniuk (1995) study, the stimuli therefore contained consistent interocular velocity differences that could have been detected by mechanisms tuned directly to this property (Brooks & Stone, 2004), and used in the same way to assess whether an object was expanding or contracting in depth extent.

Retinal size, as with disparity, needs scaling by an estimate of viewing distance. It seems likely that observers would have been scaling both angular size and disparity incorrectly, possibly by using the same distance estimate for both types of scaling (vanDamme & Brenner, 1997). This would cause physically constant cylinders moving toward the observer to expand in perceived size, and physically constant cylinders moving away from the observer to decrease in perceived size. This would also occur if observers failed to scale angular size at all, and based their decisions directly on the angle subtended by the stimuli on the retina. However, both these interpretations cannot provide an alternative explanation for the results. Firstly, as was described in Chapter 1, if angular size and disparity are scaled with the same mis-estimate of distance, this will have a greater effect on perceived depth compared to perceived size. The main component to the misestimated shape would be the misestimated depth.

Therefore, the same pattern of results for the perceived shape of the moving cylinders would be expected but of a lesser magnitude, as is true for the ACC task itself. Secondly, observers were all well briefed on the task before commencing; this included a demonstration of the depth and width of a cardboard model of the cylinder stimuli, as well as an explanation of the task. Furthermore, all cylinders across all tasks were the same size and scaled with distance appropriately, but a clear separation in the psychometric functions was found. Therefore although size is subject to distance mis-scaling too, this would have been consistent across all the stimuli, and could not account for the differential pattern of results found with different levels of cylinder expansion or contraction.

Overall the motion and matching tasks show a greater level of consistency with one another than with the ACC task. It can be seen by comparing Figures 2.5 and 2.7 that those individuals showing an effect of distance in the motion task also show an effect in the matching task. This is surprising as the ACC and motion tasks should be more consistent given that observers are likely to have been mis-scaling twice in the matching task. The majority of observers commented on how much “easier” the matching and motion tasks were compared to the ACC task, which is consistent with previous research suggesting that the ACC is an intuitively difficult task for observers to perform (Todd & Norman, 2003). Judging the shape of cylinders may be difficult because fusion cannot occur where the disparity gradient exceeds a critical value of ≈1 (Burt & Julesz, 1980), toward the horizontal edges of the stimuli. It may be that observers find this task difficult but have increased confidence in making relative shape judgments in the matching and motion tasks, this could affect either the way observers combine the available cues, or the response strategies and heuristics they adopt (Todd, 2004), and may account for the comparative subjective difficulty of the ACC task.

2.4.3 Summary: perceiving a stable world

In summary, the results provide a striking demonstration of the perceptual consequences of the failure to obtain shape constancy. Although sufficient information for the recovery of geometrically correct 3-D shape was available to the observer, physically constant disparity defined cylinders are perceived to expand in depth extent when moving towards the observer and to contract in depth when moving away (Figure 2.1b). Conversely, in order to be perceived as constant in shape, cylinders needed to modulate their shape to counter the effects of incorrect distance scaling, by contracting in depth extent when moving toward the observer and expanding in depth extent when moving away (Figure 2.1c, 2.7). This is consistent with recent research demonstrating that observers are unable to exploit information from disparity and motion in an immersive virtual reality environment to determine whether objects are moving relative to them as they move (Tcheang, Gilson & Glennerster, 2005), or that the room in which they are walking is expanding dramatically (Glennerster, Tcheang, Gilson, Fitzgibbon & Parker, 2006b).

Tcheang et al. (2005) got observers to move laterally back and forth ±1m relative to isolated rendered football viewed at 1.5m in an otherwise blank environment. The task was to judge whether the football remained static or rotated with or against their lateral movement. The football rotated about its vertical axis yoked to the observer's lateral movement with a gain between -1 and +1. A gain of +1 meant that the football rotated with the observer's movement so that the same surface of the ball always faced the observer, a negative gain meant that the football rotated against the observer's movement and a gain of zero that the ball remained stationary. In nearly all instances a football needed to have a positive gain, such that it rotated with the observers movement to be judged as static. This bias was greatly reduced by presenting other static footballs in the room or by giving the room lattice grid walls and a floor (Tcheang et al., 2005)

An even more dramatic demonstration of the visual system's inability to perceive a stable environment was found by manipulating the size of a virtual room as an observer moved within it (Glennerster et al., 2006b). Observers stood to the left of a virtual room with brick rendered walls and a checkerboard floor. While standing in this position the observer viewed a 'standard' cube positioned at 0.75m, they then walked horizontally across the room to view a 'comparison' cube. The task was to judge whether the comparison cube was larger or smaller than the standard. Each cube was only viewable when the observer was in zones to the left and the right of the room, and the room was otherwise empty. This meant that at no point in time were observers able to see both cubes at once, or any other objects. During the observer's horizontal movement across the room, whilst neither cube was visible, the room smoothly expanded in size by a factor of four around the cyclopean point mid way between the eyes, this included the floor and wall textures scaling. Subjects failed to notice this expansion, and for the comparison cube to be judged as the same size as the standard cube, it too had to expand in size. The expansion needed in the comparison cube varied from around a factor 2 when the comparison cube was viewed at 1.5m, to nearly 4 when viewed at 6m (Glennerster et al., 2006b).

The visual information available from the texture of the floor and walls of the room was ambiguous as to the room's size throughout the experiment, whereas stereo, motion and height in the field indicated the changing size of the room. Texture information could however be used to judge the sizes of the standard and comparison cubes. This is because observers could judge the size of the standard and comparison cubes *relative* to the texture of the walls and floor in the two zones where each cube was visible. There were therefore two broad categories of information that could be used by observers in judging the size of the two cubes. These are (1) *absolute* *information* about distance and shape from stereo, motion and height in the field, and (2) *relative information* from the comparison of the cubes size relative to the wall and floor textures (Glennerster et al., 2006b).

The authors demonstrate that a model with a single free parameter, which represents the relative weighting given to *absolute information* compared to *relative information*, predicts observer's size matching performance well. In this model an increase in viewing distance results in an increase in the weight given relative information about cube size, and a decrease in the weight given absolute information about cube size, as would be predicted from geometry. The fact that observers fail to notice the changing size of the room is likely to be due to the low weighting given stereo, motion and height in the field, compared to texture, given the room's dimensions. The weighting of these cues may also be reduced as they conflicted with the observer's proprioceptive information about the position of their body in the room.

Proprioceptive information suggests that the observer is standing on the ground but stereo, motion and height in the field specify that the room has expanded four fold, and that the observer is now floating above the room's floor. Texture on the other hand is ambiguous as to the size of the room, so does not conflict with preoperception. Observers in the study were to some extent aware that the room was changing in size as they felt the length of their strides to be increasing and decreasing as they moved across the room (Glennerster et al., 2006b). This suggests that observers had access to the conflicting preoperceptive and visual information (see also Ernst, 2006, Hillis et al., 2002).

The results of the current experiment complement those from Tcheang et al. (2005) and Glennerster et al. (2006b) by demonstrating that within a distance range where vergence information should be maximally reliable, observers are unable to scale disparity with an estimate of the true viewing distance to correctly estimate the shape constancy of objects moving in depth. These studies serve to demonstrate that biases in perceived shape can come about both directly, through the transformation inherent in the use of information from a single visual cue, and more indirectly, through the way in which different sources of information are integrated to form a single percept (Knill & Richards, 1996, Mamassian et al., 2002).

It is justifiable to ask why, given well-documented biases in perception (Todd & Norman, 2003), even in the natural environment (Wagner, 1985), we perceive the world to be a relatively stable place. Much will depend on the type and quality of the visual cues available (Knill & Richards, 1996). However, it may be that the visual system embodies a strong assumption toward world stability (Glennerster et al., 2006b, Koenderink, 1998). The visual system may simply not need to estimate the absolute metric structure of the environment during our everyday interactions (Bradshaw et al., 2000, Garding et al., 1995). Consequently, the visual system would routinely avoid any conflict or distortion inherent in making decisions about metric structure by using alternative processing strategies. The experimental situation might force the visual system to make an inaccurate decision regarding metric structure, but this may bear little cost in everyday life.

Chapter Three

Visual noise as a determinant of systematic distortions in the perception of distance and shape from binocular visual cues

3 Visual noise as a determinant of systematic distortions in the perception of shape from binocular visual cues

3.1 Introduction

3.1.1 Distortions in the perception of three-dimensional shape from stereopsis: vergence as a cue to object distance

As was discussed in the introduction, the perception of 3-D shape from binocular visual cues in isolation is subject to systematic distortion. Stereoscopically defined objects presented at far viewing distances tend to be seen as squashed in depth extent, and stereoscopically defined objects presented at near viewing distances stretched in depth extent. As a consequence of this perceptual distortion an observer sets more depth than is needed when matching depth intervals to width intervals at far distances and less depth than is needed at near distances (Johnston, 1991). Geometrically, horizontal retinal disparity needs scaling with an estimate of the viewing distance in order for absolute depth to be estimated. With binocular visual cues in isolation both vergence and vertical disparity could provide information about viewing distance that could be used to scale disparity for the estimation of 3-D shape.

However, practically in most experimental situations vergence will provide the predominant source of distance information because vertical disparities are only exploited by the visual system for scaling disparity when the stimuli are sufficiently large (20 degrees or more) (Bradshaw et al., 1996, Rogers & Bradshaw, 1993). In the majority of studies investigating the perception of 3-D shape from binocular cues the stimuli are typically too small to produce sufficient gradients of vertical disparities to be of use. For example, a horizontally orientated cylinder 10cm in length positioned at a distance of 40cm subtends an angle of 14°; this is below that required to provide a sufficient gradient of vertical disparity for the visual system to use (Cumming, Johnston & Parker, 1991, Rogers & Bradshaw, 1993).

In 3-D shape judgement tasks such as the ACC, where an observer sets the shape of a disparity-defined cylinder to be circular at a range of viewing distances it is possible to calculate the distance at which an apparently circular cylinder would need to be viewed in order to physically project disparities at the retina consistent with those of a circular cylinder, this is termed the scaling distance (Johnston, 1991). In calculating the scaling distance it is assumed that all error in set 3-D shape is due to misestimating object distance (see also Foley, 1980). Scaling distances are typically found to be larger than the rendered object distance for near distances and smaller than the rendered object distance at far distances, with a cross over point at approximately 80cm (Glennerster et al., 1996, Johnston, 1991).

This is a qualitatively similar pattern of results to that inferred from distance estimation tasks where vergence is the predominant distance cue, near distances tend to be overestimated and far distances underestimated (e.g. Gogel & Tietz, 1973). Because of this similarity a link has been made between the misestimation of distance from vergence and the misestimation of shape from disparity (Foley, 1980, Johnston, 1991). The geometry of vergence suggests that it should be a more reliable cue for near distances. This is because the angle of convergence becomes progressively smaller as the distance of fixation increases, with approximately 90% of the vergence range being used for distances up to 1m (Collewijn & Erkelens, 1990). Any angular error in vergence will therefore have a proportionally greater effect at far compared to near distances, for this reason vergence is thought to be of little practical use for the estimation of distance beyond approximately 2m (Howard & Rogers, 2002, Mon-Williams & Tresilian, 1999). It is therefore surprising that distance and shape estimation tasks indicate that near distances to be overestimated when vergence is the predominant distance cue.

This chapter will start by reviewing the literature relating to the estimation of distance from vergence and discuss current explanations as to why distance might be systematically misperceived when vergence is the predominant or only distance cue. A model of how distance and shape might be estimated from noisy vergence, disparity and gaze angles signals will then be presented. Data from this model will then be examined in relation to a comprehensive set of measurements of the distribution and bias of 3-D shape settings made in a simple shape judgement task.

3.1.2 Estimating distance from the vergence state of the eyes

Gogel (1969) suggested that in the face of ambiguous information about target distance, as is typical of most psychophysical studies, the visual system's estimate of distance might shift toward some default value, this was termed the 'specific distance tendency' (SDT). In this framework, as the information available to an observer from which to estimate distance is decreased, estimates shift toward a specific or intermediate distance value (Gogel & Tietz, 1973). As a consequence of this near distances are overestimated and far distances are underestimated. The distance around which estimates are thought to contract was considered to be stable for a given observer in a given viewing situation, but variable over different observers and for the same observer in different viewing environments, where the availability and quality of visual cues may vary.

Gogel and Tietz (1973) used an illusory motion parallax task and verbal distance estimation to investigate the SDT. In the illusory motion parallax task observers slid their head laterally side to side whilst fixating a point of light presented at a range of distances (30cm to 8.83m) in an otherwise dark environment. In this viewing situation illusory motion of the point of light is predicted if the observers accurately fixate and track the light but misestimate its distance. If the light's distance is underestimated it will be perceived as moving laterally with the movement of the head, whereas if its distance is overestimated it will be perceived as moving laterally against the movement of the head. The light will only appear as stationary if its estimated distance is the same as the distance of fixation. Two groups completed the motion parallax and distance estimation tasks, one group with monocular vision and one group with binocular vision.

In the motion parallax task observers more frequently perceived motion of the light against movement of the head at nearer distances and motion of the light with movement of the head at further distances. The distance at which the frequency of observers' responses to movement of the light were equally 'with' and 'against' their head movement was interpreted as the distance at which perceived distance was equal to the distance of object fixation. This was approximately 3m for monocular viewing and 5m for binocular viewing. If the rationale of the task is accepted, the results suggest that near distances are overestimated and far distances underestimated, but differentially so for monocular and binocular viewing. It is also worth noting that the inferred abathic distance differs greatly to the 80cm abathic distance inferred in 3-D shape judgement tasks (Glennerster et al., 1996, Johnston, 1991)

In contrast to the motion parallax task, verbal estimates of distance provide mixed evidence for the SDT. Gogel and Tietz (1973) report both un-calibrated and calibrated verbal distance estimates, the un-calibrated distance estimates are the raw distance estimates reported by the observer, whereas the calibrated distance estimates have been corrected by a function fitted to observers' verbal estimates of distance and physical distance, as determined under sparse fully illuminated binocular viewing. There are important differences between the two measures so both are discussed. The authors base their conclusions on median data as the distributions of estimated distance are skewed. For monocular viewing the calibrated distance estimates indicate that distances under $\approx$2m are overestimated and distances over $\approx$2m are underestimated. The uncalibrated distance estimates indicate that distances greater than $\approx$2m are underestimated, but the overestimation of near distances is inconsistent.

For binocular viewing the calibrated distance estimates indicate that distances over $\approx$2m are underestimated, with a small overestimation of nearer distances. In contrast, the uncalibrated distance estimates indicate that distances >1m are underestimated but distances <1m are estimated accurately. The assumption underlying the calibration procedure is that distances are veridically perceived in the 'full cue' calibration condition and that any errors present in this condition result from converting an accurately perceived distance into a verbal distance estimate (Gogel, 1969, Gogel & Tietz, 1973). This is analogous to the observer having an incorrect criterion for the unit of measurement, for example if an observer underestimated the length of a metre all distance estimates reported in metres would be underestimated by a fixed proportion.

Although the criterion an observer has in mind for a particular unit of measurement is an interesting issue, comparison of the calibrated and non-calibrated data show that the use of the calibration function acts to add additional bias into observers' distance estimates (e.g. Gogel, 1969, Gogel & Tietz, 1973). It is a strong assumption to suggest that all error in the calibration condition is due to observers having an incorrect criterion for the unit of measurement used for verbal report. Bias might be introduced at a number of stages during processing and could in part be due to the nature of the viewing geometry itself. Occam's razor would suggest that as the use of the calibration procedure acts to add additional bias into observer's distance estimates; the use of uncalibrated distance estimates would be the more appropriate measure of perceived distance.

It is also unclear whether the assumptions underlying interpretation of the data from the motion parallax task in terms of pure distance misestimation are valid. To be interpreted in terms of misestimated distance alone, observers would need to have accurately fixated the target and accurately tracked its position whilst making lateral head movements. They would also need to have accurate knowledge of the distance their head had moved. Illusory movement of the point of light would be predicted if any of these assumptions were not met, as was recognised by Gogel and Tietz (1973). This makes it difficult to assume that the task is a direct measure of misestimated viewing distance. The results of these tasks are therefore unclear as to biases that may exist in the estimation of distance from vergence information.

In a distance estimation task Viguier et al. (2001) got observers to set the distance of an LED cursor to be equal to, half that of, or double that of, a reference LED. The reference was placed at a range of distances along the median plane (between 20-80cm, depending on the task) and the cursor was positioned along a line parallel and 4cm to the right of the median plane. Both were presented at eye height and across all tasks the reference LED subtended a constant angular size and had equal luminance across the distances tested, whereas the cursor's angular size and luminance varied appropriately with distance. Both the cursor and reference were viewable, meaning that relative disparity information was available for setting the cursor's distance. In a fourth condition equal distance settings were made but without a concurrently seen reference. In this condition the reference was presented alone for five seconds and then turned off for five seconds, before the cursor was presented for the distance setting. This dark interval was used to allow vergence to adopt a resting state between the presentation of the reference and cursor to eliminate any relative disparity information. Finally, in a fifth condition, the reference was presented alone and oral distance estimates were made.

With concurrent viewing of the cursor and reference equal distance settings were highly accurate across the whole distance range, reflecting the availability of relative disparity information. However, half distance settings were overestimated at near distances and underestimated at far distances, and all double distance settings were underestimated. Verbal responses underestimated all reference distances and this tendency increased as the reference distance increased. The authors interpreted the underestimation of distances in the double, half and verbal judgment tasks as indicating that observers used an egocentric reference of $\approx$10cm from themselves from which to judge distances. This reference corresponds to the maximum possible vergence angle (Collewijn & Erkelens, 1990). This fits the data but it is unclear why the visual system would adopt such a strategy in the face of inconsistent feedback during tasks such as prehension and conflicting information from other visual cues. Furthermore, with more than one target in the scene clear biases in the estimated distance of the target object would be expected from previous data (Foley, 1985).

Importantly with sequential presentation of the cursor and reference distance settings were accurate up to 40cm but progressively underestimated at 60 and 80cm. Accurate estimation of near distances is what would be expected from the geometry of vergence. Eye movements were tracked for four subjects during a 3 second fixation of the cursor and reference LED's. A clear inversely proportional relationship between the vergence angle and fixation distance was seen as would be predicted from the geometry of vergence in the experimental task. This suggests that observers were accurately fixating the target across the distance range used. In general the precision of distance estimates decreased as the reference distance increased across all tasks. Accurate perception of near distances when vergence is the predominant distance cue has been reported in a number of other studies studies.

Swenson (1932) got observers to manually estimate the perceived distance of a binocularly viewed target along the median plane with a movable pointer positioned below the stimuli. Target luminance and angular size were kept constant. At each of the three near distances tested (25, 30 and 40cm) manual distance estimates were highly accurate with mean errors <2mm. Mon-Williams and Tresilian (1999) got observers manually to judge the distance to a binocularly viewed single point light source in an otherwise dark environment. The light source was presented at a fixed presentation distance directly in line with the right eye. Vergence specified distance between 20 and 60cm was manipulated with ophthalmic prisms placed in front of the left eye.

A regression of perceived distance on vergence specified distance suggested that observers tended to overestimate near distances and underestimate far distances. This interpretation of the results can however be questioned. Both base in and out prisms were used to manipulate vergence specified distance, this means that the (fixed) distance of the point light source was within the vergence specified distance range. Vergence specified distance therefore conflicted with distance information from other visual cues, such as accommodation. In a cue conflict situation such as this estimated distance has been shown to be a compromise between that signalled by the conflicting visual cues (Watt et al., 2005a), as would be predicted from models of cue combination (Landy et al., 1995). Cue conflicts could therefore explain the apparent contraction in distance estimates.

Tresilian, Mon-Williams and Kelly (1999) demonstrated similar results in a study that used prisms to manipulate vergence specified distance within a viewing box. Distance estimation was investigated in a range of viewing conditions designed to assess the role of additional visual cues in the surroundings (rich cue, reduced cue and darkness), changing angular size of the target object, and monocular versus binocular viewing. Targets were large rectangles of black card during viewing in the light and luminous tubing when viewed in the dark, with the exception of one dark condition in which a point light source was used. All targets were viewed in line with the right eye, which allowed vergence specified distance to be manipulated with prisms in front of the left eye in some conditions. Objects were placed between 25/30cm and 100cm (depending on condition) and distance was manually estimated using an indicator stick on the outside of the viewing box.

A strong linear relationship between perceived and physical distance was found across all conditions. In addition, base in prisms were shown to increase perceived distance and base out prisms to decrease perceived distance, as would be expected by the prismatic viewing geometry. The effect of the prisms increased as the available visual cues decreased, which would be predicted on the basis that vergence information now conflicted with fewer visual cues. This cue conflict was evident in the point light condition, in which the target's physical distance was fixed at 50cm, but vergence specified distance was manipulated using prisms. Here overestimation of near distances and underestimation of far distances was found. Discussion will focus on the conditions where no prisms were used, as these contained no cue conflicts.

In the rich cue condition with angular size varying appropriately with target distance, estimated distance was highly accurate. When angular size was removed as a cue to distance or targets were viewed in the reduced cue environment there was a general tendency for further distances to be underestimated, although estimated distance remained reasonably accurate under most cue conditions. When viewing luminous tube targets in the dark with angular size as a cue, distance estimates were more variable compared to the lit conditions and distances up to $\approx$80cm were slightly overestimated. When these stimuli were viewed monocularly in the dark and changing angular size was the only cue to object distance, far distances were underestimated and near distances were overestimated but to a smaller amount. These data suggest that in natural viewing conditions vergence can provide a valuable source of distance information.

Foley and Held (1972) got observers to manually estimate the position of (a) a point light source viewed in the dark and (b) short plastic rods viewed in the light, with an unseen hand. The plastic rods were viewed in the light on a gridline surface positioned 3cm below eye level and two context 'tack' objects were present, so multiple cues such as perspective, accommodation and relative size were available. The point light source was produced by viewing two laterally separated lights, presented at a fixed distance of 75cm, 1.5cm below eye level, through polarising filters in front of each light and eye. The point light sources were presented at eye level and appeared at the same height as the tops of the rods resulting in vergence being the predominant cue to distance. Targets appeared at distances up to 36cm along either the median plane or a plane running 16.2 degrees to the right or left of the median plane.

In both viewing conditions, along each viewing plane, observers tended to overestimate target distance; this overestimation was much greater (up to 25cm) for the point light source viewed in the dark. As before, the overestimation of the point light source's distance is likely to be related to conflicts between vergence and other visual cues. It is less clear why observers consistently overestimated distance by up to $\approx$11cm in the multi-cue condition where there were no cue conflicts. Manual estimation of distance has been found to be highly accurate for distances up to 50cm in other studies using a rich cues consistent environment (Mon-Williams & Tresilian, 1999, Tresilian et al., 1999).

3.1.3 Making links between the misestimation of distance and the misestimation of 3-D shape

For distance misestimation to provide an explanation of the results of 3-D shape judgement tasks the scaling distances calculated in shape judgement tasks should quantatively correspond with patterns of distance misestimation, this is generally not the case. Scaling distance calculated from 3-D shape judgement tasks are indicative of near distances (<80cm) being overestimated and far distances (>80cm) being underestimated (Glennerster et al., 1996, Glennerster et al., 1998, Johnston, 1991). In contrast, the distance perception literature suggests that near distances are estimated reasonably well when vergence is the predominant source of distance information, but that far distances are underestimated (Tresilian et al., 1999, Viguier et al., 2001). The results from shape and distance estimation tasks are therefore more consistent with one another at far distances compared to near distances.

Geometrically, vergence would be expected to be a more reliable cue for near compared to far distances (Collewijn & Erkelens, 1990, Mon-Williams & Tresilian, 1999) and to a large extent the distance perception literature in this area confirms these predictions. This suggests that the misestimation of 3-D shape at near distances may be due to some other reason than the misestimation of viewing distance. There are however large differences between studies, and between individual observers within studies, as to the perception of distance and shape they gain from experimental stimuli. Accurate perception of near distances is not always found (Foley & Held, 1972) and similarly over- and under-estimation of distance is not always inferred from 3-D shape judgement tasks (Todd & Norman, 2003). This complicates matters when trying to make links between distance and shape estimation tasks. The nature of the task the observer is performing is likely to have a large influence on the accuracy of distance perception (Howard & Rogers, 2002).

For example, when judging the distance of an object, its perceived size or an observer's prior expectations about its size as presented over different distances, may influence distance estimates (Collewijn & Erkelens, 1990). Differences between studies may also be attributable to the mode of response (e.g. Viguier et al., 2001), for example manual estimation of distance is arguably less prone to cognitive influences compared to verbal report. Observers might also adopt various heuristic strategies when making judgements about attributes such as shape, which need not be consistent across different response tasks (Ramachandran, 1985, Todd, 2004). Overall, current experimental data suggests that misperceptions of the shape of disparity-defined objects cannot be solely explained by the misestimation of object distance, as is assumed in the calculation of scaling distances (Johnston, 1991). However, it is clear that the misestimation of distance plays an important role in the misestimation of 3-D shape, especially at far viewing distances.

3.1.4 Are systematic distortions of perceived shape and distance examples of a contraction bias?

It has been suggested that the overestimation of near distances and underestimation of far distances, typically described in terms of a SDT, may simply represent a 'contraction bias' whereby in conditions of uncertainty observers responses regress to the centre of an available response range, which has the result of reducing average error (Mon-Williams & Tresilian, 1999). More specifically this is what Poulton (1989) would term a 'response contraction bias'. The response contraction bias occurs when observers have a limited range of responses available with a known centre or range, under these conditions they tend to make responses which are biased toward the centre of the available range (Poulton, 1989). In this interpretation the systematic distortions demonstrated in distance and depth perception tasks should not be taken as a failure of vergence to accurately determine object distance. Instead distance might be accurately estimated from vergence and used to scale horizontal disparity for the estimation of depth, but observer's responses regarding distance or depth might be biased to the middle of the available response range (Mon-Williams & Tresilian, 1999, Mon-Williams, Tresilian & Roberts, 2000).

In a direct distance estimation task it is clear to see how a contraction bias could operate to produce results consistent with the SDT if an observer has an estimate of the response range. However, this interpretation is inconsistent with studies that have shown near distances to be estimated reasonably accurately (e.g. Viguier et al., 2001). It is less clear how the results of 3-D shape judgement tasks could be explained by a contraction bias. The results of shape judgement tasks such as the ACC are consistent with a contraction bias in *distance estimates*. Contraction biases act on the unit of response (Poulton, 1989), in the case of shape judgement tasks this would cause observers to bias their estimates and decisions toward an intermediate shape within the range[3]. Unless observers were presented with a range of shapes centred on a theoretically expected bias in perceived shape at each distance within the range, it is difficult to see how responses in a shape judgment task could result in a pattern of results consistent with a contraction bias in *distance estimates*.

It is also hard to reconcile a contraction bias explanation with the notion that observers are completing a task on the basis of perceived distance or shape. A contraction bias suggests that observers effectively disregard perceived distance and shape in favour of biasing their judgements to the middle of the available range. This is not representative of observer's performance in experimental tasks, or observer's subjective reports of how they complete these tasks (Figure 3.1). Many shape judgement tasks are also blocked by distance in order to reduced the effects of screen cues to distance/depth (Watt et al., 2005a), 'blocking' distance estimates is known to

[3] Here "shape" could refer to un-scaled retinal disparity, depth as compared to width/height or a higher-level estimate of shape such as curvature, depending on how you wish to interpret an observer's completion of the task.

eliminate the effects of the contraction bias (Mon-Williams & Tresilian, 1999, Poulton, 1989).

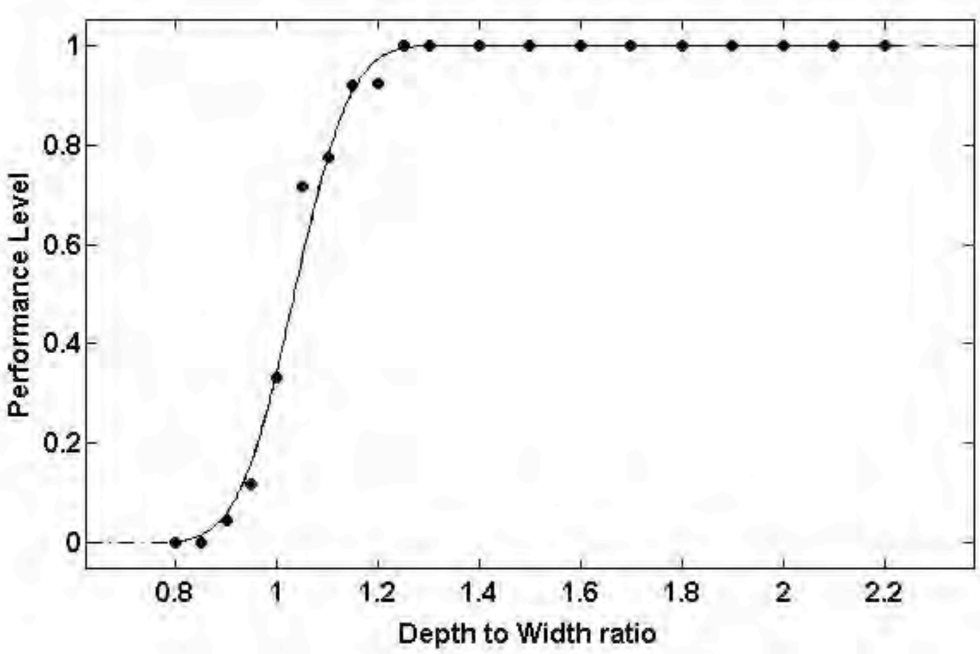

Figure 3.1: Example psychometric function for an observer performing the apparently circular cylinder task in the 'Lidded' condition of the experiment reported in Chapter 4. In brief, stimuli were presented on four interleaved adaptive staircases, with a total of 262 trials. The depth to width ratio of the stimulus is shown on the abscissa and the proportion of 'stretched in depth' responses on the ordinate. The 0.5 performance level (PSE) indicates the depth to width ratio needed for the observer to perceive the stimuli as circular. The range of the stimuli in this instance was not 'centred' around the observers PSE. It is immediately clear that observers are not simply biasing their responses to the middle of the available response range, despite remaining subjectively unsure of the accuracy of their responses. The value of the PSE was a depth to width ratio of 1.03 (upper 95% confidence interval 1.06, lower 95% confidence interval 1.02) and by comparison the centre of the stimulus range was 1.5. This suggests that observers are performing these types of task on the basis of perceived shape.

A further prediction the contraction bias makes is that as the available cues to distance increase the contraction bias should decrease, because more cues are available from which to estimate distance (Tresilian et al., 1999). However, this is also true of explanations in terms of the traditional specific distance tendency (Gogel, 1969, Gogel & Tietz, 1973). It is possible that similar results to a contraction bias could be produced if observers built up an estimate of the average object distance during the course of the experiment, and that this acted to bias perceived distance and shape. This would be somewhat similar to the action of a distance prior (Yang & Purves, 2003), but one learnt over the course of the experiment. The role of prior information about distance is discussed in a later section. For now it seems reasonable to assume that part of the bias in distance and shape estimates may be due to the way in which distance is estimated from sensory information about the vergence state of the eyes.

3.1.5 Uncertainty about the convergence state of the eyes

It was recognised by Helmholtz, that uncertainty as to the convergence state of the eyes could lead to misperceptions about the position and shape of objects in space.

> "Owing to the uncertainty of our judgements as to the degree of convergence of the eyes, we are liable to have illusions also about the forms of things in space as seen binocularly. The interpretation of the visual phenomena would be correct if the amount of convergence were different, but it is not correct of the convergence actually used." (Helmholtz, 1925 p. 318.)

The quote from Helmholtz makes reference to two features of the brain's estimate of the convergence state of the eyes, that is its accuracy and precision (Collewijn & Erkelens, 1990, Harris & Dean, 2003), both of these could have important consequences for the estimation of distance from vergence. The brain's estimate of the vergence signal could be biased/inaccurate if, for example, the eyes were more or less converged than is needed for object fixation. If this were the case vergence specified distance might be different to the physical distance of the object and if the brain used this incorrect distance estimate to scale retinal disparity, perceived 3-D shape would be subject to bias. The second issue is that of precision, the visual system is not a perfect information processing system; as a consequence its estimates of signals such as the vergence state of the eyes will be subject to noise and uncertainty.

Uncertainty in the vergence state of the eyes could be introduced in a number of ways and at multiple stages during processing, from muscular instability in the ocular-motor system to the stochastic nature of neural firing. Because of this there will be no unique estimate of a signal such as vergence, but instead a distribution of estimates. In making estimates about world properties such as distance from ocular convergence the brain must therefore make decisions based on probability distributions substantiated as neural population activity. A lack of precision in the vergence signal will inevitably lead to a lack of precision in vergence specified distance, but it is also the case that a lack of precision in the vergence signal could act to introduce bias into vergence specified distance estimates. Initially discussion will focus on the accuracy of the vergence signal before moving onto the role of precision.

3.1.6 The accuracy of the vergence signal

It is known that the visual axes of the eyes may fail to converge as intended when fixating an object; this results in the object having a fixation disparity. Fixation disparity varies from around 6 arc min in the fovea to as much as 20 arc min in peripheral vision, but this is typically not noticed by the observer as corresponding image points project to within Panum's fusional area (Howard, 2002). This area represents the retinal area in which corresponding image points can fall in the two eyes off of the horopter (the locus of zero disparity for a given fixation) and still be fused. The boundaries of Panum's fusional area are determined by the maximum disparity an image point can have with respect to fixation without becoming diplopic. The size of Panum's fusional area increases when moving from central to peripheral vision, which explains why a larger fixation disparity can be tolerated in the periphery (Howard, 2002). This increase in size may act to compensate for the increased variability of disparity likely to be encountered at increasing eccentricities from fixation in the natural environment (Hibbard, In Press).

3.1.7 The role of dark vergence and dark focus

In low luminance reduced cue settings the vergence state of the eyes is biased toward a tonic state also referred to as dark vergence, this acts to introduce a fixation disparity. The dark vergence state of the eyes is thought to arise from tonic efference from the eyes' extraocular muscles, the muscles of neck and the vestibular system (Howard, 2002). Ivanoff and Bourdy (1954; quoted in Owens & Liebowitz, 1980) showed that as the luminance of a non-fixated binocular target was decreased a fixation disparity became present. The majority of the fixation disparities were biased toward an intermediate level of vergence (see also Ivanoff, 1955). The distance corresponding to this dark vergence state has been shown to vary between 50cm and optical infinity (when the visual axes are parallel), with a mean of around 1.16m (Owens & Liebowitz, 1980). The eyes also adopt a characteristic accommodative state in the dark known as 'dark focus'; this tends toward an intermediate level of focus equivalent to 0.32m to 1.5m, with a mean of around 0.5m, and also varies greatly between individuals (Owens & Liebowitz, 1976). There is very little correlation between the distance corresponding to an individual's dark vergence and dark focus (Owens & Liebowitz, 1980), which contrasts with the normal coupling shown between accommodation and vergence during natural viewing (Howard, 2002).

Given the role of accommodation and vergence as cues to object distance (Watt et al., 2005a), attention has focused on the possible role of dark vergence and dark focus in biasing estimated distance and estimated 3-D shape. Owens and Leibowitz (1976) investigated the relationship between dark vergence, dark focus and the specific distance tendency (SDT). The specific or 'abathic' distance around which distance estimates contract was measured with a monocular motion parallax technique similar to Gogel and Tietz (1973). Dark focus was measured using a laser optometer and dark vergence was measured using a nonius line technique. Dark focus was found to be uncorrelated with an individual's abathic distance, whereas dark vergence correlated with an individual's abathic distance, this correlation was greater when individual's lateral phoria was corrected for ($r = 0.76$) compared to before correction ($r = 0.56$) (Owens & Liebowitz, 1976).

In a separate study observers adapted to base-out prisms and lenses which maintained the natural link between accommodation and vergence. Observers then manually indicated the perceived distance of a binocularly viewed point of light without visual feedback about hand position. Prism adaptation altered perceived distance as would be expected from the prismatic geometry, dark vergence was affected in a consistent manner, but dark focus was unaffected (Owens & Liebowitz, 1980). Although these data suggest that dark vergence and distance perception are both affected by prism adaptation in a consistent way, it is not clear whether the change in dark convergence can quantitatively predict the change in perceived distance.

3.1.8 The possible relationships between vergence and estimated distance

There are a number of stages in the estimation of distance from the vergence state of the eyes at which bias and uncertainty could be introduced; these are schematically illustrated in Figure 3.2. Bias and uncertainty could be introduced at any one of these stages. Discussion will focus on three possible ways in which estimated distance might be biased, for the purposes of this discussion its will be assumed that the estimated distance is reported correctly, $\hat{D} = D_R$, i.e. there is no contraction bias operating. The first way in which estimated distance might be biased is if the physical vergence state of the eyes θ_V was biased away from the correct vergence state needed for object fixation θ_F. The brain's estimate of the vergence state of the eyes $\hat{\theta}$ could be correct, such that $\hat{\theta} = \theta_V$, but due to the physical misconvergence of the eyes this would result a biased estimate of object distance $D_F \neq \hat{D} = D_V$. An example of this type of bias is the case where the dark vergence state biases the physical convergence state of the eyes resulting in a fixation disparity (Owens & Liebowitz, 1976).

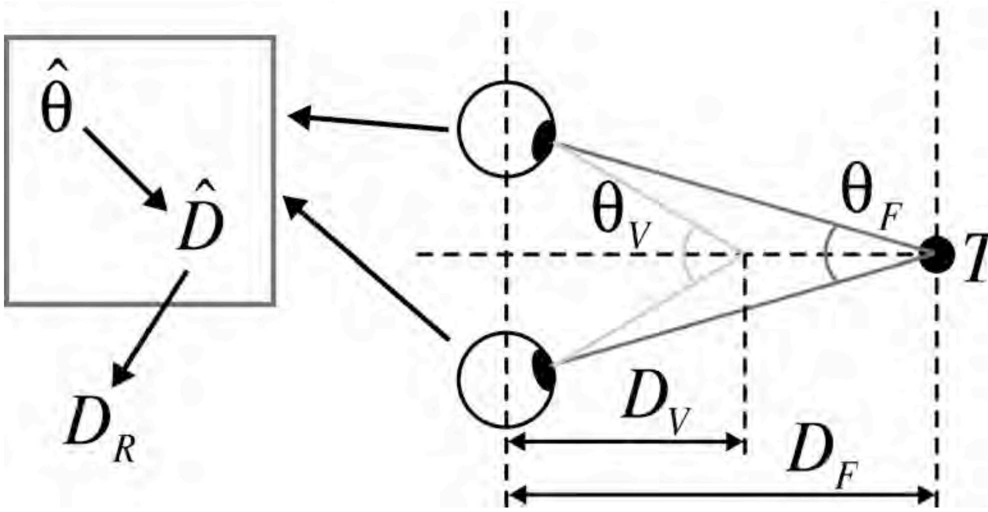

Figure 3.2: Schematic diagram showing the estimation of distance from the vergence state of the eyes. The observer is required to fixate an object T at a distance D_F this defines the vergence angle θ_F. The physical angle of convergence θ_V, which corresponds to a distance of D_V, may or may not correspond to the vergence angle needed for correct fixation. The brains estimate of the physical state of convergence is represented by $\hat{\theta}$. Given this vergence estimate an estimate of object distance $\hat{D}$ can be made. This distance estimate is then reported through a certain means of response e.g. verbally or manually, this reported distance estimate is represented by D_R.

It is likely that some of the disparities present within experimental stimuli for shape and distance judgement tasks would not be fusible for the observer. However, it is unlikely that physical misconvergence of the eyes could account for the biases found in estimates of distance and 3-D shape. The misconvergence required to account for the biases would result in the two eyes' images becoming diplopic such that the observer would be unable to do the task. For example, Johnston (1991) got observers to fixate a cross on the median plane at the same viewing distance as the stereoscopic stimulus in order to ensure correct vergence demand. If it is assumed that (1) all error in set shape is due to observers misestimating object distance and (2) observers estimated distance solely from the vergence state of the eyes, which were misconverged at this incorrect distance, it is possible to get an estimate of the tolerable fixation disparity that would be required for the observer to misconverge at the scaling distance yet still see the fixation cross as fused[4], as was required in the experiment.

The value of fixation disparity required in Johnston (1991) is approximately 136 min arc for the near viewing distance (53.5cm) and approximately 117 min arc at the far viewing distance (214cm). These values are clearly much larger than the ≈ 6 min arc fixation disparity tolerable with foveal fixation of a target (Howard, 2002). Qualitatively different impressions of depth can be gained with diplopia and observers are known to feel uncertain in making shape settings during tasks such as the ACC (Todd & Norman, 2003), but observers do not generally experience diplopia whilst competing experimental tasks and report being able to complete these tasks on the basis of perceived shape. This suggests that they make their judgements about shape on well-fused images, and are therefore converged on the object to within Panum's fusional area whilst doing the task.

A second possible source of bias could be introduced if the eyes were correctly converged on the stimulus such that $\theta_V = \theta_F$, but the brain's *estimate* of the vergence state $\hat{\theta}$ was biased such that $\hat{\theta} \neq \theta_V$, this would result in bias in the perceived distance of the stimulus $\hat{D} \neq D_F$. This type of bias could be introduced if dark vergence acted to bias the vergence *estimate*, rather than the physical convergence state of the eyes. As a final example, consider a case where the eyes were correctly converged on the object and the brain estimated this vergence state correctly, such that $\hat{\theta} = \theta_V = \theta_F$, yet estimated object distance was different to the physical distance of the object, $\hat{D} \neq D_F$. In a Bayesian framework this type of bias could result from the mapping that relates the vergence state to a distance estimate, as embodied in the likelihood function, or the combination of sensory information about object distance with prior distance information (Mamassian et al., 2002).

[4] This was accomplished by calculating the difference between the vergence angle for fixation at the distance of the fixation cross and the vergence angle for fixation at the scaling distance.

3.1.9 The precision of vergence: distance estimation as Bayesian inference

So far discussion has focused on bias/inaccuracy in the vergence signal during viewing in the reduced cue setting and the effect this would have on the estimation of distance. Alternatively, the visual system's estimate of distance from vergence could be unbiased but noisy/imprecise, this noise could introduce bias into distance estimates made by the observer. Noise could be introduced into processing at a number of stages during the formation of a distance estimate. Consider the case where the distance estimate from vergence can be represented by a Gaussian likelihood function with its peak centred on the true object distance. In this case vergence would be an unbiased but noisy estimator of distance. In the Bayesian framework this sensory information may then be combined with a Gaussian prior representing information about the statistical structure of the environment (Yang & Purves, 2003). Information about the statistical structure of the environment could be learned by the observer and used when making perceptual estimates (Adams, Graf & Ernst, 2004).

Combination of the likelihood and prior produces a posterior distribution from which a distance estimate is made (Mamassian et al., 2002, Yang & Purves, 2003). When priors and likelihoods are modelled as Gaussian distributions their relative variances determine the position of the peak (and mean) of the posterior distribution upon which estimates are made. In this way the level of noise in the distance estimate from vergence could, in conjunction with prior information, act to alter the distance estimate. Although modelling vergence specified distance as a Gaussian distribution centred on the true object distance is a reasonable approach to take, it is also possible that variability in the vergence signal could act to skew the likelihood function for distance from vergence through the way in which the vergence signal is transformed into a distance estimate. This would result in vergence no longer being an unbiased estimator of distance; noise would act to bias sensory information about distance from vergence before it was combined with prior information.

Noise in early visual signals has been shown to play an important role in predicting distortions of perceived geometry in a range of geometric illusions (Fermuller &

Malm, 2004) and in conjunction with a prior for slow velocities, the distortions present in a number of motion illusions (Weiss et al., 2002). Hogervorst and Eagle (1998) investigated the role of noise in early visual measurements as a determinant of systematic distortions of perceived 3-D shape in monocular structure from motion sequences. When viewing a structure from motion sequence of an 'open book' dihedral angle rotating around its vertical axis systematic misperceptions of its shape are found. Specifically, depth is underestimated with small dihedral angles and small angles of rotation, but becomes more veridical as the angle of rotation is increased. With increasing rotation angle depth becomes overestimated, especially for larger dihedral angles (Hogervorst & Eagle, 1998).

Hogervorst and Eagle (1998) formulated a Bayesian model for the estimation of 3-D shape from these structure from motion sequences. The model assumed realistic estimates of zero-mean Gaussian noise in retinal image measurements of direction and speed; this introduced no bias into image measurements, just variation around their true value. With the introduction of measurement noise there is a distribution of possible 3-D dihedral angles, represented by the posterior probability distribution, which are consistent with the image measurements. The mean of the posterior distribution was taken as the observer's predicted percept. The model was able to quantitatively predict the observed bias and variability in estimated 3-D shape. The bias predicted by the model is a result of the nonlinear mapping between the measurement space and space of 3-D scenes, this nonlinear mapping causes the Gaussian noise in retinal measurements to be distorted into a non-Gaussian posterior probability distribution for 3-D shape (Hogervorst & Eagle, 1998). The elegance of this approach is that bias in perceived 3-D shape can be predicted on the basis of noise in early visual processing without the need to introduce biases, defaults or priors.

3.1.10 Summary: the role of vergence in the estimation of distance and three-dimensional shape

Observers are found to systematically misperceive the 3-D shape of disparity-defined objects, such that the same object presented at different distances is perceived to have a different three-dimensional shape (Bradshaw et al., 2000). If it is assumed that all error in set shape is due to misestimating object distance, this misperception of shape is consistent with the overestimation of near distances (below $\approx$1m) and the underestimation of far distances (beyond $\approx$1m) (Glennerster et al., 1996, Johnston, 1991). A link has been made between the misestimation of 3-D shape and the misestimation of distance from vergence. Geometrically vergence should be a reliable cue to near distances and become progressively less reliable as distance increases (Collewijn & Erkelens, 1990). Overall the distance perception literature supports these predictions. Many instances where vergence is suggested to be a poor distance cue can be explained by conflicting visual cues (Tresilian & Mon-Williams, 1999) or assumptions about how observers are completing the task at hand (Gogel & Tietz, 1973).

The accurate perception of near distances from vergence suggests that biases in perceived shape cannot be solely accounted for by the misestimation of distance. There is however large variation in the results of distance and shape perception studies (Todd & Norman, 2003), which may suggest other processes, such as the adoption of task dependent heuristic strategies (Glennerster et al., 1996, Todd, 2004), may be at work. There are a number of stages at which the estimation of distance from vergence might be subject to bias and uncertainty. One such source of bias, a physical misconvergence of the eyes, cannot satisfactorily account for biases in perceived distance and shape. A number of other factors such as the influence of dark vergence (Owens & Liebowitz, 1976) or prior distance information (Yang & Purves, 2003) could contribute to the misperception of distance. However, before introducing priors it is important to assess the role that early visual noise has on perceptual estimation. The role of noise in sensory signals as a source of perceptual distortions was highlighted; this is the focus of the modelling and experiment detailed in the remainder of this chapter.

3.2 Modelling systematic distortions in the estimation of distance and three-dimensional shape

3.2.1 Description of the model

Monte-Carlo simulations were run in the Matlab programming environment to investigate the effect of noise in vergence, version and disparity signals on the visual system's estimation of distance from vergence, depth from scaled disparity and height from scaled angular size. Estimates of each of these variables must be made if the visual system is to make judgements about the 3-D shape of objects, such as when comparing the depth of an object to its width or height (e.g. Bradshaw et al., 2000). Simulations were carried out to model performance in a task where an observer is required to set the depth of a vertically orientated triangle defined by three points to match its height (Figure 3.3). The triangle has a depth d and height h and is situated at a distance D from the observer, who has a half interocular distance I. The geometrically correct half-vergence angle θ, half-disparity η and half-angular size α, are given by equations 3.1, 3.2 and 3.3 respectively[5].

$$\theta = \arctan(I/D) \tag{3.1}$$

$$\eta = \arctan\left(\frac{Id}{D(D-d)+I^2}\right) \tag{3.2}$$

$$\alpha = \arctan\left(\frac{h/2}{D}\right) \tag{3.3}$$

Simulated noise took the form of n samples of zero-mean Gaussian noise added to the geometrically correct vergence, disparity or angular size signal. This resulted in

[5] A derivation of equation (2) can be found in Appendix 1.

the vergence, disparity and angular size signals being represented by Gaussian distributions of signal estimates centred on the geometrically correct signal value with standard deviations corresponding to an estimate of that signals uncertainty (Figure 3.3). The use of zero-mean Gaussian noise introduced no bias into the signals, just variability around their true value. It is an empirical question as to the value of noise to use in the simulations; this will be addressed as the role of noise in each signal is discussed.

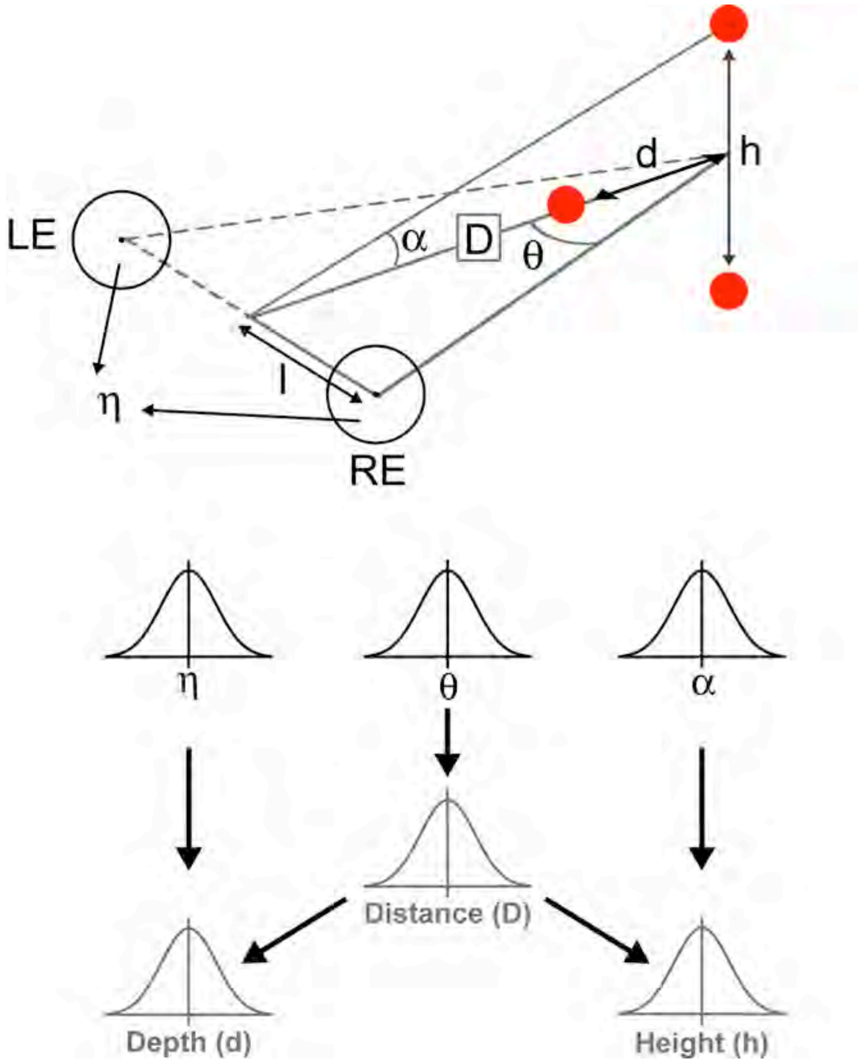

Figure 3.3: Diagram showing details of the Monte-Carlo simulations. The upper portion of the diagram shows the task modelled, which was to set the depth, d, of a vertically orientated triangle defined by three points to match its height, h. The observer's left and right eyes are labelled LE and RE respectively and define the half-interocular distance I. When viewed at a distance D the depth of the triangle projects a half-disparity, η, at the retina and results in the half-vergence angle, θ. The half height of the triangle subtends an angle, α. The effect of

noise in vergence, disparity and angular size signals was investigated for the estimation of distance, depth and height. A plan view of this process is shown in the lower portion of the diagram, visual signals are shown in black and perceptual estimates shown in grey. Noisy signals were simulated as Gaussian distributions centred on the geometrically correct signal values. Noise in early visual signals results in a distribution of estimates for world properties such as distance, depth and height, this process was investigated separately for each of the three visual signals. For more information see the accompanying text.

In order to estimate distance from vergence, depth from scaled disparity and height from scaled angular size it is assumed that the visual system has estimates of ocular convergence $\hat{\theta}$, retinal disparity $\hat{\eta}$, the angle of gaze $\hat{\alpha}$ and the interocular distance $\hat{I}$. For the simulations reported it is assumed that the visual system has accurate knowledge of the interocular distance i.e. $\hat{I} = I$. If the visual system has estimates of these values it could estimate distance $\hat{D}$, depth $\hat{d}$ and height $\hat{h}$ using the relationships shown in Equations 3.4, 3.5 and 3.6 respectively. It is clear from these equations that vergence noise will affect the estimation of distance and also the estimation of depth and height, whereas noise in the gaze angle will effect only the estimation of height, and noise in disparity only the estimation of depth. This is clearly a simplification as vergence, gaze angle and disparity are unlikely to be completely independent signals. For example, the disparity of an object will depend on the distance of fixation. However, it is worthwhile to understand how noise in each of these signals will effect the estimation of distance and 3-D shape.

$$\hat{D} = I / \tan\hat{\theta} \tag{3.4}$$

$$\hat{d} = \frac{\tan\hat{\eta}\left(\hat{D}^2 + I^2\right)}{\hat{D}\tan\hat{\eta} + I} \tag{3.5}$$

$$\hat{h} = 2\left(\hat{D}\tan(\hat{\alpha})\right) \tag{3.6}$$

Simulations were carried out for noise in each of the signals independently, for instance, when the role of noise in the vergence signal was investigated disparity and the angle of gaze were assumed to be geometrically correct and unaffected by noise. This strategy allows the relative importance of noise in each signal to be determined. Noisy visual signals will result in a distribution of possible values for world properties such as distance, depth and height; these distributions were simulated using Equations 3.4, 3.5 and 3.6. For instance, from a distribution of vergence signal estimates it is possible to calculate the corresponding distribution of distance, depth and height values that this noisy signal would produce when combined with geometrically correct estimates of disparity and gaze angle.

Taking noise in the vergence signal as an example, in the simulations there is a Gaussian distribution of n samples corresponding to the vergence signal. For each of these n samples it is possible to calculate the corresponding distance value using Equation 3.4. This produces a distribution of distance values, each of these values can then be used to calculate the corresponding depth and height values using Equations 3.5 and 3.6 respectively. This results in a distribution of depth and height values consistent with the noisy vergence signal. For the simulations reported here $n = 500{,}000$, unless otherwise stated. The distributions of distance, depth and height are produced by constructing histograms of the n samples of each signal with a bin size equal to 1mm. The frequency of samples in each of these bins is then plotted against the bin value to form the distributions shown.

No restrictions were placed on the possible values of distance, depth and height, to ensure that no bias was introduced into the distributions as a result of imposing limits on the values these variables could take. This means that in some instances these values may take on unrealistic values, this is discussed further below. With noisy visual signals there is a distribution of possible world states consistent with the sensory information, from this distribution the visual system must make a single perceptual estimate regarding a property of the world. This process was modelled as choosing the peak of the distance, depth or height distributions.

Choosing the peak of each distribution has a simple interpretation, consider a distribution of distance estimates from vergence, the peak of the distribution represents the most likely distance to have produced the current sensory information available from the vergence cue. This is equivalent to maximum likelihood estimation (MLE) where the peak of the likelihood function is chosen as the perceptual estimate (Ernst & Banks, 2002). However, the brain might use alternative decision rules depending on the gains and losses associated with making a given decision on the likelihood or posterior distributions.

For example, the least squares loss function results in a loss equal to the square of the difference between the estimated world state and the actual world state. With this loss function the MLE (or MAP) rule is no longer the best rule to follow in order to maximise expected gains, choosing the mean of the posterior distribution now results in the maximum gain (Mamassian et al., 2002). When a distribution is Gaussian its peak and mean are the same, but when the distribution is skewed, the peak and the mean signal different estimates. Typically estimation and cue integration have been modelled with Gaussian probability distributions (Ernst & Banks, 2002, Hillis et al., 2004), however, this is likely to be a simplification as sensory signals may be non-linearly related to world properties (Hogervorst & Eagle, 1998). This may have important implications for models of cue combination and for explaining biases found in the perception of 3-D shape.

3.2.2 Simulation results

3.2.2.1 Estimating variability in the vergence signal.

Estimating the variability of the vergence signal used for scaling is not an easy task and is out of necessity going to be indirect. Currently, estimation of its value can be approached in two ways. The first is by making physiological measurement of the variability in the convergence state of the eyes. The second is by inferring vergence variability from tasks that require an absolute estimate of the vergence signal, such as the estimation of absolute distance. As was discussed above, variability in the vergence state of the eyes needs to be within Panum's fusional area for diplopia to be avoided, this can be achieved simply by measuring the signed disparity of a "fixated" point and shifting the eyes accordingly to minimise this value. This means that no absolute estimate of the vergence is required. Measuring eye stability during fixation is therefore likely to underestimate the level of variability in the vergence signal. Direct measurement of eye orientation also takes no account of noise introduced through efferent feedback from the muscles and subsequent processing of the vergence signal.

Taking the second route, converting variability in absolute distance estimates into variability in the vergence signal is also problematic. The foremost problem is namely that being investigated here, the mapping that relates vergence and estimated distance, and the process that is used to make a perceptual estimate. For example, the vergence signal could be relatively noisy, resulting in a broad distribution corresponding to the vergence signal. However, once this is mapped into a distribution of distance values, and an estimate is made from this distribution, there is no reason why the distribution of these distance estimates should reflect the variability in the vergence signal well. This represents a problem for the estimation of vergence variability. At the very least increased noise in vergence should result in increased noise in distance estimates, as the estimation of absolute distance requires an estimate of the vergence state of the eyes to be made. So arguably variability in attempts to reproduce a target's distance provide a better indication of the variability of the ocular convergence signal (Brenner & van Damme, 1998).

Brenner and van Damme (1998) got observers to make equal, half or double distance settings between a reference and target stimulus presented at eye height. The reference was presented 10 degrees to the left of the median plane, and the adjustable target 10 degrees to the right. The presentation of the reference and the target was contingent on the horizontal direction of gaze. This ensured that the reference and target were never viewable at the same time, so as to remove relative disparity as a cue. The same settings were also made but in conditions without a seen reference, these conditions directly followed the block of trials where the observer was making the same distance setting but with gaze contingent presentation of a reference and target. Therefore in these conditions observers were reproducing the distance of reference seen in a just-completed block of trials. All conditions were carried out at both 1 and 2m viewing distances.

With gaze contingent presentation, standard deviations for distance settings, expressed in angular terms, averaged across observers and viewing distances, were approximately 10, 24 and 19 min arc for equal, half and double distances settings. Without the reference present standard deviations increased to about 55 min arc for the equal and half distance settings and about 30 min arc for the double distance settings. Using a similar design Brenner and Smeets (2000) got observers to align two vertically separated lines laterally and in depth. The lines were computer generated and presented at screen distances of 30, 60 and 130cm. The upper reference line appeared in one of four positions, 8% nearer or further than the screen distance and 2.6 degrees to the left of the median plane or 2.1 degrees to the right of the median plane. Again gaze contingent presentation ensured that the lines were never viewable at the same time. The average standard deviation of depth settings across subjects, collapsed over condition, was approximately 6 min arc and remained constant across the distances tested, suggesting that angular error in vergence is relatively independent of the convergence state of the eyes in this distance range.

It is likely that in both Brenner and van Damme (1998) and Brenner and Smeets (2000) observers were able to make use of information about the change in eye posture across sequential fixations of the target and reference. This 'delta vergence' mechanism could provide either relative disparity information, if observers were able to accurately register a change in vergence, or absolute distance information, if the absolute vergence posture of the eyes was known for at least one point in the scene (Backus & Matza-Brown, 2003, Brenner & van Damme, 1998). In a task where an observer is matching the distance of a reference by adjusting the distance of a second target it would be possible to complete the task by nulling the relative disparity of the two points as gathered over sequential fixations. An observer could also make double and half distance settings if they had knowledge of the relative disparity that corresponds to a given difference in distance, taking into account changes in version.

When matching the distance of a reference point with a target it is an open question as to the length of time to leave between the presentations of each in order to eliminate relative information from sequential fixation. For example, stereoacuity thresholds have been shown to rise from around 1 min arc with an interstimulus interval of 0 seconds to around 30 min arc with an interstimulus interval of 32 seconds (Foley, 1976). The variability of verbal or manual distance estimates when vergence is the only distance cue might be more representative of the variability in ocular convergence, because here observers are not comparing the distance of a reference and target over time.

Variability of distance estimates is known to depend on the method of response and the more precise method of response is known to depend on the individual. This indicates that verbal and manual responses are not a pure indicator of the variability of distance estimates or ocular convergence, but in part reflect variability inherent in the response process. For example, variability of vergence was estimated for a distance estimation task where manual and verbal distance estimates were made[6]. Mean standard deviations for distance estimates across observers, expressed in angular terms, were 155 for verbal estimates and 143 min arc for manual estimates. Standard deviations differed greatly between the three observers tested, 39-385 min arc for manual estimates and 78-216 min arc for verbal estimates. For a given observer, low variability with one method of response did not necessarily correspond to low variability in the other method.

Overall, large differences in estimated vergence variability can be gained from different response tasks, the inferred variability is dependent on factors such as whether relative disparity or distance information is available over sequential fixations of a reference and target, whether a distance estimate is made to a previously seen target or one that is continuously visible and the mode of response by which the observer is making their estimate. Variability across observers is relatively unsurprising and may in part account for the large individual differences found in shape judgement tasks (Helmholtz, 1925, Todd, 2004, Todd & Norman, 2003).

[6] In the task the observer was presented with two targets. The upper target was presented at eye level at one of 11 distances between 15 to 45cm in steps of 3cm, the lower target was presented 3 degrees below eye level at either 21, 30 or 39cm. Observers manually and verbally estimated the distance of the lower target in randomly interleaved trials (n=198 manual, n=66 manual). There was no interaction between the upper and lower target's distance. Unpublished data Drga (2006).

However, it adds to the difficulty inherent in making an estimate of vergence variability from a distance estimation task.

3.2.2.2 Vergence noise: Monte Carlo simulation results

Monte Carlo simulations were carried out to investigate the effect of vergence noise on the estimation of object distance, depth, height and depth to height ratio. Six viewing distances were simulated from 30 to 130cm in steps of 20cm, this is roughly equivalent to the range of distances typically used in 3-D shape judgement tasks and covers the range over which vergence should be of maximum use as a cue to distance (Collewijn & Erkelens, 1990, Mon-Williams & Tresilian, 1999). Object depth and height were both 5cm. A noise level of 100 min arc (50 min arc in the half vergence angle) was considered a reasonable value to use to initially demonstrate the effect of vergence noise on perceptual estimates. A broader range of vergence noise values from 20 to 180 min arc were used to more systematically investigate the effect of noise level on perceptual estimates. The simplifying assumption was made that vergence variability in angular terms was independent of the viewing distance.

A set of distance, depth and height distributions are shown in Figure 3.4, a systematic effect of vergence noise can be seen in the distribution of distance values (Figures 3.4a and 3.4b). At the nearest viewing distance of 30cm the distribution of distance values is narrow and peaks close to the true object distance. However, as viewing distance increases the distance distributions become flattened and systematically skewed so that the peak in the distributions shift to an underestimation of object distance. This shift progressively increases with increasing viewing distance and can be more clearly seen in Figure 3.4b where the distributions have been normalised to each simulated object distance in order to highlight the shift, negative values indicate distance underestimation and positive values distance overestimation. The peaks of the distributions shown in Figure 3.4a are shown in Figure 3.5a, from this it can be seen that near distances are estimated reasonably accurately but distances are progressively underestimated as the viewing distance increases.

The effect of vergence noise on distance estimates has predictable effects on both the depth and height distributions (Figures 3.4c and 3.4d respectively). At near viewing distances the distributions are peaked close to the actual object depth and height, but with increasing viewing distance the distributions become progressively flattened and skewed toward lower values. The peaks of both distributions shift from near correct estimation at 30cm to progressive underestimation of depth and height at larger viewing distances (Figure 3.4b), this can be seen more clearly in Figure 3.5b. The underestimation of depth with increasing distance is more pronounced than the underestimation of height, this is due to depth being proportional to the square of distance ($d \propto D^2$) and height being directly proportional to distance ($h \propto D$). It is assumed that the same estimate of distance is used to scale both depth and angular size (vanDamme & Brenner, 1997). The differential effect of distance misestimation on depth and height results in perceived depth to height ratios less than one across the whole distance range (Figure 3.5c), this is more pronounced at far compared to near viewing distances.

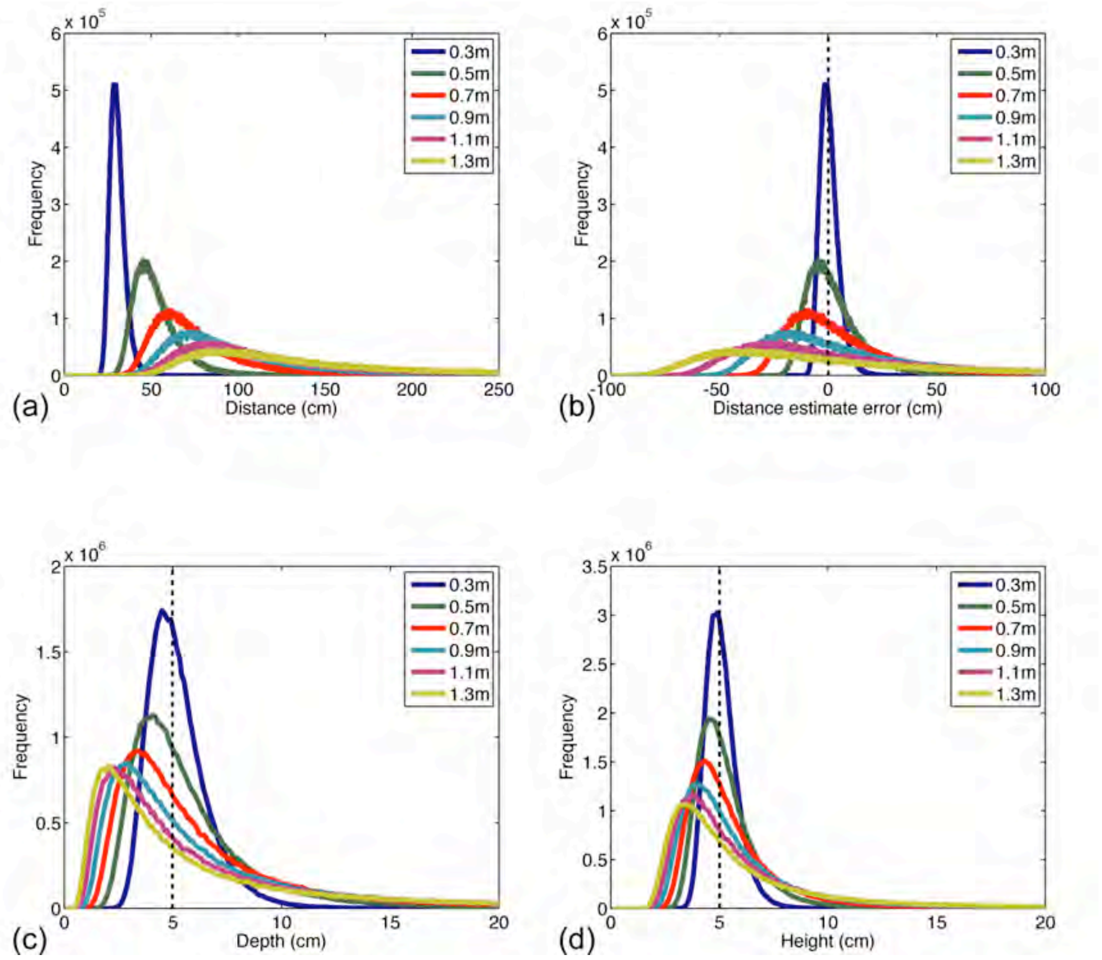

Figure 3.4: Example distance, depth and height distributions from Monte Carlo simulations investigating the effect of vergence noise. (a) Shows the distribution of distance values for each of the six simulated viewing distances. (b) Shows the same distance distributions but each distribution has been normalised to the viewing distance used in the simulation. This highlights the types of error present in the distance values; the dashed line on the plot indicates no error, negative values indicate underestimation of distance and positive values overestimation of distance. (c) Shows the distribution of depth values produced by scaling the geometrically correct disparity value with the distribution of distance values for each simulated viewing distance, the dashed line shows the correct depth. (d) Shows the distribution of height values produced by scaling the geometrically correct angular size with the distribution of distance values for each simulated distance, the dashed line shows the correct height. Note that for clarity the distance distributions have been clipped at 1mm and 2.5m and the depth and height distributions at 1mm and 20cm, for details see the accompanying text.

The distance distributions shown in Figure 3.4a have long flat tails at larger viewing distances, this is a result of the way in which noise affects the geometry of the vergence signal. At larger viewing distances the vergence angle becomes smaller and as a result is more affected by a constant level of angular noise, this leads to a much greater change in estimated distance at far distances. For example at a viewing distance of 50cm a positive 50 min arc error in half vergence results in a change in distance of $\approx$3cm, whereas at 1m the same error in vergence results in a change in distance of $\approx$30cm. As was pointed out by Mon-Williams and Tresilian (1999) the effect of vergence error is asymmetrical about the true object distance. A fixed error in vergence ε has a greater effect on estimated distance if the error is negative (making the vergence angle smaller) compared to if the error is positive (making the vergence angle larger), this is shown in Equation 3.6. Using the notation of Mon-Williams and Tresilian (1999), D is the true viewing distance, θ the half vergence angle, I the interocular distance and ε an angular error in the half vergence angle.

$$D - I\cot(\theta + \varepsilon) < I\cot(\theta - \varepsilon) - D \tag{3.6}$$

The relationship shown in Equation 3.6 means that an angular error in vergence that decreases the vergence angle ($\theta - \varepsilon$) has a greater effect on estimated distance than the same level of angular error that increases the vergence angle ($\theta + \varepsilon$). This asymmetry in the effect of vergence noise is small at near viewing distances but becomes larger as the viewing distance increases (Mon-Williams & Tresilian, 1999). This results in distance distributions having a long tail for large viewing distances, corresponding to small vergence angles. The tail for large values in the distance distribution results in a corresponding tails for large values in the distance and depth distributions (which is the reason of the distribution being clipped). Mon-Williams and Tresilian (1999) did not consider the distribution of distance estimates that would be produced by a noisy vergence signal, but pointed out that on *average* noise would lead to an overestimation of far distances. They suggested that the visual system might have knowledge of the asymmetrical effect of vergence noise and could attempt to compensate for this by incorporating a bias toward underestimating more distant targets. This would be able to account for the underestimation of far distances that is found when vergence is the predominant cue to distance (Viguier et al., 2001).

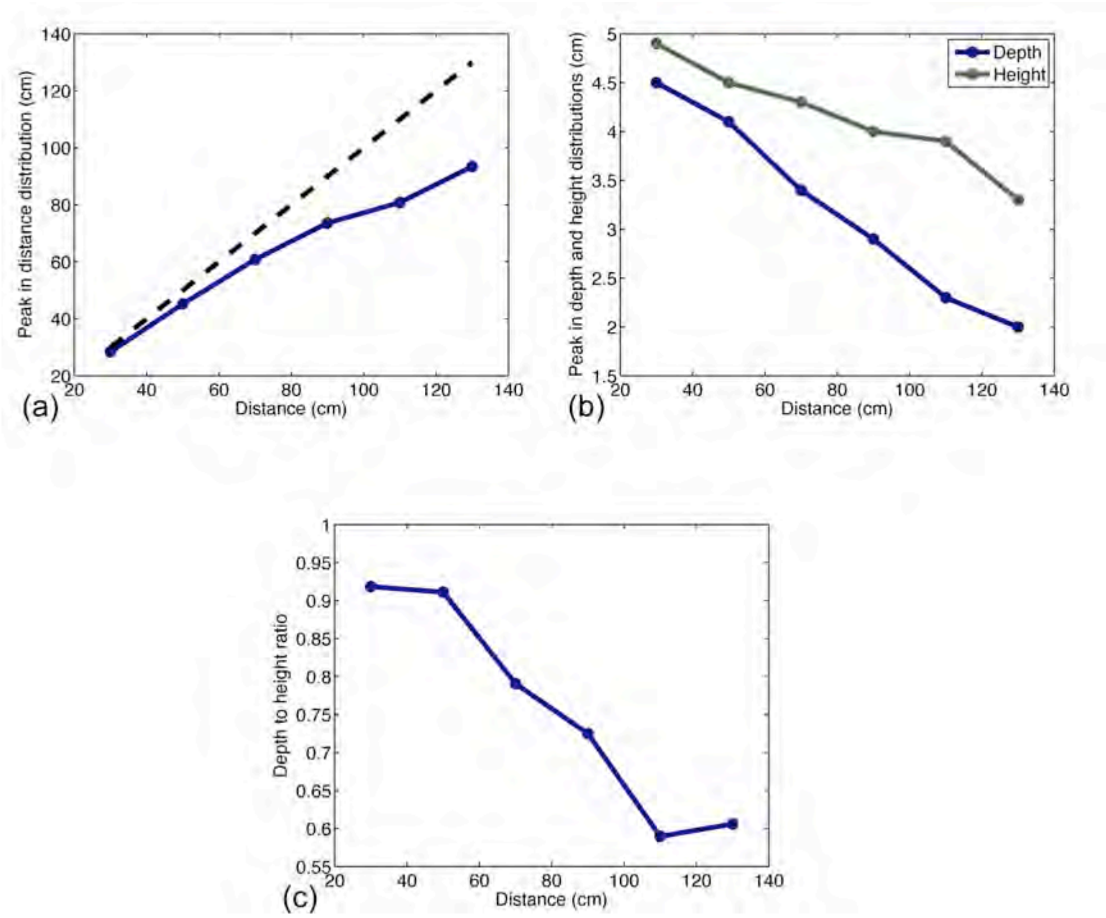

Figure 3.5: The effect of vergence on the peaks in the distance, depth and height distributions and depth to height estimates, (a) shows the peaks of the distance distributions shown in Figure 3.4a, (b) shows the peaks of the depth and height distributions shown in Figures 3.4c and 3.4d, and (c) shows depth to height ratio's calculated by dividing the peak in the depth distribution by the peak in the height distribution at each simulated viewing distance.

Whilst it is true that on average distance estimates increase with increasing noise in the vergence signal i.e. the mean of the distance distribution increases, this is not the case for the peak of the distance distribution. The mean of the distance distribution increases because of the large distance values produced by the asymmetric effect of noise on the vergence angle, but the peak of the distance distribution shifts toward smaller distance values. It is an empirical question as to the processing strategy the visual system adopts when dealing with distributions corresponding to world properties. The decision rule that the visual system uses to make estimates from these distributions will depend on the costs and gains associates with making a given error (Mamassian et al., 2002). The peak in the distance distribution represents the most likely distance to have produced the noisy vergence signal, and if taken as the visual system estimate of distance predicts the underestimation of far distances that is observed experimentally. This underestimation is not predicted by taking the mean of the distribution, unless a post hoc compensation mechanism is implemented (Mon-Williams & Tresilian, 1999).

To get a more comprehensive picture as to how vergence noise affects the estimation of distance, depth, height and depth to height ratio, simulations were run for vergence noises ranging from 10 to 90 min arc in steps of 20 min arc, otherwise the parameters for the simulations were the same as those used previously. Peak estimates of distance, depth and height are shown in Figure 3.6, as are depth to height ratios. A clear effect of vergence noise on estimates of distance, depth and height is seen for a wide range of noise values. For each noise value estimated distance, depth and height are progressively underestimated with increasing viewing distance. This bias systematically increases as the level of simulated vergence noise is increased. The different noise curves converge at near distances; this is expected from the geometry of the vergence signal as the effect of vergence noise becomes less asymmetric at nearer fixation distances.

With the 10 min arc noise very little bias is seen across the whole distance range, by comparison the greatest bias in estimates is seen for the largest noise value at the largest viewing distance, here distances are underestimated by more than half their true value. As was seen previously, depth is affected more than height by underestimates of object distance, which results in depth to height ratios less than one across the whole distance range for most levels of noise. These simulations demonstrate that with increasing viewing distance a progressively greater underestimation of distance, depth and height is predicted if the visual system were to estimate the most likely world values to have produced these noisy sensory signals. The progressive underestimation of distance and depth relative to height as distance increases is consistent with the psychophysical literature (Johnston, 1991, Viguier et al., 2001), but the simulations do not predict the overestimation of depth intervals relative to height intervals found at near viewing distances (Johnston, 1991).

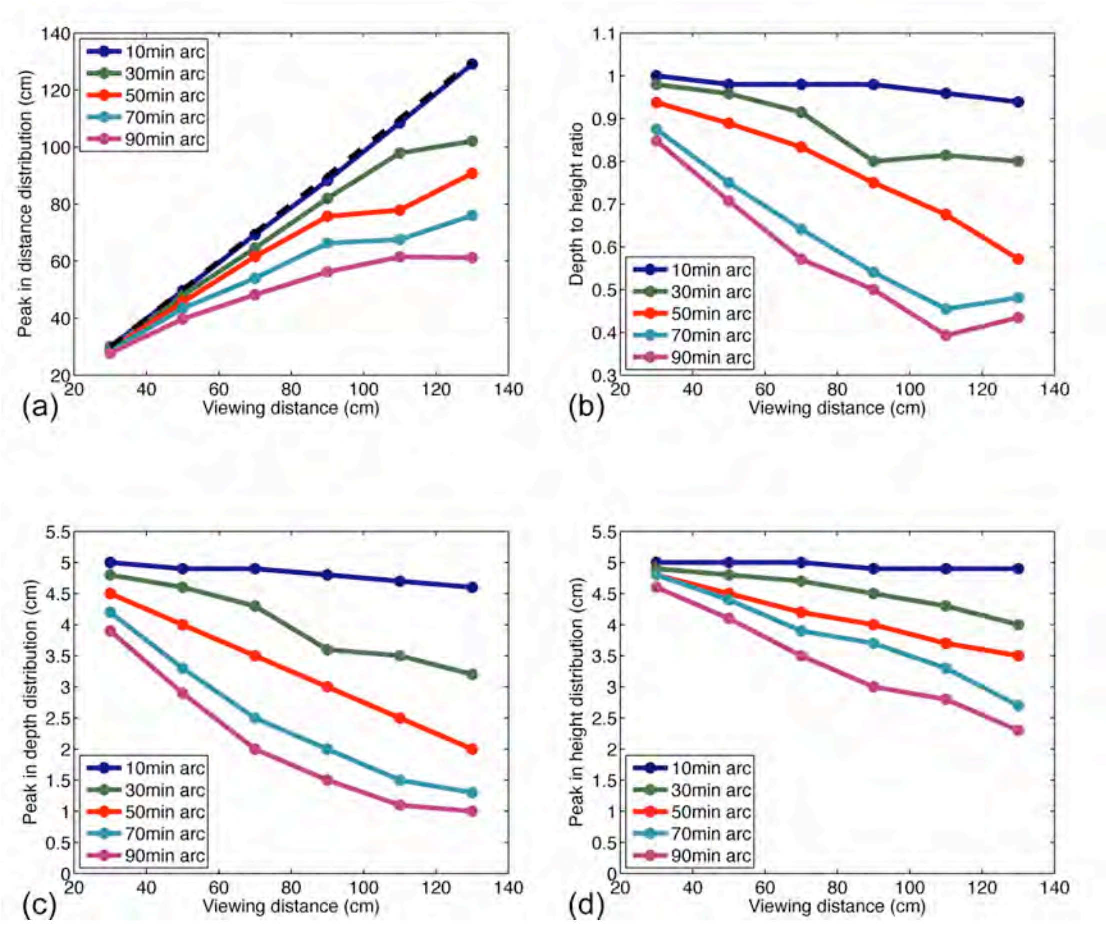

Figure 3.6: The effect of varying levels of vergence noise (10 to 90 sec arc in the half vergence angle) on the peak of the distributions for (a) distance, (c) depth, and (d) height. Peak depth to height ratio (b) was calculated by dividing the peak in the depth distribution by the peak in the height distribution for each simulated viewing distance.

3.2.2.3 The role of disparity noise: Monte Carlo simulations

A second source of noise that could effect the perception of 3-D shape is that in the disparity signal. Measurements of stereoacuity can be taken as a measure of the variability present in the disparity signal. Stereoacuity thresholds measure the ability to detect the depth of a test stimulus with respect to the depth of a reference. Stereoacuity can either be measured relative to a baseline of zero disparity, if the reference is fixated, or measured with fixation on a third nearer or further point, in this instance thresholds are said to be measured with a disparity pedestal (Howard & Rogers, 2002). Stereoacuity thresholds measured on a disparity pedestal are known to be much larger than those measured with fixation of the reference point (Westheimer, 1979). Gaze is typically uncontrolled during 3-D shape judgement tasks, allowing the observer to fixate the object (Glennerster et al., 1996) or in some instances observers are required to fixate a fixation cross to maintain correct vergence demand (Johnston, 1991), this suggests that thresholds measured relative to fixation are the more applicable measure.

Stereoacuity is known to vary with a number of parameters such as spatial frequency and orientation (Bradshaw & Rogers, 1999), pedestal disparity, luminance contrast, horizontal and vertical eccentricity (Howard & Rogers, 2002), the nature of the configuration of the test and reference stimuli (Westheimer, 1979), and concurrent versus sequential presentation of test and reference (Westheimer, 1979). Evidence also suggests that stereo-thresholds may be dependent on the depth of a target point from a local reference plane, such as a surface, rather than the relative depth of two points in the Cartesian z axis running orthogonal to an axis connecting the two eyes (Glennerster, McKee & Birch, 2002). There are numerous common tests of stereoacuity such as the Frisby stereo test, the Keystone stereo test and the Randot stereo test. As accessed by the Keystone stereo test 97% of adults have a stereoacuity of 2 min arc or better whereas 80% have a stereoacuity of 30 sec arc or better (Howard & Rogers, 2002).

Bradshaw and Glennerster (2006) found stereo-thresholds for discriminating whether a horizontal corrugation was concave or convex to be 2-8 sec arc depending on the viewing distance and the observer. Using a similar task Bradshaw and Rogers (1999) found thresholds ranging from 2-3 sec arc with horizontally orientated corrugations with spatial frequencies of 0.5cpd, up to 80-100 sec arc with horizontal or vertical corrugations with lower spatial frequencies. Using a much sparser stimulus consisting of a thin test line subtending 15 min arc and either one or two reference lines, stereo-thresholds were found to be 4-20 sec arc depending on the exact spatial and temporal configuration of the test and reference lines (Westheimer, 1979). Stereo-thresholds measured as the displacement of a central column of dots from a slanted dot grid reference plane have been shown to be around 25 to 36 sec arc depending on the observer (Glennerster et al., 2002).

As can be seen, estimates of disparity noise are much smaller than estimates of vergence noise, typically measured in seconds rather than minutes of arc. Thresholds also vary greatly depending on the precise stimulus characteristics and the nature of the task the observer is performing. Stereoacuity has also been shown to vary with viewing distance, which suggests that both a minimum detectable disparity and a minimum detectable depth may determine its precision (Bradshaw & Glennerster, 2006). A representative value of 20 sec arc for variability in the disparity signal was used to initially investigate the role of disparity noise the estimation of depth. Because the simulations are based on half disparity values this noise value is halved i.e. 10 sec arc. The viewing distances used were the same as those in the vergence simulations, in these simulations vergence and angular size took their geometrically correct values. A more comprehensive range of disparity noises ranging from 5-85 sec arc in the half-disparity signal were used to systematically investigate the role of disparity noise in the estimation of 3-D shape. As with vergence variability, the simplifying assumption was made that the level of variability was independent of the viewing distance.

Example depth distributions are shown in Figure 3.7a (only depth distributions are shown as disparity noise will not effect estimates of distance and height in the simulations). It is clear from the distributions shown in Figure 3.7a that 10 sec arc of noise in disparity does not act to skew the depth distributions and therefore introduces no bias into the position of the peak of the distributions. At the shortest viewing distances the depth distributions are highly peaked at the true depth value, as viewing distance increases the distributions remain peaked at this value but the variance of the distribution increases symmetrically about its peak. This lack of bias in the peak of the distribution is replicated over a wide range of disparity noises from 5-85 sec arc (Figure 3.7b). At the highest values of noise at the largest viewing distances a very small bias is seen (maximum of 3mm), but this is caused by the instability inherent in estimating the peak of very shallow distributions, rather than any bias introduced by the way the noisy disparity signal is transformed to a depth estimate.

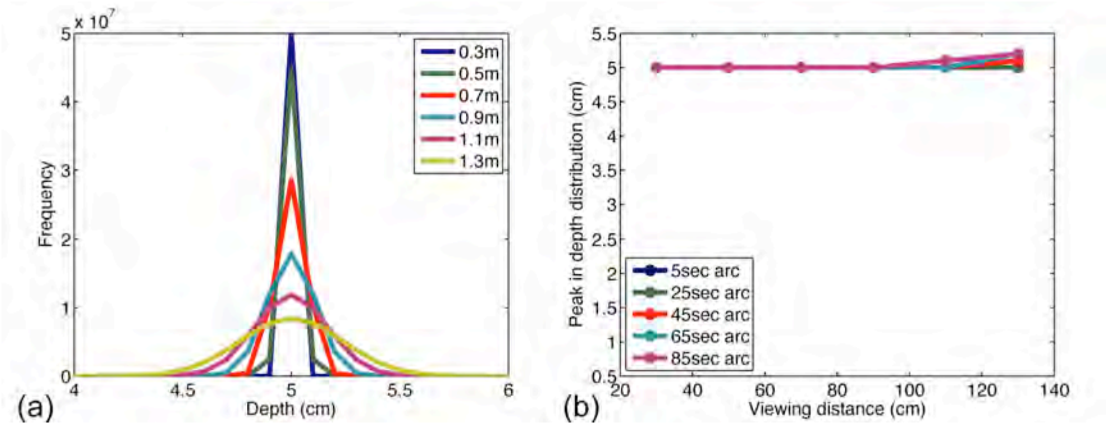

Figure 3.7: (a) Example depth distributions from Monte Carlo simulations investigating the effect of disparity noise on the estimation of depth from disparity. The same six viewing distances were simulated as for the vergence simulations, disparity noise was set to 10 sec arc. The depth distributions have been clipped at 4 and 6cm for clarity as the distribution frequency values drop to (and remain at) zero beyond these values. (b) Shows the peak in depth distributions across distance for varying levels of disparity noise (5 to 85 sec arc).

3.2.2.4 The role of gaze angle noise: Monte Carlo simulations

A third source of noise that could affect performance in estimating 3-D shape is noise in gaze angle, this will alter the angular size of the object, which is scaled to estimate object height. Evidence suggests that in angular terms variability in direction estimates is the same as variability in distance estimates (Brenner & Smeets, 2000). The same values of variability were therefore used for simulations investigating gaze angle noise as for vergence noise. The level of noise again remained constant across distance. Example height distributions are shown in Figure 3.8 (only height distributions are shown as noise in gaze angle will not affect estimates of distance and depth in the simulations). The results of the Monte Carlo simulations show the same pattern of results to that found with disparity noise and depth.

In Figure 3.8a it can be seen that 50 min arc noise in the gaze angle signal did not result in bias in the peak of the height distributions. At the nearest viewing distance the distribution is highly peaked around the true height of the stimulus, as viewing distance increases the variability of the distributions increases symmetrically around this value. This is replicated over a wide range of noise values from 10-90 min arc (Figure 3.8b). Any shift in the peak of the distributions is again small, compared to that found with vergence noise, and due to the instability of the peak of the distribution rather than any inherent bias. The shift in the peaks of the height distributions are larger then those for the depth distributions because the level of simulated noise was higher.

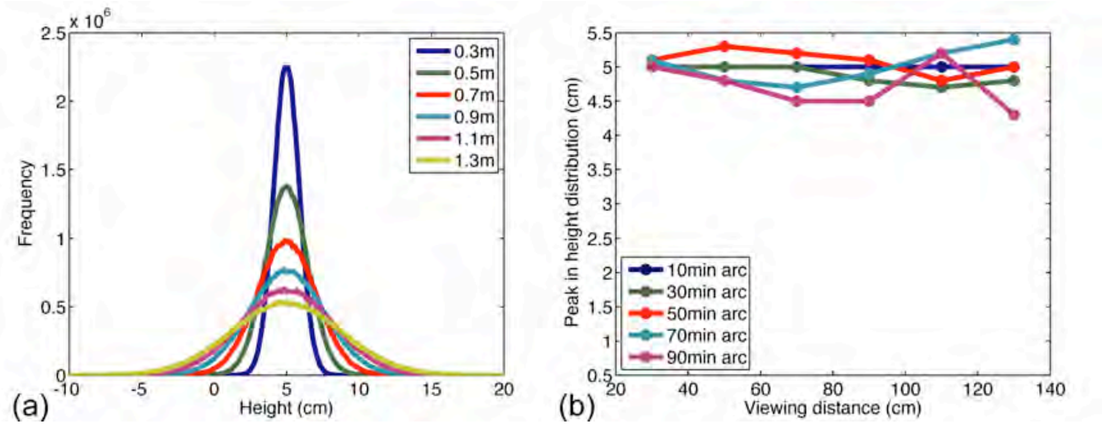

Figure 3.8: (a) Example height distributions from Monte Carlo simulations investigating the effect of gaze angle noise on the estimation of height. The same six viewing distances were simulated as for the vergence simulations, gaze angle noise was set to 50 min arc. The height distributions have been clipped at -10 and 10cm for clarity as the distribution frequency values drop to (and remain at) zero beyond these values. (b) Shows the peak in height distributions across distance for varying levels of gaze angle noise (10 to 90 min arc).

Arguably a more appropriate measure of noise to use in modelling the estimation of object height would be variability in size judgements. The variability in size judgements is typically much smaller than the variability in gaze direction that was used in the modelling presented here (Brenner & Smeets, 2000, Howard & Rogers, 2002). It is therefore possible that the level of noise used in the modelling overestimated that which would be present in the estimation of object height in the task modelled. Given that no biasing effect of noise was found in the Monte Carlo simulations with a potentially overestimated level of noise, this indicates that errors in estimating object height would play little role in producing observed distortions in the perception three-dimensional shape (Bradshaw et al., 2000, Johnston, 1991).

3.2.3 *Monte Carlo simulation summary*

In summary the Monte Carlo simulations suggest that noise in the vergence signal could play an important role in the aetiology of systematic misestimation of distance from vergence and shape from disparity. Specifically, the most likely distance to have produced a noisy vergence signal is a progressive underestimation of the true object distance as viewing distance increases. This underestimation is due to the distribution of distance values consistent with a noisy vergence signal becoming progressively more skewed with increasing distance. This bias increases as the level of vergence noise increases. Increased vergence noise results in increased bias. Near distances are estimated reasonably accurately with most levels of simulated noise, but with higher levels of noise these too are underestimated, although to a much lesser extent than far distances. The underestimation of far distances is consistent with the experimental literature on distance estimation from vergence (e.g. Viguier et al., 2001).

The simulations also show that if the distance values produced from a noisy vergence signal are used to scale disparity or angular size, both depth and height are progressively underestimated with increasing object distance. This underestimation is greater for object depth, which results in depth to height ratios progressively less than one as object distance increases. The underestimation of depth intervals relative to height intervals at far distances is consistent with previous research on shape constancy, however, noise in the vergence signal did not predict the overestimation of depth intervals relative to height/width intervals at near distances (Bradshaw et al., 1996, Johnston, 1991). The role of noise in disparity and gaze angle signals was also investigated. In contrast to vergence noise, the simulations showed that noise in either disparity or angular size signals did not predict any systematic biases in the perception of object depth or height, just variability around their true values. A comprehensive data set was collected using the same 3-D shape judgement task to that modelled for comparison with the results of the Monte Carlo simulations.

3.3 Psychophysics: The triangle task

3.3.1 Methods

The task was similar to that used by Bradshaw et al. (2000) and consisted of matching the depth of a vertically orientated triangle defined by three points to its height (Figure 3.19). There were three main aims of the psychophysics, (1) determine the distribution of shape settings made across distance in a standard shape judgement task, (2) to gain an understanding of the effect of changing stimulus height on 3-D shape settings and (3) to get an idea of the individual differences in the perception of shape from disparity that any model must account for. Although not predicted by the simulations, stimulus size (height/width) has been shown to affect the estimation of object shape previously (Champion et al., 2004, Collett, Schwarz & Sobel, 1991, Johnston, 1991), this is discussed at greater length below. To the above aims, extensive data sets were collected from two observers, these allowed the distribution of shape settings across distance to be estimated, four additional data sets were obtained with a less extensive data collection regime to investigate the effect of changing stimulus height and individual differences.

3.3.1.1 Apparatus

Stimuli were viewed on a 21in. monitor with a screen resolution of 1024 by 768 pixels, through Stereographics LCD shutter goggles synchronised to the 100 Hz refresh rate of the screen. Across all conditions the viewing distance to the monitor screen was matched to that of the base of the rendered triangle stimulus (Figure 3.9), this minimised effects of conflicting focus cues which are known to effect the perception of 3-D shape from disparity-defined stimuli when the screen distance is different to that of the rendered stimulus (Watt et al., 2005a). Eye height matched the height of the apex dot of the triangle stimuli (Figure 3.9), head movements were minimised with the use of a chin rest.

3.3.1.2 Stimuli

The stimuli were disparity-defined vertically orientated triangles, consisting of three red Gaussian blobs (Figure 3.9). Two blobs defined the 'base' of the triangle, which had a height h; these were rendered at the physical distance of the monitor screen. A third blob (labelled A in Figure 3.9) defined the triangle's apex and could be adjusted along the z dimension to vary the triangle's depth, d. The height of the triangle's base was 2, 4 or 6cm[7]. Stimuli were drawn in red and only the red gun of the monitor was used, as the shutter goggles have minimum crosstalk at longer wavelengths. The blobs had a constant standard deviation of 2.9 arc min across distance in order to eliminate changing blob size as a cue to object distance. Vergence was therefore the predominant cue from which observers could estimate the triangle's distance. Although it is possible that the constant angular size of the blobs could be taken as a cue indicating that the distance of the stimuli is not changing, it was considered important to isolate vergence as a cue to distance in order to be concordant with the modelling. Stimuli were rendered taking into account the interocular distance of the observer and were presented along the median plane (Figure 3.9).

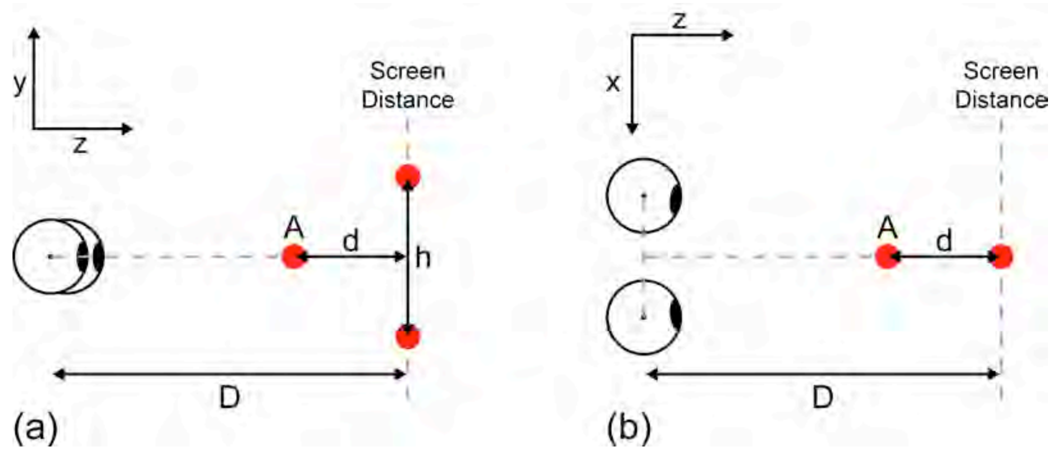

Figure 3.9: Schematic diagram of stimulus presentation seen from (a) the right hand side and (b) from above. Observers viewed a vertically orientated triangle defined by three red Gaussian blobs, positioned at a distance D from their eyes. The task was to match the depth of the triangle d to the height of its base h. This was accomplished by adjusting the position

[7] The term 'size' will be used to refer to the height of the triangles. Both these terms are used in the same way i.e. to refer to the 2, 4 or 6cm height of the triangle stimuli.

of the apex blob of the triangle, labelled A, in depth (z axis) along the median plane (the dashed horizontal grey line in (a) and (b)). The Gaussian blobs forming the base of the triangle were rendered at the same distance as the monitor screen.

3.3.1.3 Observers and training

There were six observers in total; five were experienced psychophysical observers (PS, PBH, HN, RG and VD) and one was a naïve observer (GA), who had limited familiarity with psychophysical research. All had normal or corrected to normal vision and good stereopsis, at least 40 sec arc as measured with the Randot stereo test. Initial screening found that the naïve observer had difficultly fusing the experimental stimuli and experienced diplopia at all viewing distances for all triangle base heights. As a result of this GA participated in two training sessions (30-45 minutes each) prior to data collection. These consisted of tracking the position of the apex blob of the triangle with the eyes as it was moved back and forth in depth (z axis) along the median plane (Figure 3.9). Movement of the apex dot was under control of the observer, who was instructed to move the apex dot from a position fronto-parallel with the triangle's base towards them until they could no longer fuse it. At this point they were instructed to move the apex dot back to fronto-parallel with the triangles base and repeat the process. The stimuli used for training were rendered in the same way as those used in the experiment proper and training took place over the whole distance range, using each of the three triangle base heights. After training the naïve observer reported being able to perceived depth without diplopia and was able to complete the experimental task.

3.3.1.4 Procedure

Observers used the method of adjustment to set the depth of the triangle d to match its height h (Figure 3.9). The depth of the triangle could be varied between plus and minus eight times the height of the triangle's base i.e. plus or minus eight times the geometrically correct depth setting. Prior to starting the experiment the experimental task was clearly explained to the observers with the help of a diagram, this included an explanation of the ‘height’ and ‘depth’ of the triangle stimulus. Observers also had the opportunity to have a few practice trials in which to familiarise themselves with the task if necessary. Observers PS and PBH collected data at six viewing distances, these were 30/40 (PBH/PS), 50, 75, 100, 125 and 139cm. The difference in the shortest distance used for PS and PBH was due to PS being unable to fuse the stimuli at 30cm. The remaining observers collected data at five viewing distances; these were 30, 50, 70, 90 and 110cm, with the exception of observer GA who collected data at 40cm instead of 30cm as the stimuli could not be fused at 30cm. There were therefore 18 viewing conditions for observers PS and PBH (3 base heights x 6 distances) and 15 viewing conditions for the remaining observers (3 base heights x 5 distances).

PS and PBH completed 200 trials for each viewing condition, this allowed a distribution of settings in each viewing condition to be obtained. The remaining observers (with the exception of GA) completed 25 trials for each viewing condition, this allowed distance and base height effects to be looked at across observers. GA was the exception and completed 50 trials for viewing condition; this was because after collecting 25 trials in each viewing condition his data remained quite erratic compared to that obtained from the other observers. GA also expressed a high degree of subjective uncertainty in his settings despite reporting that the stimuli remained fused throughout. A second set of data collection was therefore carried out for this observer. This is discussed further in the results section. Overall the data collection resulted in a total of 3600 trials for PS and PBH, 750 trials for GA and 375 trials for HN, RG and VD.

Trials were blocked by viewing distance and completed in a randomised order. Trials within each distance block were completed in sessions of 75 randomly interleaved trials, 25 for each base height. For observers HN, RG and VD this resulted in one session per distance, for observers PS and PBH eight sessions per distance and for observer GA two sessions per distance. All sessions were completed at one distance before moving onto the next, although because observer GA carried out two data collection runs he was the only observer to repeat data collection at each distance. Within a session trials were self paced, observers pressed the space bar of a computer keyboard to trigger the presentation of a triangle stimulus and adjusted its depth using the up and down arrow keys. The initial depth of each triangle stimulus took a random value between zero i.e. fronto-parallel with the blobs defining the base of the triangle, and four times the height of the triangle's base. When their adjustment was complete observers pressed space once to remove the stimulus from the screen and again for presentation the next stimulus. Observers were encouraged to take as long as needed to make their settings and took breaks whenever needed to avoid dark adaptation and to take a rest.

3.3.2 Results

3.3.2.1 Distributions of depth settings: Observers PBH and PS

Figure 3.10 shows the distribution of shape settings for observers PBH and PS. Histograms were made by binning observer's settings (200 per distance by height combination) into bins spanning 5% steps between zero and the observers maximum depth to height ratio setting (across that observers whole data set) rounded up to the next largest 0.1. For both observers the distributions are reasonably symmetrical and the bias in perceived 3-D shape is characterised as a shift in the distribution of shape settings along the abscissa rather than by a skew in the distributions of shape settings. The variance of the distributions increases with increasing viewing distance for all base heights, as would be expected from noise having a greater effect on observer's settings as distance increases. The distributions of observer PBH have a much greater variability compared to those of PS (note the increased range on the abscissa and decreased range on the ordinate), observer PBH also shows a greater level of bias in perceived shape (Figure 3.11). This is indicative of a link between the level of noise in signal estimates and the level of bias in the perception of 3-D shape, as is suggested by the Monte Carlo simulations.

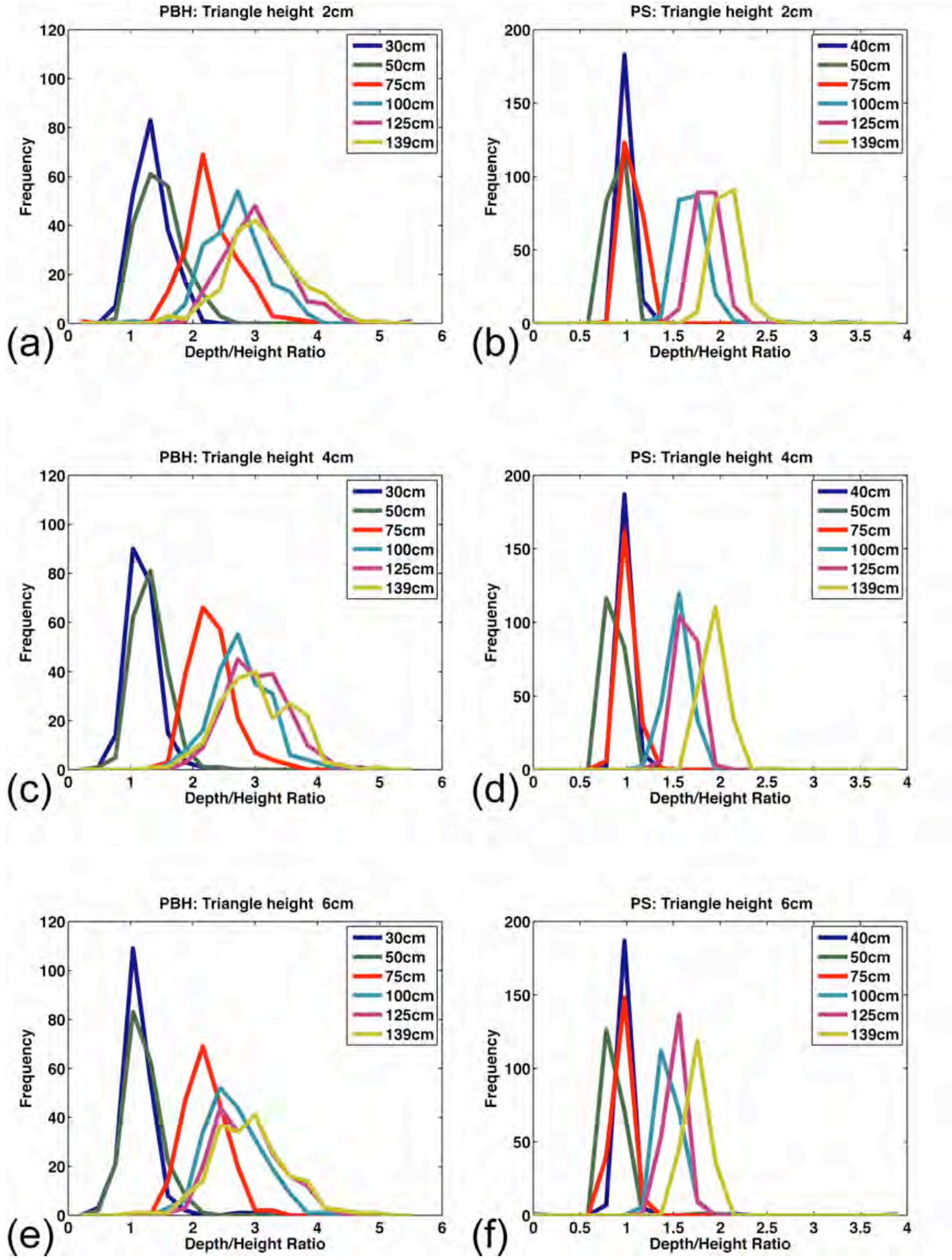

Figure 3.10: The distribution of depth to height ratio settings made by observers PBH (right column) and PS (left column), for the 2cm (top row), 4cm (middle row) and 6cm (bottom row) height triangles. Note the difference in scale for on the ordinate and abscissa axes between observers.

3.3.2.2 Inferential statistics

Individual observer's mean depth to height ratios and scaling distances are plotted in Figures 3.11 and 3.12. Figure 3.11 shows the data from observers PBH and PS who completed the more extensive data collection procedure (as shown in Figure 3.10) and Figure 3.12 shows the data from the remaining observers who completed the standard data collection procedure. Scaling distances shown in these figures represent the distance at which a stimulus with the observer's mean depth to height ratio setting would need to be presented in order to project disparity at the retina corresponding to a triangle with a physical depth to height ratio of one (see Appendix A4). Large individual differences can be seen when observer's data are plotted separately (Figures 3.11 and 3.12), this is common finding in 3-D shape judgement tasks (Todd & Norman, 2003). Data were therefore analysed separately for each observer.

All data were log transformed before analysis to reduced the effects of non-homogeneity of variance and non-normality. Individual data were entered into separate univariate ANOVA with distance and triangle height as fixed factors. Main effects and interaction statistics are shown in Table 3.1. Significant main effects and interactions were followed up with pairwise comparisons, where quoted in the text these comparisons are significant at the p<0.05 level. The data for observer GA showed clear differences between the two experimental runs (Figure 3.13). An initial ANOVA with experimental run entered as an additional fixed factor showed that experimental run interacted with the effect of both distance ($F_{(4,\ 720)}$ = 48.89, p <0.001) and triangle height ($F_{(2,\ 720)}$ = 3.71, p <0.05). The three-way interaction between experimental run, distance and height was non-significant ($F_{(8,\ 720)}$ = 1.45, *ns*). As a consequence of this GA's data were also analysed separately for each experimental run. Initial discussion will focus on GA's overall data as similar trends are present in both experimental runs, the differences between experimental runs is discussed in a separate section.

3.3.2.3 Effects of viewing distance on set 3-D shape

A clear effect of distance on 3-D shape can be seen across all observers (Figures 3.11 and 3.12), a triangle with a perceived depth to height ratio of one has progressively more depth set in it as the viewing distance increases, this is true for all stimulus sizes. If observers had correct estimates of disparity and angular size and were scaling these with a geometrically correct estimate of distance, depth to height ratios of one would be expected for each distance and size of triangle. This is indicated by the horizontal dashed line in the depth to height ratio plots. If all error in set shape is accountable for by the use of an incorrect distance estimate used to scale disparity, a depth to height ratio less than one is consistent with observers scaling disparity with an overestimate of object distance, and a depth to height ratio greater than one is consistent with observers scaling disparity with an underestimate of object distance.

Observers HN, PS and VD show a pattern of settings consistent with the underestimation of far distances and the overestimation of near distances. This is much clearer for observer HN compared to observers PS and VD. For PS settings are only consistent with the overestimation of distance at the 50cm distance, whereas for VD settings are only consistent with the overestimation of distance at the nearest 30cm viewing distance. Observers PBH and GA set more depth than needed across the whole distance range, consistent with them underestimating all distances. Settings become more accurate at nearer distances, but for observer GA these near distance settings remain far from veridical. Both observer's scaling distances are indicative of them taking very little account of changing object distance when making their shape settings (Figures 3.11 and 3.12). In contrast to the results of all other observers, RG makes settings consistent with the overestimation of all but the furthest 110cm viewing distance, at which settings are near accurate (Figure 3.12). Statistical analysis of observer's data confirms these findings.

All observers showed a main effect of object distance (Table 3.1), this was investigated further for each observer with pairwise comparisons of set shape across distance, these confirm that depth to height ratios increase significantly with increasing object distance and are detailed below. For HN all pairwise comparisons were significant and consistent with an increase in depth to height ratio with increasing distance, except for the furthest two distances (90 and 110cm) where depth to height ratios significantly decreased at the furthest distance (Figure 3.12). For PBH all pairwise comparisons were significant and consistent with an increase in depth to height ratio with distance, except the two furthest distances (125 and 139cm), which do not differ. For PS all pairwise comparisons were significant and consistent with an increase in depth to height ratio with distance, except for the nearest 40cm distance at which depth to height ratio increases and is not significantly different to that found at 50cm (Figure 3.11).

For observer GA all pairwise comparisons were significant and consistent with an increase in depth to height ratio with increasing distance except for the nearest 40cm distance where depth to height ratio increases relative to that found at 50cm (Figure 3.12). For RG all pairwise comparisons were significant and consistent with an increase with depth to height ratio with distance except for the 70 and 90cm distances, which do not differ (Figure 3.12). For VD all pairwise comparisons were significant and consistent with an increase in depth to height ratio with distance except for the furthest two distances (90 and 110cm), which do not differ (Figure 3.12). Overall, across observers the pairwise comparisons confirm that increased depth to height ratios are found with increasing distance, there were no trends across observers as to where non-significant pairwise comparisons were found e.g. at far or near distances.

The main effect of distance was moderated by a significant distance by size interaction for all observers bar VD (Table 3.1). It is clear from the data that similar effects of distance were found for each size of triangle and that it is differences in the nature of this effect that accounts for the significant distance by size interaction i.e. some triangle sizes show a larger effect of distance than others (Figures 3.11 and 3.12). This interaction is therefore discussed in the next section on the effects of triangle size on set depth to height ratio.

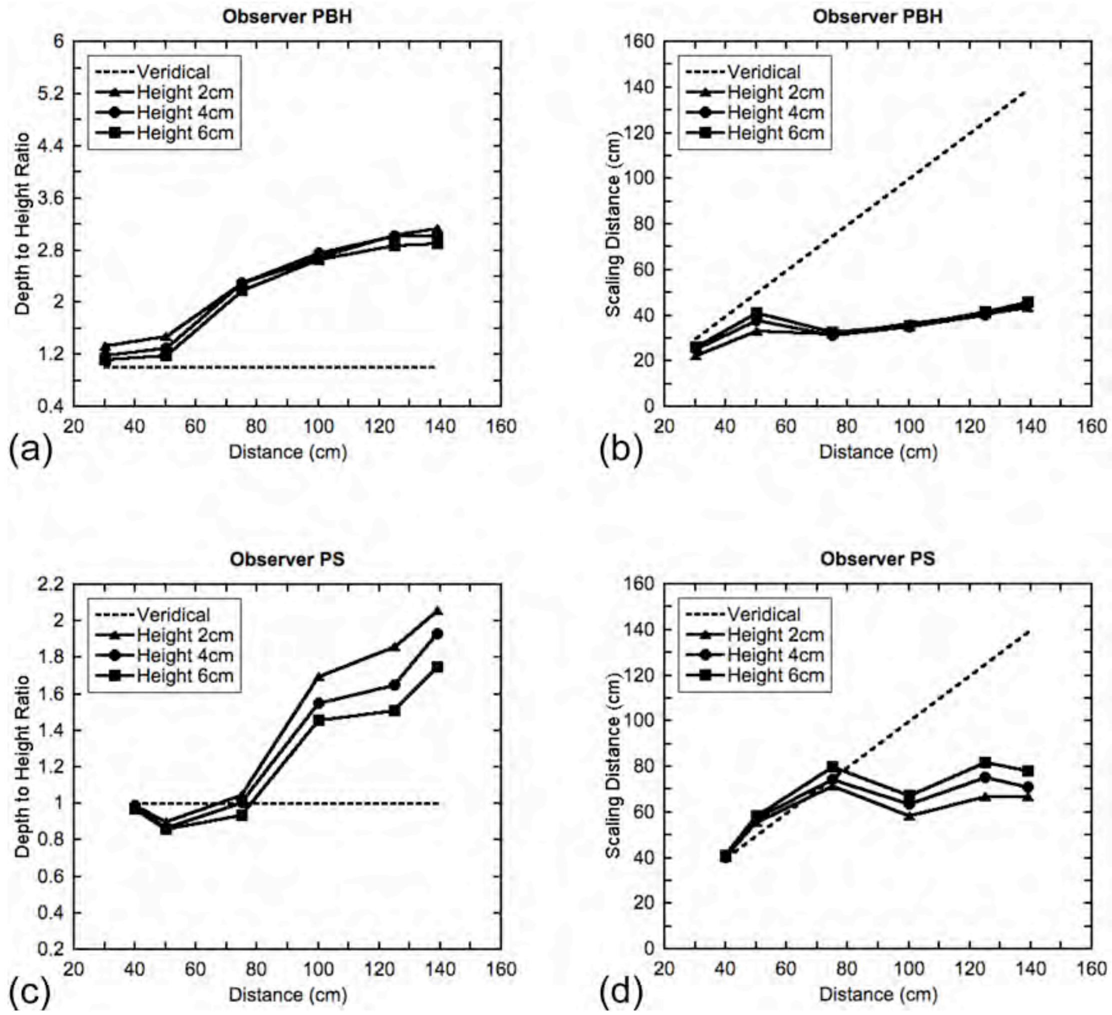

Figure 3.11: Mean set depth to height ratio and scaling distances for observers PBH (a, b) and PS (c, d). Depth to height ratios are shown in the left column (a, c) and scaling distances in the right column (b, d). Error bars show standard error of the mean, these are typically smaller than the symbols. Note the different scale on the ordinate axis for PBH and PS in the depth to height ratio plots.

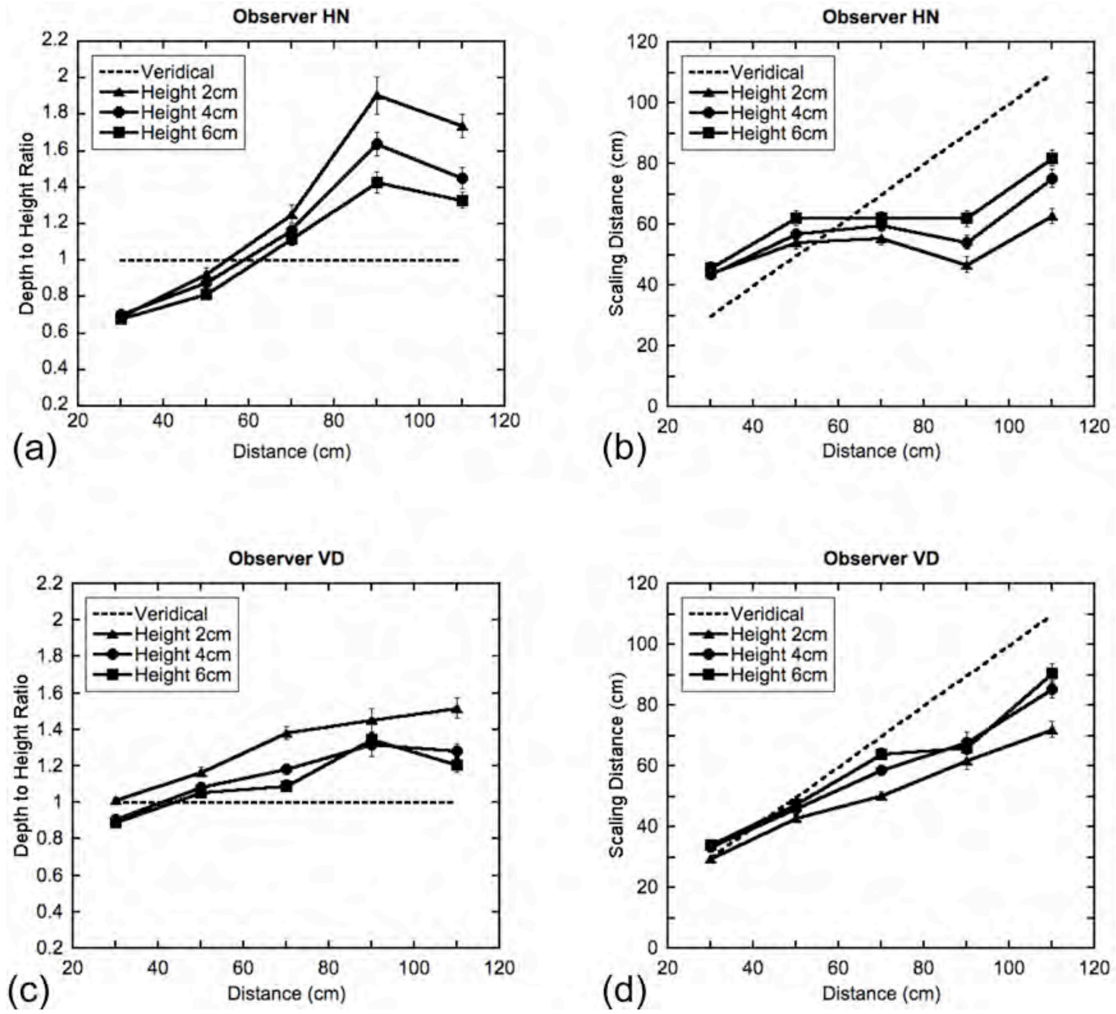

Figure 3.12: Mean set depth to height ratio and scaling distances for observers HN (a, b) and VD (c, d). Depth to height ratios are shown in the left column (a, c) and scaling distances in the right column (b, d). Error bars show standard error of the mean.

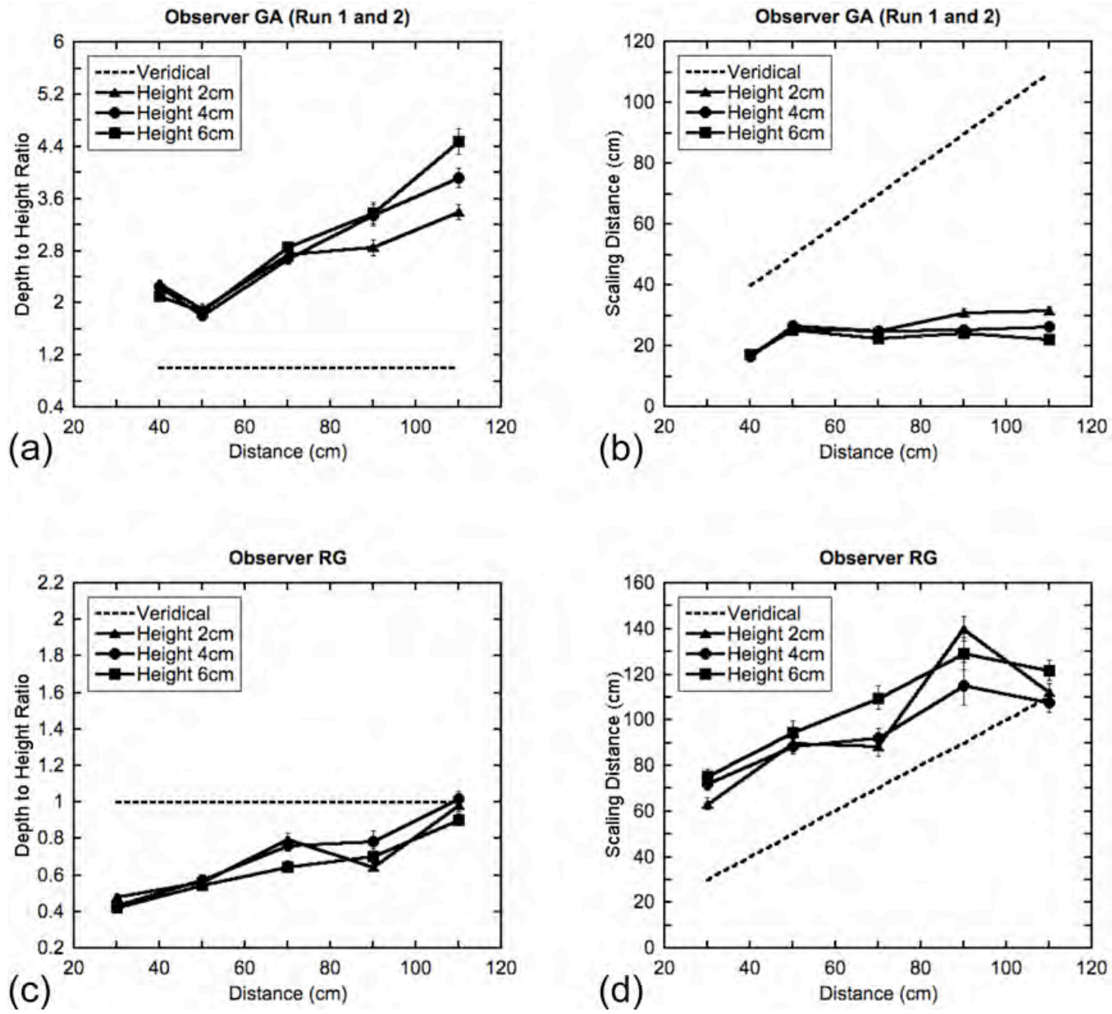

Figure 3.12 (continued): Mean set depth to height ratio and scaling distances for observers GA (a, b) and RG (c, d). Depth to height ratios are shown in the left column (a, c) and scaling distances in the right column (b, d). Data for GA show combined data for both data collection sessions. Error bars show standard error of the mean. Note the different scale on the ordinate axis for GA and RG for the depth to height ratio plots.

3.3.2.4 Effects of triangle size on set 3-D shape

All observers showed a main effect of triangle height on shape settings (Table 3.1), for five out of six observers (HN, PBH, PS, RG, VD) greater depth to height ratios were set as triangle size decreased (Figures 3.11 and 3.12). Observer GA showed the opposite pattern of results, that is increased depth to height ratios with increasing triangle size (Figures 3.12). For all observers, with the exception of VD, the effect of height was moderated by a significant height by distance interaction (Table 3.1), these interactions were investigated further using pairwise comparisons. Observer VD, who showed no interaction, set significantly greater depth to height ratios for the 2cm triangles compared to both the 4cm and 6cm triangles across the whole distance range, depth to height ratios for the 4 and 6cm triangles did not differ (Figure 3.12).

For observers HN, PS, VD and GA the effect of triangle size on set depth to height ratios tended to increase with distance (Figures 3.11 and 3.12). For observer PS pairwise comparisons showed that across all distances, bar the nearest 40cm distance, progressively larger depth to height ratios were set as triangle size decreased. At the 40cm distance only the 2cm and 6cm height triangles differed from one another, with a greater depth to height ratio being set in the 2cm triangle. For observer HN larger depth to height ratios were set with decreasing triangle size at the furthest two distances (90 and 110cm). No effect of triangle height was seen at the 30 or 70cm distances, and only the 2 and 6cm heights differ at the 50 cm distance, with a greater depth to height ratio being set with the 2cm height stimuli.

Observer PBH set greater depth to height ratios as triangle height decreased at the nearest two distances (30 and 50cm), but this effect became less clear at the remaining distances (Figure 3.11). At 75cm, only the 4 and 6cm heights differ with a greater depth to height ratio being set with the 4cm height triangle, no effect of height was found at the 100cm distance. At 125cm a smaller depth to height ratio is set with 6cm triangles compared to either the 2 or 4cm heights, which do not differ. Finally, at the furthest 139cm distance, only the 2 and 6cm distances differ with greater depth to height ratios being set with the 2cm height.

Observer RG showed no consistent effect of triangle size, no effect of size was found at the 30, 50 and 110cm distances. At the 70cm distance a decreased depth to height ratio is set with the 6cm size triangles compared to the 2 and 4cm size triangles, which do not differ. At the 90cm distance only the 2 and 4cm heights differ, with a greater depth to height ratio being set with the 4cm triangle. Finally, for observer GA there was an effect of size but only at the two furthest distances (90 and 110cm). This was opposite to that found in all other observers, at 90cm smaller depth to height ratios are set in the 2cm triangles compared to both the 4 and 6cm triangles, which did not differ, and at 110cm progressively smaller depth to height ratios are set with decreasing triangle height (Figure 3.12)

Observer	**Distance**	**Height**	**Distance x Height**
GA (both sessions)	F(4, 720)=294.29 p<0.001	F=(2, 720)=7.85 p<0.001	F(8, 720)=6.21 p<0.001
HN	F(4, 360)=285.73 p<0.001	F(2, 360)=23.12 p<0.001	F(8, 360)=2.24 p<0.05
PBH	F(5, 3582)=2730.63 p<0.001	F(2, 3582)=74.01 p<0.001	F(10, 3582)=9.81 p<0.001
PS	F(5, 3582)=8587.13 p<0.001	F(2, 3582)=556.83 p<0.001	F(10, 3582)=33.54 p<0.001
RG	F(4, 360)=140.40 p<0.001	F(2, 360)=6.99 p<0.001	F(8, 360)=2.05 p<0.05
VD	F(4, 360)=58.17 p<0.001	F(2, 360)=26.58 p<0.001	F(8, 360)1.53 ns
GA (session 1)	F(4, 360)=83.59 p<0.001	F(2, 360)=0.51 ns	F(8, 360)=2.17 p<0.05
GA (session 2)	F(4, 360)=489.73 p<0.001	F(2, 360)=24.85 p<0.001	F(8, 360)=9.80 p<0.001

Table 3.1: Showing the main effects and interaction statistic for all observers. Cell shading denotes the significance level, dark grey indicates at least $p<0.01$, light grey indicates $p<0.05$ and clear cells indicate non-significance.

3.3.2.5 The effect of experimental run in observer GA

As was discussed above, observer GA's data differed significantly between experimental runs (Figure 3.13). A main effect of distance is found in both sessions (Table 3.1), but this was more consistent and of greater magnitude in the second experimental run. In the first experimental run, depth to height ratios at 40, 70 and 90cm do not differ from one another. But, set depth to height ratio is greater at the 110 and 50cm distances compared to all others. In the second experimental run pairwise comparisons showed that depth to height ratios increase with distance except for at the nearest two distances, which do not differ. A main effect of triangle size was found in only the second experimental run, but a size by distance interaction was found in both runs (Table 3.1).

In the first experimental run the only effect of size was at the nearest 40cm distance where greater depth to height ratios were set in the 2cm compared to 6cm triangles. In the second experimental run the effect of size was more pronounced, at the furthest 110cm distance progressively greater depth to height ratios were set as triangle size increased. At 90cm smaller depth to height ratios were set in the 2cm triangles compared to the 4 and 6cm triangles, which did not differ. At 70cm only the 2 and 6cm triangles differ, with a larger depth to height ratio being set in the 6cm triangles. Observer GA therefore showed a much greater effect of distance and triangle height in the second experimental run. This is evident in the range of depth to height ratios set across distance, in the first run this was $\approx$1.5-3.5 whereas in the second run this was $\approx$2.2-5.4, this was the greatest level of bias found across observers.

Naïve observers have been shown to be more affected by manipulations of the viewing conditions and the range of available disparities in a shape constancy task, and show a greater level of bias (Glennerster et al., 1998). But it is unclear why observer GA should make such different settings in the second run compared to the first. GA was more confident in completing the experimental task in the second run, but practice effects normally act to improve performance rather than result in greater bias, as is the case with GA. It is possible that GA was better able to fuse the stimuli and therefore was able to make greater depth settings in the second experimental run, but this is inconsistent with GA's report of being able to fuse the stimuli and complete the task in the first experimental run.

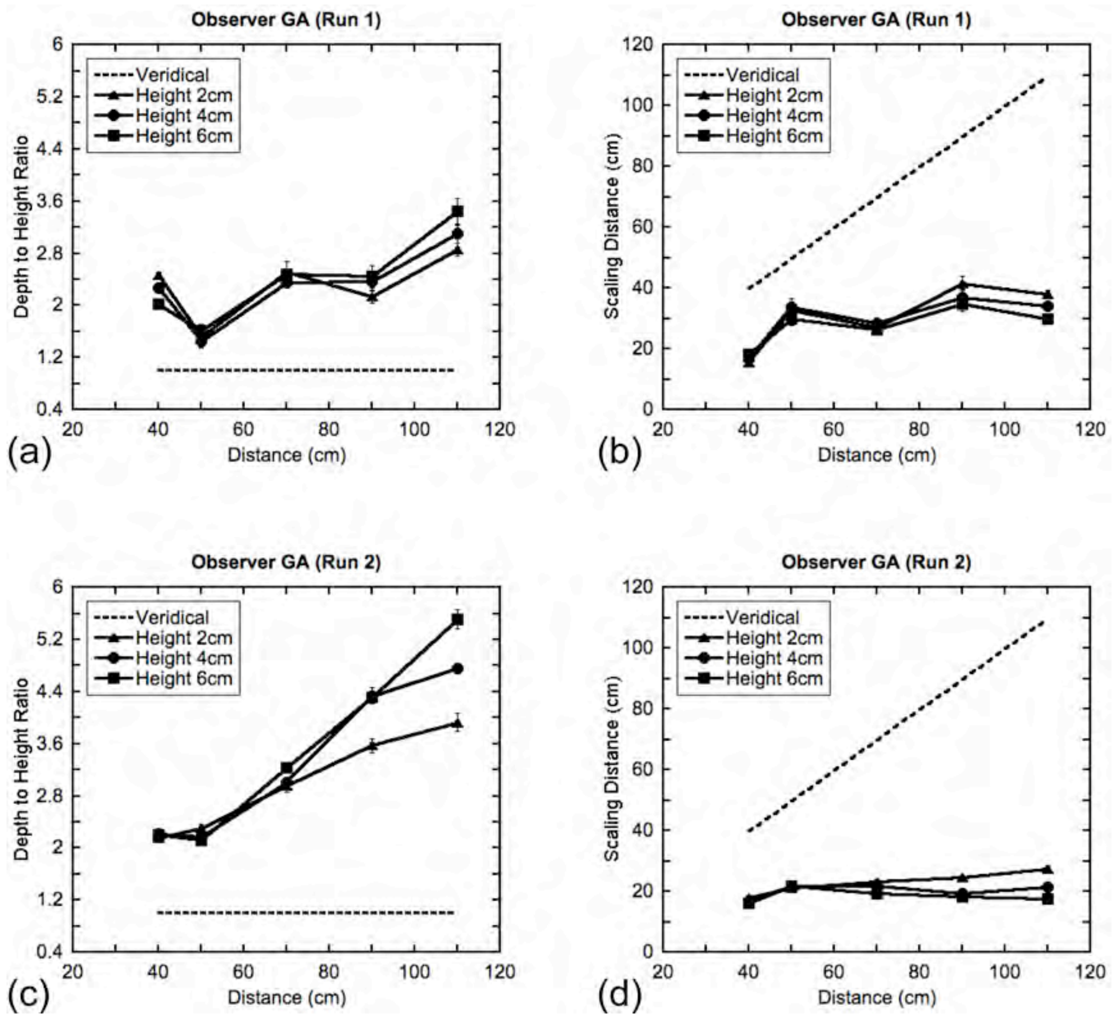

Figure 3.13: Mean set depth to height ratio and scaling distances for observer GA broken down over the two experimental sessions. Depth to height ratios are shown in the left column (a, c) and scaling distances in the right column (b, d). Session one is shown on the top row (a, b) and session two on the bottom row (c, d). Error bars show standard error of the mean.

3.3.3 Discussion

The experimental data demonstrate the typical pattern of results found when observers are required to set the 3-D shape of a disparity-defined object at different distances. For all observers a triangle with a perceived depth to height ratio of one had progressively more depth set in it as the viewing distance increased. For all observers bar RG this resulted in depth to height ratios greater than one at far distances, some also showed depth to height ratios less than one at near distances, whereas for others settings were close to veridical. By contrast observer RG made depth to height ratio settings less than one across the whole distance range except for the furthest distance, where depth to height ratios were close to one. Four out of the six observers (PBH, PS, HN and VD) also showed an effect of stimulus height, this was broadly consistent with depth to height ratios increasing as stimulus height decreased, although the presence and magnitude of this effect was dependent on the viewing distance. Out of the remaining two observers RG showed no effect of stimulus height and GA the opposite effect of stimulus height. These data are discussed more thoroughly in relation to the Monte Carlo simulations below.

3.4 Overall discussion: Modelling and the psychophysics

3.4.1 Comparing the Monte Carlo simulations and the psychophysical data

To get an idea of the level of vergence noise that would be needed to account for the bias found in observers shape settings, inverse shape settings (1/depth to height ratio) were calculated for two representative observers (HN and PBH) (Figure 3.14). Inverse shape settings represent the level of depth underestimation that would be required in order that observer's depth to height ratio settings corresponded to a triangle with a depth to height ratio of one. These inverse shape settings can be compared to the results from the Monte Carlo simulations (compare Figure 3.14 with Figure 3.6) to assess the level of vergence noise that would be required to account for the observer's perceptual bias. Comparison of Figure 3.14b and 3.6 suggests that the data from observer PBH can be accounted for well by the vergence noise model, the level of noise required to account the bias in this observers shape settings would be around 80-90 min arc (in half the vergence angle). This level of noise predicts the observer's perceptual bias well across the whole distance range.

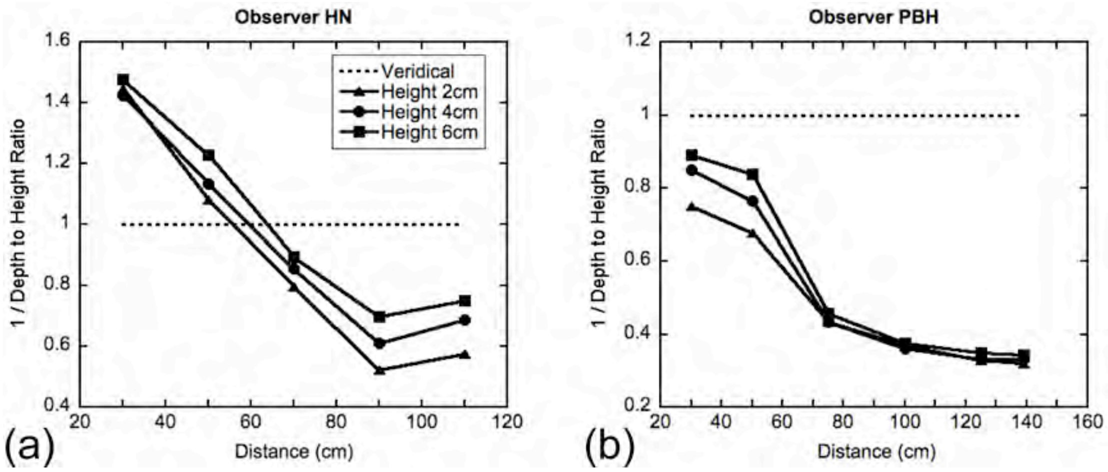

Figure 3.14: Inverse shape settings (1/depth to height ratio) for observers HN (a) and PBH (b). The key is the same for both observers. The Horizontal line in both graphs represents veridical settings. Note the difference in scale on the ordinate and abscissa for the two observers.

For observer HN the model can account less well for the data. Firstly the model does not predict a triangle with a depth to height ratio of one to be perceived as having a depth to height ratio greater than one at near distances, as seen in this observer's data. This is because the model does not predict the overestimation of distance from vergence at any viewing distance. Whilst the overestimation of depth intervals relative to height intervals at near distances is not found in all observers, it is a consistent feature of these types of task (Glennerster et al., 1996, Johnston, 1991, Johnston, Cumming & Parker, 1993). Furthermore, in the present study observer RG made settings consistent with the overestimation of depth relative to height at all bar one distance. The model detailed above cannot account for this observer's data. The maximum bias shown by HN is at the 110cm stimulus distance, to account for the level of bias found here would require around 70 min arc noise in half vergence. However, this level of noise does not account well for this observer's data across the rest of the distance range. In particular, inverse shape settings seem to increase to larger depth to height ratios with decreasing distance too rapidly to be accounted for well by the model. This results in inverse depth to height ratio settings being greater than one at near distances.

It is difficult to gauge whether the level of noise required in the vergence signal to predict the magnitude an observers perceptual bias is realistic. This is because there is no clear way to estimate of the level of noise in the vergence signal accurately at present. Estimates of the amount of noise in the vergence signal are out of necessity indirect, and therefore may be unrepresentative of the true level of noise in the signal. For example, for observer PBH an 80 min arc level of noise would account for the bias in shape settings well. This corresponds to a distance range of 13.5cm (38.2-24.7cm) at 30cm and 8055cm (8125-69.9cm) at 139cm (the nearest and furthest viewing distances for this observer). However, this does not mean that this observer would actually make this range of estimates in a distance estimation task. With this level of noise the most likely distance to have produced this noisy vergence signal at a viewing distance of 139cm is a distance of 65cm. Estimates would therefore cluster around this value, which is more realistic in terms of the level of bias demonstrated in distance estimation tasks (Viguier et al., 2001).

3.4.2 The effect of stimulus height

The present study found effects of stimulus size on depth to height ratios. Across most observers larger depth to height ratios were needed in smaller size stimuli for them to be perceived as having a depth to height ratio of one. This effect was distance dependent, with some evidence suggesting that the effect increased with increasing distance, but not across all observers. Previous studies have found effects of stimulus size on perceived depth and shape. Collett et al. (1991) had observers judge the depth between two surfaces defined by stereo and texture. They varied either distance alone, so the angular size of the stimuli decreased naturally with increasing distance, or both distance and size, so that the angular size remained constant across viewing distance. When stimuli remained constant in angular size over distance perceived depth decreased with increasing viewing distance, whereas when the stimuli changed angular size the effect of distance was diminished in most observers.

Johnston (1991) also found an effect of stimulus size in the apparently circular cylinder task. With a far 214cm viewing distance *greater* depth to half-height ratios were required in *larger* sized stimuli for them to be perceived as circular, whereas with a near 53.5cm viewing distance *greater* depth to half-height ratios were needed in *smaller* stimuli to be perceived as circular. At an intermediate 107cm distance there was very little effect of stimulus size. Although there is some correspondence in the direction of the effect of stimulus size between the data presented here and that of Johnston (1991), the interaction between size and distance seems to differ between the studies, so the correspondence is not all that convincing.

Champion et al. (2004) investigated the effect of surface shape and stimulus size on the perceived 3-D shape of objects presented at 80cm. They found mixed results across observers and different surface shapes. There seemed to be a general trend for greater depth to be needed with larger stimuli for them to be perceived as having a depth equal to their half-height, but the reverse effect was seen with some observers for some surface shapes. Overall, there is little consistency across studies as to the effect of stimulus size on the perception of 3-D shape. The simulations detailed above do not predict any differences in the perception of shape for differently sized stimuli. At present it therefore remains unclear whether (a) consistent effects of stimulus size can be found in shape perception tasks, and whether (b) the effects of stimulus size can be predicted from any theoretical basis.

3.4.3 Concluding remarks

This chapter presented the results of simulations run to investigate the role of noise in vergence, disparity and gaze angle signals in the aetiology of the misestimation of object distance and 3-D shape. By assuming that observers estimate the most likely distance, depth and height to have been produced by a noisy vergence signal, in conjunction with estimates of disparity and gaze angle, it was possible to predict both the progressive underestimation of far distances (Viguier et al., 2001) and the progressive underestimation of depth relative to height intervals with increasing distance. The simulations were not however able to predict the overestimation of depth relative to height/width found at near distances in both the present and previous studies (Bradshaw et al., 2000, Glennerster et al., 1998, Johnston, 1991). A review of the experimental literature on the estimation of distance from vergence suggested that vergence could provide accurate information about object distance at near distances, but that far distances are progressively underestimated as the viewing distance increases (Mon-Williams & Tresilian, 1999, Swenson, 1932, Tresilian et al., 1999, Viguier et al., 2001).

This is consistent with the results of the simulations and indicates that the overestimation of depth relative to height/width found at near distances is not a consequence of the overestimation of distance. In contrast to the effects of noise in the vergence signal, noise in both disparity and angular size signals did not predict any bias in perceived shape. The psychophysical data showed some evidence for an effect of stimulus size on perceived shape, but when the effect of stimulus size is looked at across studies there is no clear pattern of results. The simulations did not predict any effect of stimulus size. In summary, vergence noise may be an important determinant of systematic distortions in perceived distance and shape.

Chapter Four

The role of height in the field and a view of an objects upper surface and contour in the perception of shape

4 The role of height in the field and a view of an objects upper surface and contour in the perception of shape

4.1 Introduction

4.1.1 Viewing an object from above

In many instances 3-D shape from binocular visual cues in isolation is systematically misperceived, consistent with the overestimation of near distances and the underestimation of far distances (Bradshaw et al., 2000, Glennerster et al., 1996, Johnston, 1991). Although these studies are valuable in demonstrating the circumstances under which depth constancy fails they do so by removing valuable sources of information that are normally present in the natural viewing situation. For example, we rarely view and interact with objects directly at eye height; in natural tasks such as prehension objects are typically viewed from above and on a supporting surface. This provides improved information from which to estimate 3-D shape; two such sources of information are (1) height in the visual field and (2) a view of an object's upper surface and contour. The poor shape constancy found when observers view disparity-defined stimuli at eye height might therefore be ameliorated if objects are presented in more naturalistic viewing conditions, where additional and potentially more reliable cues to viewing distance and shape are available.

4.1.2 Height in the field as a cue to distance

Height in the field has been recognised as a cue to distance since the time of Euclid (Howard & Rogers, 2002). By standing in an open space and looking at near and far objects on the ground plane it is readily apparent that the further an object is away the higher it appears in the field of view. Objects on the ground plane reach a maximum height within the field of view at the horizon and a minimum height within the field of view directly below the eye. An identical but opposite relationship exists with objects on a flat plane above the eyes running parallel to the ground, here nearer objects appear higher in the visual field reaching a maximum directly above the eyes and further objects appear lower in the visual field reaching a minimum at the horizon. Height in the visual field has also been termed 'vertical gaze angle' (Gardner & Mon-Williams, 2001), 'slope of regard', 'angular declination below the horizon' (Ooi et al., 2001, Ooi, Wu & He, 2006) and 'angular elevation'.

In isolation height in the visual field allows only relative distance judgments to be made. However, it also offers the possibility of estimating absolute distance if additional information is available to the observer and certain plausible assumptions are made. Consider an eye positioned orthogonally at a height H above a flat surface which has an object placed on it at a distance D from the observer (Figure 4.1a). The vertical gaze angle to this object, as measured relative to a line of sight running perpendicular to gravity, is denoted by θ. Given an estimate of eye height $\hat{H}$ and vertical gaze angle $\hat{\theta}$, estimated distance $\hat{D}$ along the surface plane is given by,

$$\hat{D} = \hat{H} \cot \hat{\theta} \tag{4.1}$$

If the surface plane is not flat, but has a known constant slope δ (Figure 4.1b) and an observer has an estimate of this slope, $\hat{\delta}$, distance along the surface plane can be estimated by (Ooi et al., 2006),

$$\hat{D} = \frac{\hat{H} \cos \hat{\theta}}{\sin\left(\hat{\theta} + \hat{\delta}\right)} \tag{4.2}$$

For the visual system to make use of these geometric relationships to estimate object distance it must have access to accurate estimates of eye height, the angle of gaze and the slope of the supporting surface (Gibson, 1950, Howard & Rogers, 2002, Ooi et al., 2006). It must also make the assumption that the ground plane is opaque and that viewed targets are resting upon it (Cutting & Vishton, 1995). Gardner and Mon-

Williams (2001) added to this list that the nervous system should also have knowledge of ocular position with respect to the head and head orientation with respect to the shoulders. This seems implicit in assuming that the visual system can make use of gaze angle, vergence or eye height information at all. Cutting and Vishton (1995) suggested that the supporting surface needed to be planar and orthogonal to gravity for the use of height in the field. However, if observers have accurate knowledge of the constant slope of the supporting surface this is not necessary (Equation 4.2).

Evidence suggests that observers are accurate in judging their angle of gaze relative to a supporting surface (Ooi et al., 2006), but tend to underestimate their eye height in both light and dark conditions (Ooi et al., 2001, Sinai et al., 1998, Stoper & Cohen, 1986). This underestimation is much reduced in the light (0.29 degrees) compared to the dark (2.79 degrees) (Stoper & Cohen, 1986). Studies investigating the accuracy of eye height judgements have tended to use an adjustment method whereby the eye was moved relative to the target or vice versa. The observer indicated to stop this movement when the eye and target were at the same height from the supporting surface. It is possible that the underestimation of eye height could partly be due to observers pre-empting when to stop the movement to avoid overshooting their estimate (see Stoper & Cohen, 1986).

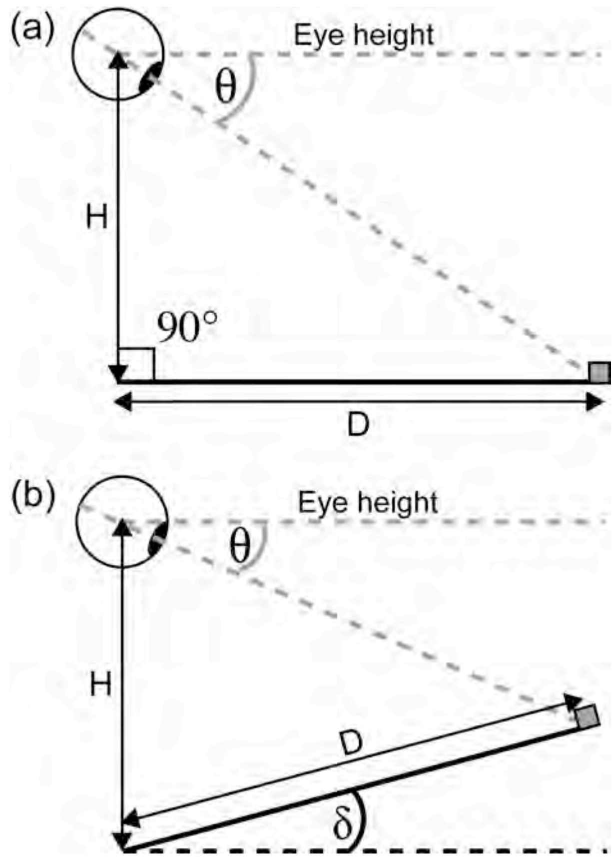

Figure 4.1: A schematic diagram showing how vertical gaze angle can be used to estimate absolute distance on a flat or slanted surface. An eye is positioned at a height H above a supporting surface which has an object placed on it at a distance D from the observer, the vertical gaze angle to this object, relative to a line of sight running perpendicular to gravity, is denoted by θ. In (a) the ground plane is flat and orthogonal to gravity, whereas in (b) the ground plane has a constant slope δ. For further details see the accompanying text.

The assumption that objects are resting upon an opaque ground plane or supporting surface is a plausible inference to make about the world and under monocular viewing when the perceived distance of an object is ambiguous observers have been shown to apply such an assumption (Gibson, 1950, Meng & Sedgwick, 2001). To a first approximation many ground planes and surfaces that humans interact with are relatively flat. There are also numerous cues in a natural viewing situation from which an observer could estimate the slant of the supporting surface, for example, binocular disparity, perspective and texture gradient. Studies have shown that observers optimally weight visual cues of texture and disparity in the estimation of surface slant over a range of viewing distances ($\approx$19-171cm) (Hillis et al., 2004). Optimal estimation does not necessarily result in veridical perception, but in a cues-consistent environment where multiple cues such as texture, disparity, perspective and haptics are available the estimated slant could be reasonably accurate.

As with vergence as a cue to distance, vertical gaze angle will become a progressively unreliable cue as object distance increases; this is because any error in its measurement will correspond to a progressively greater distance range at larger viewing distances (Mon-Williams & Tresilian, 1999). In most natural situations the baseline for triangulation with vertical gaze angle (eye height from the supporting surface) will be much greater compared to the baseline for triangulation with vergence (the interocular distance). Angular resolution for changes in eye direction has also been shown to be similar to angular resolution for changes in vergence (Brenner & Smeets, 2000). This suggests that vertical gaze angle may be a more reliable cue to distance than vergence, because for any given distance the vertical gaze angle will be larger than the vergence angle and therefore affected less by the same level of angular noise. If this is the case vertical gaze angle may be weighted more highly when combined with other cues such as vergence in the estimation of object distance (Jacobs, 2002, Knill & Richards, 1996, Landy et al., 1995).

4.1.3 Effects of manipulating height in the field

Attempts have been made to selectively manipulate vertical gaze angle to gauge its role in the estimation of object distance. Ooi et al. (2001) manipulated the vertical gaze angle of a standing observer in an open field situation using base-up and base-down ophthalmic prisms. Wearing base-up prisms, which act to increase θ, was shown to decrease the perceived distance of an object on the ground plane and adaptation to these prisms increased perceived distance to the object. Ooi et al. interpreted the results in terms of a change in perceived eye height, in this interpretation perceived eye height is increased when wearing base-up prisms, which causes an increase in θ and therefore a decrease in perceived distance. Similarly, after prism adaptation, which causes a decrease in θ perceived distance is increased (Ooi et al., 2001). It is also possible that the causality is reversed i.e. a change in θ causes a change in perceived eye height.

Gardner and Mon-Williams (2001) have shown that height in the field influences the reaching and positioning movements of an observer seated at a table surface. Observers carried out a visuo-motor positioning task and a prehension task, whilst wearing either base-up prisms, which increase θ, base-down prisms, which decrease θ, or no prisms. The positioning task required observers to place a pin in the underside of a table surface directly below a target viewed within arm's reach. The prehension task required observers to reach out and grasp, but not pick up, an object. The prism manipulation had the predicted effect, visuo-motor positioning was reasonably accurate with no prisms, but overshot the target with base down prisms, and undershot the target with base up prisms. In the prehension task peak velocity, which has been shown to increase linearly with object distance (Jeannerod, 1988), was greater with base-down prisms and smaller with base-up prisms, when compared to reaches with no prisms. This indicates an increase and decrease in perceived distance respectively.

An issue related to using prisms to manipulate height in the field is that the prisms also act to alter the perceived slant of the support surface. Surfaces slant upwards with base-down prisms and downwards with base-up prisms (Gardner & Mon-Williams, 2001). If the ground plane is perceived to be slanting, height in the field will signal a different object distance to when the surface is perceived to be flat (as can be seen from Figure 4.1 and by comparing Equations 4.1 and 4.2). For the same gaze angle, θ, estimated distance, $\hat{D}$, will increase when the supporting surface is perceived to be sloping away from the observer ($\hat{\delta} < 0$), and decrease when the supporting surface is perceived to be sloping toward the observer ($\hat{\delta} > 0$). However, the effect of a prism-induced change in slope on distance estimates is opposite to that demonstrated with the use of prisms for the manipulation of vertical gaze angle.

Studies using prisms to manipulate height in the field in the natural open field setting have not highlighted or reported a change in perceived slope of the supporting surface (Ooi et al., 2001). This suggests that the effect of manipulating vertical gaze angle on the perception of slope could be restricted to near distances (Gardner & Mon-Williams, 2001). A prism-altered estimate of slope from vertical gaze angle will conflict with other cues to slope such as the texture gradient, which will continue to signal the true slope of the supporting surface. It is possible that the effect of vertical gaze angle on perceived slope might be reduced at far distances because whereas the reliability of vertical gaze angle is expected to decrease with distance (Gardner & Mon-Williams, 2001) the reliability of texture remains approximately constant (Hillis et al., 2004). This means that the weight given to vertical gaze angle will decrease relative to the weight given other cues as the viewing distance increases. However, manipulations of vertical gaze in a full cue open field setting have been shown to influence the perception of far distances, this suggests that the effect of vertical gaze angle on perceived slope might be less reliable or robust than its direct effect on perceived distance.

In addition to changing perceived slope, prisms can also introduce conflict regarding the height of the supporting surface, between visual and haptic information, and curvature distortion (Gardner & Mon-Williams, 2001). This highlights the fact that prisms used to manipulate vergence or vertical gaze angle (Gardner & Mon-Williams, 2001, Tresilian & Mon-Williams, 1999) affects more than just the angle of gaze, they also introduce large conflicts between cues for many perceptual attributes. This makes it difficult to interpret any effect they have unambiguously. In a situation with large cue conflicts complex interactions may occur between cues in the estimation properties such as distance, slant and shape. Under these situations the visual system may act as a robust estimator and veto the information provided by highly discrepant cues (Landy et al., 1995). It is also worth noting that the effects of prisms on perceived distance have also been demonstrated when viewing point light objects in the dark (Ooi et al., 2006). Here many, but not all, of the conflicts associated with the use of prisms will be eliminated.

Under many situations in the natural environment objects are not viewed on a single flat planar surface below eye level. For example, a person standing by an office desk viewing a mug of coffee positioned on a book resting on the desk surface. If an observer is to use height in the field to estimate the distance to the mug of coffee they must take into account the changing relationship between the angle of gaze and distance for each of the relevant surfaces (the office floor, the desk top and the book). These types of relationships have been termed ‘nested contact relations’ and evidence suggests that observers are able to take into account these geometrical relationships when estimating object distance (Meng & Sedgwick, 2001). This suggests that height in the field could be a useful and accurate cue for estimating object distance in more diverse naturalistic circumstances. Somewhat contrary to this is evidence showing that distances are overestimated near dips/drops in the ground surface and underestimated over changes in ground plane texture (Sinai et al., 1998).

Overall, manipulations of height in the field, although complicated by the introduction of cue conflicts, indicate that it is a potentially useful source of information regarding absolute distance from the observer. However, there has been little direct assessment of its role in scaling binocular disparity. Given its potentially improved reliability compared to vergence, height in the field might be weighted more highly in the estimation of object distance during natural viewing (Jacobs, 2002, Landy et al., 1995). It could therefore be a valuable cue to distance for scaling disparity in estimation of 3-D shape. In a viewing situation where height in the field is removed, such as when viewing isolated disparity defined objects at eye height, the visual system would be forced to use potentially less reliable distance information from vergence. Systematic distortions in perceived 3-D shape have been demonstrated under these viewing situations (Johnston, 1991).

4.1.4 Object contour as a cue to object shape

In addition to improved distance information from height in the field, additional information about an object's shape is provided by the projected view of the object's upper surface and boundary contour, and by improved disparity information across a disparity discontinuity. Gillam, Flagg and Finlay (1984) investigated the role of disparity discontinuities in determining perceived slant in stimuli subject to the geometric effect. In the geometric effect a horizontal magnifier placed in front of an eye causes a frontoparallel surface to be perceived as slanted about its vertical axis. Magnification in one eye introduces conflict between binocular cues produced by the magnification and monocular cues such as perspective, with the result that the perceived slant of the surface is less than that predicted by the magnification. Gilliam et al. presented observers with a regular frontoparallel dotted grid surface. In one condition a horizontal magnifier was placed over the full field of view in one eye, causing the whole surface to appear slanted about its vertical axis. In another condition, the magnifier was placed over just the upper half of the field of view; this caused only the upper half of the surface to appear slanted around its vertical axis and created a disparity discontinuity across the upper and lower half of the stimulus.

In the full field magnification condition slant was less than that predicted from image magnification, as would be expected from the visual system combining information from the conflicting binocular and monocular cues to slant in the stimulus. However, when only the upper half of the image was magnified its slant was much closer to that predicted from the magnification (Gillam et al., 1984). This indicates that relative disparity over the disparity discontinuity provided additional information for the perception of surface slant (van Ee et al., 1999). Observers are also faster to perceive structure in random dot stereograms containing a disparity discontinuity than in full field slant stimuli, or hinge ‘open book’ stimuli containing two abutting slanted surfaces (Gillam, Chambers & Russo, 1988). This suggests that disparity discontinuities are salient image features that can aid fusion and provide relative disparity information for the estimation of surface properties. This may be beneficial as the visual system has increased sensitivity to relative disparity compared to changes in absolute disparity (Howard & Rogers, 2002).

Consider viewing a cylindrical object standing on a table surface within near visual space from above. Viewing the cylinder from above provides a projected view of its upper surface and boundary contour that are not available when the cylinder is viewed at eye height (see the ‘Lidded’ diagram in Figure 4.2). These sources of information can provide the observer with additional information with which to estimate 3-D shape. The aspect ratio of the elliptical projection of the cylinder’s upper surface at the retina can be used to determine the shape of the cylinder if the orientation, distance and height of the cylinder are known. This is because the elliptical projection of the cylinder’s upper surface changes both with the shape of the cylinder itself, and its position and orientation with respect to the observer. Information about the distance and height of the cylinder is available from both vergence and height in the field. Orientation information is provided by the disparity-defined slant of the cylinder’s upper surface and relative disparity information over the disparity discontinuity across the cylinder’s sides and upper surface.

Now consider a situation in which the cylinder is still viewed from above at the same distance and height, but now information from the upper surface of the cylinder is not available to the observer (see the ‘Contoured’ diagram in Figure 4.2). Information about the cylinder’s aspect ratio and boundary contour is still available, but in a different form, as only half the extent of the cylinder’s depth can now be seen. The observer also has the same information available from vergence and height in the field about object distance and height. However, in this situation there is no information about the slant of the cylinder’s upper surface and no relative disparity information across a disparity discontinuity. This may make it more difficult for the observer to determine the orientation and shape of the cylinder compared to when a view of the cylinder’s upper surface was available.

In summary, viewing an object from above can provide valuable information about 3-D shape that is not available when viewing an object at eye height. Eye height viewing is the typical situation in most experiments that have investigated the perception of 3-D shape from binocular visual cues. It is therefore possible that the perceptual distortions of 3-D shape demonstrated in such tasks is in part due to the unnatural viewing situation that observers are subject to. The current study focuses on understanding the role of height in the scene and a view of an object’s upper surface and boundary contour in the estimation of 3-D shape of disparity defined objects. Height in the scene can provide improved information about object distance, from which disparity can be scaled and a view of an objects upper surface and contour can provide improved information about object shape.

4.2 Methods

4.2.1 Observers

Three observers took part in the experiment. All were experienced psychophysical observers, two (PBH and PS) were aware of the purpose of the experiment. All observers had stereoacuity of at least 40 seconds of arc as measured with the Randot stereo test.

4.2.2 Apparatus and monitor calibration

Stimuli were viewed on a 21in. monitor through Stereographics LCD shutter goggles synchronized to the 100 Hz refresh rate of the screen. The monitor had a screen resolution of 1152 $\times$ 864 pixels and was housed on a custom built moving platform that allowed monitor distance to be adjusted. The frame of the monitor housing was covered with matt black card. Stimuli were generated and presented in the Matlab programming environment using the Psychophysics toolbox extensions (Brainard, 1997, Pelli, 1997). The screen was spatially calibrated using a technique similar to that described in Backus, Banks, van Ee and Crowell (1999). A detailed description of the calibration equipment and calibration routine can be found in Appendix A3. In brief, observers were positioned so that the centres of rotation of their eyes were at a known position and distance from the monitor screen and frontoparallel with respect to its surface. In this position the median plane passed through the centre of the monitor screen perpendicular to its surface. This was achieved by using a custom-built bite bar and sighting device. Individuals had their own bite bar, which was used to maintain head position throughout the experiment. During calibration a nylon filament loom was placed in a frame attached to the front of the monitor so that it was frontoparallel to the screen and 1cm from its surface. The small distance between the loom and the monitor screen minimised the effects of conflicting focus cues, which are known to affect the perception of simulated 3-D shape (Watt et al., 2005a).

A grid of Gaussian blobs was presented on the screen and observers aligned individual blobs of this grid with 2cm spaced intersections on the loom. This process was carried out separately for each observer and separately for each eye and for each viewing distance and viewing height used in the experiment i.e. 16 calibrations per person (2 eyes x 4 distances x 2 heights). Two-dimensional second-order polynomial equations were used to fit the x and y coordinates of the centres of the Gaussian blobs (in pixels) to the x and y coordinates of the loom intersection to which they were aligned (in metres). These equations were used to convert between stimulus coordinates and screen coordinates when rendering the experimental stimuli. The plane of the loom thus provides a geometrically correct surface in which to present stimuli to the observer. All calibrations for a given viewing height were carried out before experimental data was collected at that height.

4.2.3 Stimuli

The stimuli were random dot stereograms of vertically orientated elliptical cylinders. The cylinders were 10cm in height and had a fixed width of 6cm. The surfaces of the cylinders were rendered with red Gaussian blobs and were rendered taking into account individual interocular distances. Red was used as the shutter goggles have minimum crosstalk at longer wavelengths. The standard deviation of the Gaussian blobs was 1mm and scaled in accordance with the monitor calibration. The surface of the cylinder was rendered such that regions of its surface could be occluded from an eye depending on that eye's viewpoint of the surface, as would occur naturally when viewing a real object (Gillam & Borsting, 1988). The standard deviations of the blobs in their x and y dimension scaled with the position of the blob on the cylinder's surface. This reflected the projective foreshortening of the blob that would naturally occur given the view of the blob from each eye. Foreshortening was applied uniformly over the blob such that the blob was rendered tangent to the cylinder's curved surface. The stimuli therefore contained consistent textural information about 3-D shape. Typically textural information in stereograms is consistent with a flat frontoparallel surface and therefore conflicts with that of the stereoscopically rendered shape (Zabulis & Backus, 2004), this in itself could act to introduce bias in perceived 3-D shape.

Stimuli were presented at distances of 40, 60, 80 and 100cm centred on the median plane. In each case the monitor was positioned at a distance such that the plane of the calibration loom on which the stimuli were rendered was situated at the simulated object distance. The loom itself was only present during the screen calibration. Cylinders were also presented at two viewing heights depending on the experimental condition, these were eye height and 20cm above the centre of the cylinder i.e. 15cm above the top of the vertically orientated cylinder. There were four viewing conditions which are illustrated in Figure 4.2, these were as follows, (1) ***Lidded***: cylinders were viewed from above and a view of the cylinder's upper surface/lid was provided, this was rendered in the same way as the rest of the cylinder's surface (Figure 4.2a), (2) ***Contoured***: cylinders were viewed from above but the cylinder's upper surface/lid was not rendered (Figure 4.2a), (3) ***Rotated***: cylinders were viewed from above but were rotated around their vertical axis so that the surface of the cylinder was frontoparallel to the observer (Figure 4.2b) and (4) ***ACC***: cylinders were presented at eye height i.e. in a standard apparently circular cylinder (ACC) configuration (Figure 4.2c).

In the ACC condition vergence provides the predominant source of distance information for scaling disparity in the estimation of 3-D shape, this is the standard stimulus configuration in which biases in perceived disparity-defined shape have been demonstrated (e.g. Johnston, 1991). In the Rotated condition vertical gaze angle provides additional information about object distance that could be used to scale disparity, but the region of the cylinder's surface in view is the same as that of the ACC condition. In the contoured condition, in addition to vertical gaze angle, a view of the elliptical contour of the top edge of the cylinder provides information about 3-D shape. Finally, in the lidded condition the upper surface or 'lid' of the cylinder is also rendered. In addition to the information provided in the rotated condition, this condition provides a salient disparity discontinuity (between the cylinders body and lid) and disparity information about the slant of the cylinder's upper surface. both could be used to better estimate 3-D shape.

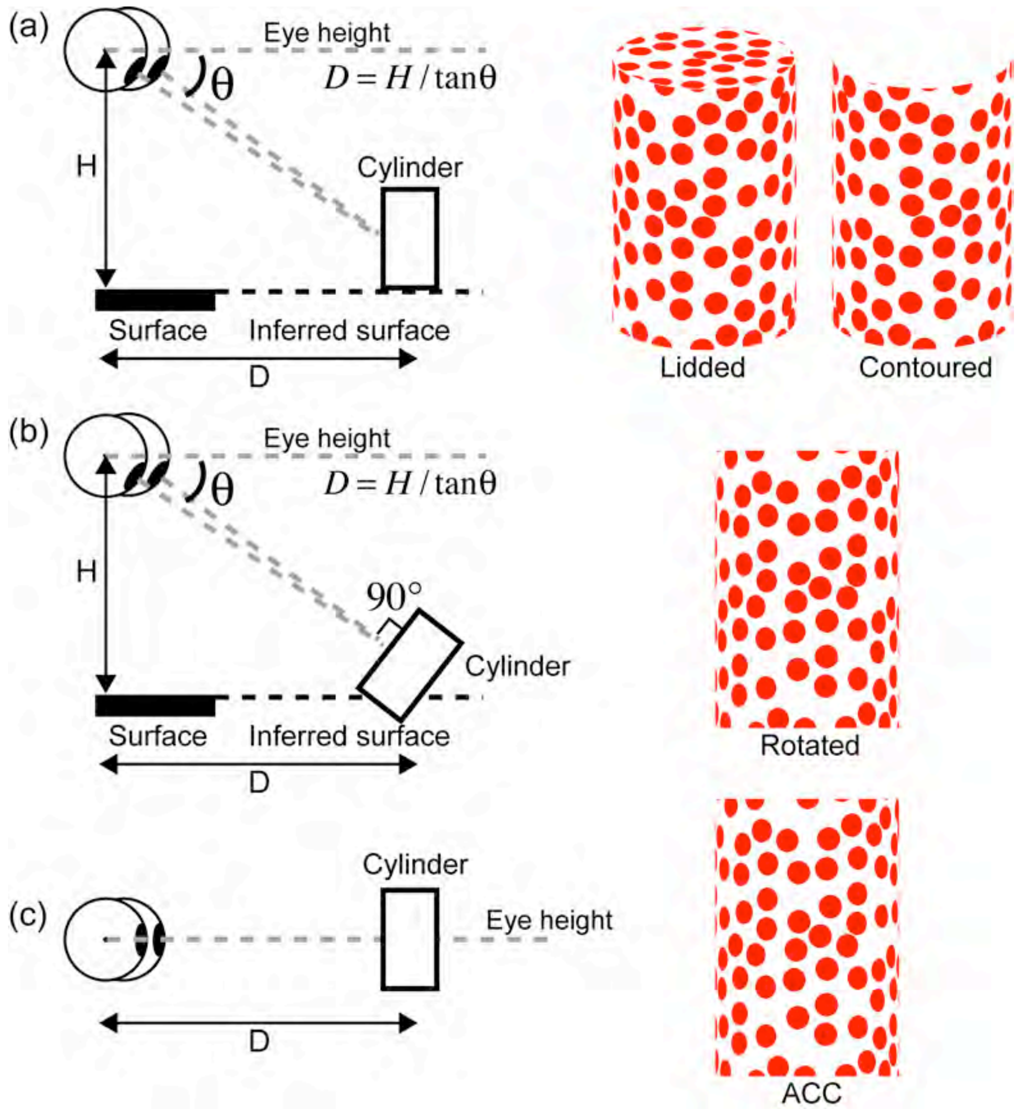

Figure 4.2: Diagram showing the experimental setup (left column) and a schematic representation of the stimuli (right column) for each of the four viewing conditions. (a) For the 'lidded' and 'contoured' conditions the stimuli were viewed from above. In the 'lidded' condition the cylinders top/lid was rendered whereas in the 'contoured' condition the top/lid was absent. (b) In the 'rotated' condition, stimuli were again viewed from above, but the stimuli were rotated around the vertical axis such that they appeared frontoparallel to the observer at each viewing distance. (c) In the ACC condition stimuli were viewed at eye level, this is the standard configuration for this type of task. In all conditions stimuli were viewed along the median plane.

During the ACC condition the response keyboard was positioned on the table surface to which the monitor platform was attached, this is the standard experimental viewing situation. Throughout all viewing conditions the table surface and monitor platform were covered with matt black cloth. In the lidded, contoured and rotated conditions the keyboard was positioned on a matt black raised table surface in front of the observer (Figure 4.2). The height of the raised surface was matched to that of the bottom of the rendered 3-D stimuli. The raised surface fitted around the bite bar mount and consisted of three sections in a 'C' shaped configuration with its open side facing the monitor. This configuration ensured that the observer's view of the stimuli on the monitor screen was not occluded by the table surface at any viewing distance.

The raised table surface served to provide haptic information consistent with the computer rendered stimuli standing on a real support surface, information about the height of an observer's eye(s) above the support surface is required to estimate distance from vertical gaze angle (Cutting & Vishton, 1995). Observer VD completed two versions of the 'Contoured' condition; one with the table surface present and one when that table surface was absent. This served as a check as to the role of appropriate haptic feedback about the support surface for the estimation of distance from height in the field. The stimuli were viewed in a dark room with no other sources of illumination and periodic breaks were taken to minimise dark adaptation. Observers were aware that they were viewing computer generated stimuli and that the raised table surface did not extend all the away to the monitor.

4.2.4 Procedure

Observers performed a standard apparently circular cylinder (ACC) task (Johnston, 1991). An elliptical cylinder was presented to the observer for a period of three seconds; the observer's task was to respond with a button press as to whether the cylinder they were presented with was 'stretched' or 'squashed' in depth extent relative to that of a circular cylinder. The cylinders had a constant width but variable depth; the exact depth range depended on the observer, but incorporated stimuli with depth profiles clearly larger and smaller than that of a circular cylinder. Due to practical constraints of the experimental apparatus viewing conditions were blocked by eye height, the 'eye level' block consisted of the standard ACC condition alone and 'view from above' block consisted of the lidded, contoured and rotated conditions. The order in which the observers completed the 'eye level' block and ‘view from above’ block was determined at random. Within the view from above block the order in which the lidded, contoured and rotated conditions were completed was randomised[8].

Stimuli were presented on four randomly interleaved adaptive staircases, implemented with custom written Matlab code. Two staircases (1-down 3-up and 3-down 1-up) started at high stimulus values (cylinders with a large depth) and worked their way down the psychometric function toward the point of subjective equality (PSE). The other two staircases (3-up 1-Down and 1-up 3-down) started at low stimulus values (cylinders with a small depth) and worked their way up the psychometric function toward the PSE. The 1-down 3-up and 1-up 3-down staircases converge on the PSE relatively quickly, whereas the 3-down 1-up and 3-up 1-down staircases converge on the PSE more slowly, sampling the edges of the psychometric function to a greater extent. Sampling data points at high performance levels, low and high on the psychometric function, more tightly constrains the fit of the function to the data as points at high performance levels carry more information regarding the shape of the underlying function (Wichmann & Hill, 2001a, Wichmann & Hill, 2001b).

The 1-down 3-up and 3-down 1-up staircases converge at a performance level of 0.794, whereas the 3-up 1-down and 1-up 3-down staircases converge on a performance level of 0.206, these are relatively high and low performance levels, so should provide sufficient data to constrain the fit of the psychometric function. Two step sizes were used for each staircase, initially the stimulus step was large (10 or 20% steps depending on the observer) in order to avoid spending a large number of trials at very high performance levels, after three stimulus reversals the step size decreased (3-5% steps depending on the observer) in order to move more slowly along the function. Each staircase terminated after twelve reversals, a reversal was

[8] After data collection it was found that observer PBH had globally misaligned the grid of Gaussian blobs to the loom during calibration at the 40 and 60cm viewing distances for the 20cm viewing height. This misalignment consisted of setting the position of the whole calibration grid 1cm below central loom alignment in one (40cm) or both (60cm) eyes. This necessitated recalibration at the 40 and 60 cm viewing distances with an eye height of 20cm and recollection this observer’s data at for the Lidded, Contoured and Rotated conditions at these distances. Data recollection followed the order in which initial data collection had occurred.

defined as two consecutive stimulus steps on the staircase that were in differing directions. A representative example of the convergence characteristics of the staircase procedure is shown in Figure 4.3a, the psychometric function fit to this data is shown in Figure 4.3b. Details of the fitting procedure can be found in the results section.

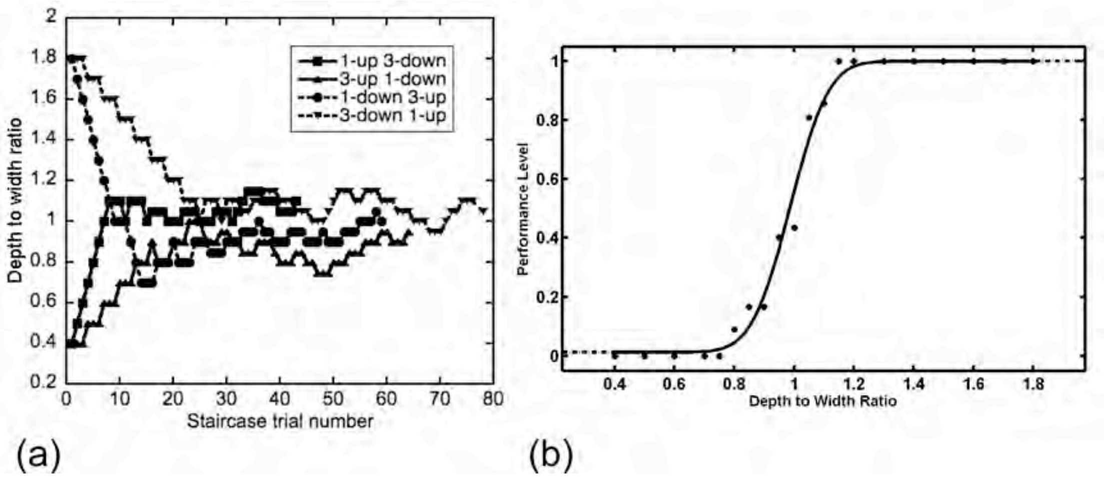

Figure 4.3: (a) A representative example of the convergence of the four interleaved adaptive staircases for observer VD in the 'Lidded' 80cm condition, the key identifies the data from each staircase (b) The psychometric function fitted to the staircase data shown in (a).

4.3 Results

4.3.1 Psychometric functions and scaling distances

For each viewing condition a Cumulative Gaussian function was fit to the data from all four staircases and the PSE and slope estimated using a maximum likelihood procedure (Wichmann & Hill, 2001a, Wichmann & Hill, 2001b). A representative psychometric function from observer VD is shown in Figure 4.3b. The PSE served as an estimate of the depth to width ratio of a cylinder needed for it to be perceived as 'stretched' and 'squashed' in depth extent relative to circular an equal number of times i.e. the depth to width ratio of a circular cylinder. Scaling distances were calculated for each condition (Appendix A4). These indicate the distance at which the cylinder represented by the PSE would need to be viewed in order to project disparities at the retina consistent with those of a circular cylinder.

The calculation of scaling distances assumes that all error in set shape is due to scaling disparity with a misestimate of the viewing distance. If observers correctly perceived 3-D shape a plot of scaling distance against physical viewing distance should result with a diagonal straight line with a slope of one and intercept of zero. Individual depth to width ratios and scaling distances are plotted in Figure 4.4. Error bars on the PSE values are 95% confidence intervals from the psychometric function fit, those of the scaling distances are 95% confidence intervals calculated by converting the limits of the confidence intervals around the PSE into scaling distance values in the same way as the PSE itself was converted (see Appendix A4 for details).

4.3.2 Observer performance and differences across conditions

On initial inspection two things are striking about the data, the first is the high accuracy with which observers are able to complete the task under a number of viewing conditions, including the least informative condition of the ACC. The second is the large individual differences between observers. Observers PBH and PS show much more accurate performance across conditions compared to observer VD, the PSE values of these observers cluster around the horizontal dashed line indicating shape constancy (Figure 4.4a and 4.4c). When these shape settings are plotted as scaling distances the data straddles the diagonal line representing a one to one relationship between scaling distance and physical viewing distance (Figure 4.4b and 4.4d). These observers' data contrast dramatically with that of observer VD, who shows large biases in all viewing conditions except for the lidded condition (Figure 4.4e). When plotted as scaling distances it is clear that VD's shape settings are consistent with this observer using a large underestimate of viewing distance to scale disparity across the whole distance range (Figure 4.4f). Note the difference in scale on the PSE and scaling distance plots of observer VD.

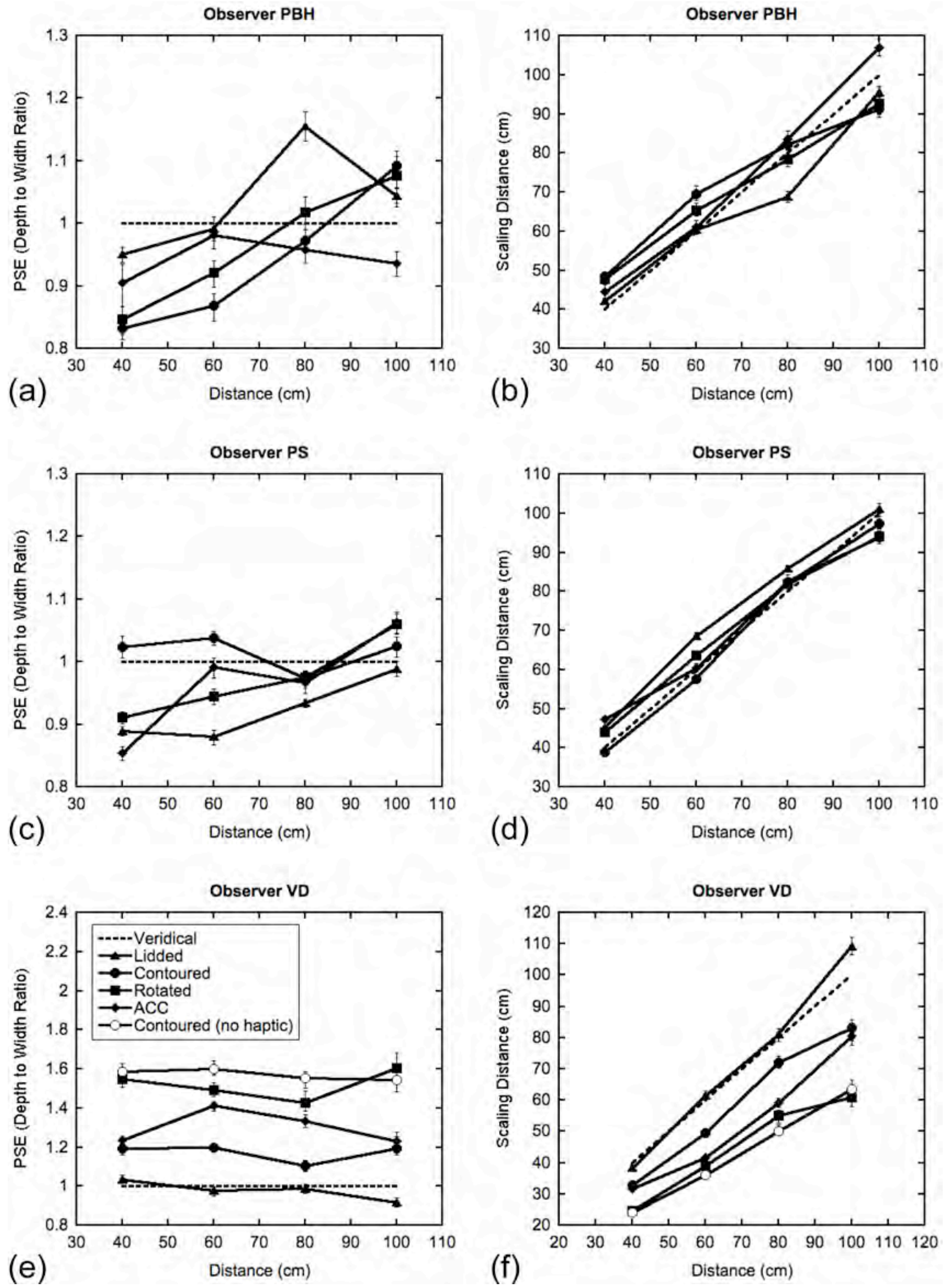

Figure 4.4: Showing the point of subjective equality (PSE) and scaling distance for each of the viewing conditions. Rows show the data from the three observers, the left column shows the PSE data and the right column the scaling distances. Error bars show 95% confidence intervals. The dashed line in both sets of graphs shows performance if observers were able to correctly estimate 3-D shape. Note the difference in the scales of the axes for observer VD.

A plot of scaling distance and physical viewing distance is typically well fit by a linear relationship (Johnston, 1991). A one to one mapping between physical viewing distance and scaling distance would result in a linear fit with a slope of one and an intercept of zero. The slope characterises the way in which the distance used for scaling changes as the physical viewing distance changes, for example, a slope less than one indicates that observers are not taking changing distance into account as much as they should, conversely, a slope greater than one indicates that observers are overcompensating for changing distance. For a linear fit with a slope of one, the intercept characterises overall bias in the distance estimates, for example, a change in the intercept of $\pm x$ would indicate that distance is over- or under-estimated by a constant amount, x, across the whole distance range.

Linear fits to scaling distance data typically result in a line with a slope less than one and an intercept greater than zero, indicating that disparity-defined shape was scaled with an overestimate of near distances and an underestimate of far distances (Johnston, 1991). The point where the linear fit crosses a one to one relationship between scaling and physical distance is taken as the physical distance at which distance is estimated correctly. This has been termed the "specific" (or "abathic") distance around which distance estimates are said to contract (Gogel & Tietz, 1973, Johnston, 1991). If shape constancy can be characterised by the extent to which near distances are overestimated and far distances underestimated, a linear fit to physical and scaling distance will rotate about the abathic distance with the extent of shape constancy. In this instance the value of the intercept is not independent of the slope, as a change in slope also results in a change in the intercept.

A least squares linear fit was made to the scaling distance data plotted against physical viewing distance and 95% confidence intervals of the slope and intercept estimated, these are shown in Figure 4.5. The slope of the fitted line is taken as a measure of depth constancy, a slope of one indicating 100% constancy and a slope of zero indicating 0% constancy (Glennerster et al., 1996, Glennerster et al., 1998). The intercept of the fitted line is typically not reported or interpreted, but it is clear that for overestimation of near distances and underestimation of far distances to occur there must be both a decrease in slope and an intercept greater than one.

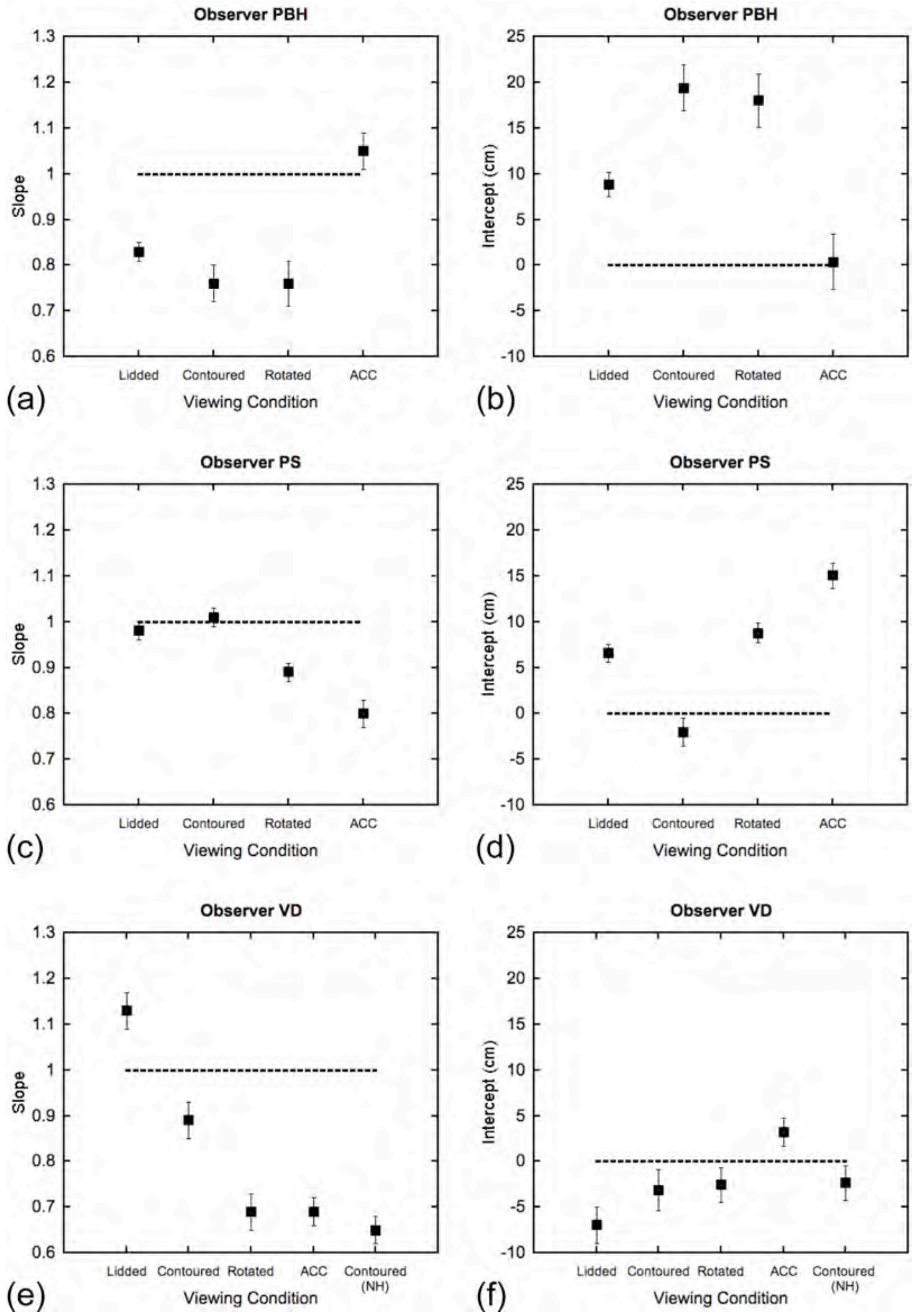

Figure 4.5: Shows the slope and intercept of the least squares fit to a plot of scaling distance against physical viewing distance. Rows show the data from the three observers, the left column shows the slopes and the right column the intercepts. Error bars show 95% confidence intervals. The horizontal dashed line in both sets of graphs shows predictions for a one to one relationship between scaling distance and physical viewing distance (i.e. a slope of one and intercept of zero).

Across observers and viewing conditions slopes ranged between 0.65 and 1.1, representing 65 and 110% constancy, these values are comparable to those found in previous studies. Glennerster et al. (1996) found mean constancy across observers to be around 75% in an eye height ACC task with cylinders and dihedral angles, and near 100% with a shape matching task. In an ACC task with cylinder stimuli Glennerster et al. (1998) found constancy of 56-71% for experienced psychophysical observers and 46-62% for naïve observers, depending on stimulus size and the available distance cues. Using an ACC type task with sparse real world objects Bradshaw et al. (2000) found constancy of 76-80% depending on the viewing condition (binocular static viewing, binocular and motion parallax or monocular and motion parallax). By contrast Johnston (1991) using an ACC task with cylinder stimuli found constancy of 27 and 26% for the two observers tested (see also Johnston et al., 1993).

For observer PS the slopes of the scaling distance to physical distance relationship are very close to one for the lidded and contoured viewing conditions, for the contoured condition a slope value of one is within the 95% confidence intervals. By comparison slopes are significantly less than one for the rotated and ACC conditions, but overall performance is very high with the shallowest slope in the ACC condition showing scaling of around 80%. A similar pattern of results is found for observer VD, slopes are nearest one for the lidded and contoured condition and less than one for the remaining viewing conditions. Slopes for this observer generally show less constancy compared to those of observer PS and a slope greater than one is found in the lidded condition. This pattern of results is consistent with the lidded and contoured conditions containing the most information about object distance and shape. Observer PBH shows the opposite pattern of results to PS and VD, for this observer the ACC condition shows a slope close to one and the remaining conditions slopes show around 75-80% constancy. This is surprising as the ACC condition contains the least information about distance and shape.

Although the slopes of observers PS and VD were closest to one for the lidded and contoured conditions, which contained the most information about distance and shape, it is not clear whether a change in slope is the best way to characterise the data of observer VD. In addition to a significant change in slope across conditions this observer also shows large shifts in shape settings in most viewing conditions, consistent with a more global underestimation of object distance across the whole range. A lack of scaling is typically characterised as the extent to which near distances are overestimated and far distances underestimated (Johnston, 1991), this is indicated by a decrease in slope and an increasingly positive intercept for the scaling distance function. Those conditions with least bias have a slope closer to one and an intercept closer to zero. This theoretical pattern of results is not consistent with shape settings of observer VD, but is generally consistent with the settings of PBH and PS. For PBH and PS those conditions with a shallower slope also have a more positive intercept (Figure 4.5).

Observer VD completed two versions of the contoured condition, in one version the raised table surface providing haptic feedback consistent with the computer rendered objects was present and in another it was absent (these are labelled “Contoured” and “Contoured no haptic (NH)” in Figures 4.4 and 4.5). It can be seen that this observer’s shape settings and scaling distances in the contoured condition with the raised table surface were far closer to veridical than when the table surface was absent. This is reflected in the global shift of shape settings toward veridical (Figure 4.4) and an increased slope of the scaling distance function (Figure 4.5). This improved performance could be due to the raised table surface providing consistent haptic feedback needed for the estimation of absolute distance from vertical gaze angle or to practice effects. The contoured condition without haptic feedback was the first condition to be completed and the contoured condition with consistent feedback the last, the intervening conditions could therefore have provided practice at the task, which improved performance, this will require further investigation.

4.4 Discussion

4.4.1 General discussion: comparison with previous studies

Overall the performance of observers was generally high across most viewing conditions, especially so for observers PBH and PS. Shape constancy, as measured by the slope of the scaling distance function, varied from 65-110% across observers and viewing conditions. Average shape constancy was 85% in the ACC condition, 78% in the rotated condition, 89% in the contoured condition and 98% in the Lidded condition. The average shape constancy of 85% found in the ACC condition is comparable to that found in previous studies that have used ACC tasks, the average constancy of 98% found in the lidded condition is comparable to that found in tasks that do not require an estimate of absolute distance for accurate performance (Glennerster et al., 1996). This suggests that the additional information available when an object is viewed from above is sufficient to dramatically improve shape constancy. Standard shape judgement tasks that present objects at eye height might therefore underestimate the level of constancy that can be achieved from binocular visual cues in isolation.

There were however large differences in the level of bias exhibited between individuals across conditions, as has been found with many previous studies (Todd & Norman, 2003). Both PBH and PS showed changes in constancy across conditions consistent with differences in the extent to which near distances were overestimated and far distance underestimated (Johnston, 1991). Observer VD also showed significant differences in constancy across conditions, but this observer's data were also characterised by a more global shift in shape settings, consistent with the underestimation of all viewing distances across most viewing conditions. Overall no clear benefit of having height in the field available in the rotated condition was found across observers, in fact average shape constancy was reduced compared to that found in the ACC condition.

Performance did however improve for observers PS and VD, when additional information about object shape was provided in the lidded and contoured conditions. Observer VD showed a dramatic improvement in the lidded condition characterised by a large shift in shape judgements from those consistent with an underestimation of distance to near veridical performance. This was also the case of the contoured condition but to a much lesser extent. The scaling distance functions of observer VD also showed an increased slope in these conditions. For observer PS the improved performance in the lidded and contoured conditions was not as dramatic as for observer VD, but resulted in slopes very near to one and intercepts closer to zero.

In contrast, the performance of observer PBH was not dramatically improved in either the lidded or contoured conditions, instead this observer showed near veridical performance in the ACC condition. This is surprising as the ACC condition provided the least information about object distance and shape and has been shown to produce biases in perceived 3-D shape in numerous other studies (Glennerster et al., 1996, Glennerster et al., 1998, Johnston, 1991). It is possible that observer PBH was at a ceiling performance level in the ACC condition, in which case the addition of distance and shape information would not be expected to improve performance. However, additional information about distance and shape provided by height in the field and a view of the cylinders upper surface and contour, actually made performance worse. This is difficult to reconcile with the information available in the stimuli.

Large individual differences are a common finding in studies investigating the perception of 3-D shape (Johnston et al., 1993, Todd & Norman, 2003). It is likely that these differences reflect the differential way in which observers weight and combine noisy sensory data with prior information about distance and shape (Hillis et al., 2004, Landy et al., 1995, Yang & Purves, 2003). However, it seems unlikely that more reliable information about 3-D shape was available in the ACC condition for PBH and that this was somehow degraded in combination with less reliable information from height in the field and a view of the cylinders upper surface and contour. Viewing an object from above should provide more reliable information. Todd (2004) has suggested that observers might adopt different heuristic strategies to complete 3-D shape judgement tasks when faced with reduced cue information. Although this is possible, the nature of the heuristic would need to be specified in order to account for the pattern of data and understand observer's performance.

4.4.2 The role of information from height in the field

It is interesting that no effect of the addition of height in the field was observed in the rotated condition. During our everyday interaction with objects in our environment eye height above the supporting surface upon which objects rest is typically greater than the 25cm used in the current experiment. However, a baseline for triangulation of

25cm with height in the field is much greater than the $\approx$6.5cm baseline for triangulation with vergence (the interocular distance). Assuming roughly equivalent levels of noise in eye direction and vergence signals, height in the field would be expected to be the more reliable cue. Effects of height in the field on visuomotor positioning and prehension have also been demonstrated with a comparable 35cm eye height above the supporting surface (Gardner & Mon-Williams, 2001). It is possible that the raised table surface did not provide sufficient information about the position of the support surface for the estimation of absolute distance from height in the field (Cutting & Vishton, 1995).

Observers were aware that the table surface did not extend the full way to the screen across viewing conditions and that the stimuli were computer generated and could not be physically located on this surface. Contrary to this suggestion is the improved constancy of observer VD in the contoured condition when the raised table surface was present, compared to when it was absent. Although, as has been stated, whether this was solely due to the availability of consistent haptic feedback will need to be investigated further. It is possible that the use of height in the field as a cue to absolute distance was selectively hindered in the rotated condition due to the way in which the stimuli were rendered. In order to equate the information available to the observer from the surface of the cylinder in the ACC and rotated conditions, the cylinders in the rotated condition were rotated around their vertical axis to be gaze-normal (Figure 4.2).

To be interpreted as an object resting upon a support surface the stimuli would need to have been sitting in a dip in the surface or be somehow supported so as not to fall over, especially so at the nearest distance where the rotation angle was greatest. Although we readily interact with objects that are not resting flat on a supporting surface, in the experimental viewing situation, where there was no visual information from the inferred support surface, this could have interfered with the use of height in the field information. An effect of height in the field on perceived shape might be found if additional visual information about the position and orientation of the support surface were available, such as would be available with real objects on a full table surface These points reflect the fact that the increased control gained over stimulus information by using computer rendered stimuli is generally at the cost of lost ecological validity.

4.4.3 Differences between studies in the level of perceptual bias

Previous studies have tended to focus on how multiple cues are involved in the estimation of shape when stimuli are presented at eye height, rather than on the additional cues that are available for the estimation of distance and shape when stimuli are viewed in a more naturalistic viewing situation. Although the results of the ACC condition presented here are comparable with much of the experimental literature (e.g. Glennerster et al., 1996, Glennerster et al., 1998), some studies using eye height ACC tasks with disparity-defined stimuli have found greater biases than those found here. Similarly, studies that have used eye height ACC tasks but with stimuli defined with multiple visual cues such as disparity and texture have found greater biases to those found in the present study where viewpoint was altered and a view of an objects upper surface and contour were available.

For example, Johnston (1991) got observers to perform an ACC task with horizontally orientated disparity-defined cylinders presented at distances of 53.5 to 214cm, and found shape constancy of $\approx$25%. Johnston et al. (1993) had observers complete an ACC task with horizontally orientated cylinders defined by texture and disparity, and found depth to width ratios ranging from 0.87 at 50cm to 1.85 at 200cm. This represents an over- and under-estimation of depth, by 15 and 54% respectively. Todd and Norman (2003) had observers complete an ACC-like task at viewing distances of 57cm and 171cm with hinge stimuli rendered with either (1) disparity and texture, (2) motion and texture or (3) disparity, motion and texture. Unlike Johnston et al. (1993) they found depth to be overestimated at both viewing distances, set depth to width ratios ranged from 0.95 to 0.57, with an average of 0.79. This corresponds to an overestimation of depth ranging from 5 to 75%, with an average of 27%.

Part of this difference could be due to height in the field and a view of an object's upper surface and contour providing more reliable information about shape compared to cues such as disparity, texture and motion. Differences between the data presented here and previous studies may also be in part due to the increased distance range used in these studies. The distance range used in the present study was 50 to 100cm, binocular cues should be particularly useful within this distance range as 90% of the vergence range is used for distances below 100cm (Collewijn & Erkelens, 1990). The studies detailed above used similar near distances but have included distance up to $\approx$200cm, this is the limit at which vergence is thought to be an informative distance cue (Mon-Williams & Tresilian, 1999).

Distance judgement tasks in which vergence is the predominant distance cue have shown that distances are progressively underestimated as the physical viewing distance increases (Viguier et al., 2001). As was shown in Chapter 3 this can be predicted from geometric consideration of distance estimation from a noisy vergence signal. This increased distance range explanation cannot however account for instances of comparable shape constancy to that found here in an ACC task across a distance range between 38 and 228cm (Glennerster et al., 1996). Part of the answer might be in conflicting visual information present in the experimental setting.

4.4.4 The role of conflicting visual information

Although increased experimental control can be gained by using computer rendered stimuli, most computer rendered stimuli are unintentionally cue conflict stimuli; this is because there are typically numerous uncontrolled cues available to the observer which conflict with those cues intentionally used to portray the distance and shape of the stimuli. Conflicting cues arise from both the rendering of the stimuli and the equipment that is used for stimulus presentation. For example, the elements of a stereogram also carry a conflicting texture cue that normally indicates a frontoparallel surface (Zabulis & Backus, 2004). Similarly, many studies present computer-generated stimuli at a range of simulated distances on a monitor positioned at a fixed physical distance from the observer (Brenner & Landy, 1999, Glennerster et al., 1998). In this situation focus cues indicating the physical distance and orientation of the monitor will conflict with the simulated distance and shape of the experimental stimuli, these cues can act to introduce bias in estimates of 3-D shape (Watt et al., 2005a).

The visual system is a robust system and is likely to use information from these conflicting sources of information when making estimates about object or surface properties. Combining these conflicting sources of information with those used to intentionally render the stimuli, through for example weighted averaging (Landy et al., 1995), could contribute to the perceptual bias and lack of shape constancy observed in many studies. However, numerous methodological differences between studies make it difficult to unambiguously attribute some aspects of reduced shape constancy to uncontrolled conflicting cues, and some to the processing by which the visual system makes estimates from visual information in the natural environment. For example, conflicting texture information could account for the increased bias seen in the ACC task of Johnston (1991) compared to that found in the present study, however, some studies have used similar stimuli with conflicting textural information but have demonstrated similar levels of bias to that found here (Glennerster et al., 1996, Glennerster et al., 1998). In these latter studies the stimuli were blurred in the horizontal dimension for anti aliasing, so it is possible that this might have decreased the effectiveness of the conflicting texture cue.

Similarly, although conflicting focus cues have been shown to contribute to the systematic distortions of shape found when stimuli are presented at multiple simulated distances but viewed on a monitor at a single fixed physical distance (e.g. Tittle et al., 1995), this is unlikely to fully account for biases demonstrated when the monitor distance and simulated object distance are concordant (Todd & Norman, 2003), despite conflicting focus cues of a lesser magnitude remaining (Watt et al., 2005a). The use of some cues is also dependent on statistical assumptions the observer makes in interpreting the visual information, for example to estimate slant/shape from texture the visual system would need to assume that the texture is homogenous and isotropic (Knill, 1998).

Additional methodological differences between studies which could, or have been shown to, contribute to differences in shape constancy include, (1) the size, density and luminance of the dots defining the stimulus (2) whether appropriate monocular regions of a stimulus are rendered (Gillam & Borsting, 1988) (3) whether isolated stimuli are presented or whether stimuli are presented ‘embedded’ in a surface surround (Econopouly, 1996) (4) the precise task the observer is required to perform (Glennerster et al., 1996, Todd & Norman, 2003) (5) whether or not the monitor is spatially calibrated (Backus et al., 1999) (6) the overall size of the stimulus (Collett et al., 1991) (7) the level of ambient lighting and therefore the availability of other distance and shape cues (Glennerster et al., 1998) (8) knowledge the observer has of the experimental setup (Glennerster et al., 1998).

4.4.5 Summary

This study sought to investigate the role of height in the field and a view of an object's upper surface and contour in the perception of 3-D shape. Whilst there was little evidence for improved performance when height in the field was the only additional cue, there was evidence to suggest that a view of an object's upper surface and contour resulted in less bias in observer's perceptions of 3-D shape. Average shape constancy, as measured by the slope of the scaling distance function was 98% when an observer viewed a disparity-defined object from above with a view of its upper surface rendered. This suggests that systematic biases in perceived shape demonstrated when stimuli are presented at eye height may in part be accounted for by the removal of many cues to distance and shape that would typically be present in natural viewing situations. There remained some level of bias in some conditions, this varied considerably between observers, but was consistent with the range of bias demonstrated in previous studies.

Efforts were made to make the stimuli as realistically rendered as possible and to minimise conflicting cues to object distance and shape. The presence of uncontrolled cues to distance and shape, along with restricting the available visual cues, will contribute to systematic biases demonstrated in previous studies. Distortions in perceived shape have however been demonstrated with real world stimuli (Watt et al., 2005a) and can in principle be predicted from the way in which the visual system makes estimates from stochastic visual information (Hogervorst & Eagle, 1998). It is possible that the level of bias that remains in natural viewing may have little consequence for the control of behaviour in the environment, such as when reaching to pick objects up, especially if the control of such actions does not require the metric information about object distance and shape (Bradshaw & Elliott, 2003).

Chapter Five

Integrating information from multiple objects in the control of prehension

5 Integrating information from multiple objects in the control of prehension

5.1 Introduction

5.1.1 The importance of disparity for reaching and grasping

Horizontal binocular disparity could potentially provide metric information about three-dimensional shape and position of objects in the world. However, numerous studies have shown that the 3-D shape of objects is systematically misperceived when objects are defined by horizontal disparity alone. This misperception is consistent with observers misestimating object distance and using this incorrect distance estimate to scale disparity (Johnston, 1991, Tittle et al., 1995, Todd et al., 1995). It is an open question as to the consequences these systematic distortions in perceived shape will have for the planning and control of prehensile movements. It has been argued that prehension is one task for which accurate information about 3-D shape and distance might be needed (Milner & Goodale, 1995). During prehension there is direct interaction with objects in the world, so any misestimation of 3-D shape could result in a greater cost. For example, misestimating the distance or shape of a coffee cup when picking it up from a table could result in fumbling the grasp and knocking the mug over (Cuijpers, Brenner & Smeets, 2006).

Because disparity offers the possibility of recovering metric information about the structure of the scene it has traditionally been seen as the primary cue for the control of prehension, the visuomotor system has been described as "falling back on" monocular cues when disparity information is unavailable (Servos et al., 1992). Equation 5.1 shows the geometric relationship that exists between depth from disparity and object distance, in this equation $\hat{d}$ represents estimated depth, $\hat{\eta}$ the measured retinal disparity, $\hat{D}$ the estimated viewing distance and $\hat{I}$ the visual systems estimate of the interocular distance. It is normally assumed that both $\hat{I}$ and $\hat{\eta}$

are estimated correctly (see Chapter 1 and 3 for a more detailed discussion). From equation 5.1 it can be seen that to estimate depth from disparity an estimate of the viewing distance is required, estimated depth from disparity is proportional to the square of estimated distance. More broadly, this geometric relationship shows that if object distance is underestimated depth from disparity will also be underestimated, and similarly if object distance is overestimated depth from disparity will be overestimated.

Previous research has shown that the kinematic parameters of prehensile movements are systematically related to object properties such as distance and size. Specifically, maximum wrist velocity has been shown to be linearly related to object distance and maximum grip aperture linearly related to object size (Jeannerod, 1988). If maximum grip aperture and maximum wrist velocity can be taken as indirect measures of perceived distance and size/depth respectively, one would expect there to be relationship between the indices such that increased wrist velocities would be associated with increased grip apertures and decreased wrist velocities would be associated with decreased grip apertures (e.g. Servos et al., 1992). This follows from the geometrical relationships that existed between depth-from-disparity and distance, and width-from-angular-size and distance (Equations 5.1 and 5.2 respectively). From these equations it can be seen that depth-from-disparity scales with the square of distance whilst width-from-angular-size scales directly with distance.

$$d \approx \frac{\eta D^2}{I} \tag{5.1}$$

$$h = 2\left(D\tan(\alpha)\right) \tag{5.2}$$

However, plots of maximum wrist velocity against distance and maximum grip aperture against depth/width typically show slopes less than one and intercepts less than zero. This means that the relationships between maximum wrist velocity and maximum grip aperture will not be well fit by the relationships shown in Equations

5.1 and 5.2. Although this is the case, if these indices can be taken as measures of perceived distance and size (depth or width), there should be some weaker relationship between the indices such that increases in maximum wrist velocity should be coupled with increases in maximum grip aperture. Furthermore, if the difference in the scaling of disparity and angular size with distance is preserved in these indices, this relationship should be stronger when grasping the depth of and object as compared to its width.

The role of binocular vision in the control of prehension has typically been assessed by contrasting kinematic indices of prehensile movements carried out under binocular and monocular vision. As was reviewed in the introduction there is inconsistent evidence as to the precise role of disparity in the control of prehensile movements. Servos et al. (1992) found that reaches carried out under monocular vision exhibited lower peak velocities and smaller grip apertures compared to binocular reaches (see also Servos, 2000). This is consistent with a coupling perceived distance and depth, as indexed by maximum wrist velocity and maximum grip aperture, but it is not clear how this can be related to disparity scaling, because with monocular vision there is no retinal disparity to scale. In contrast to these results a number of studies have found little or no effect of the removal of binocular vision on early indices of prehension such as peak wrist velocity, but significant effects on latter indices of prehension such as maximum grip aperture and time in the slow phase of the movement (Bradshaw & Elliott, 2003, Churchill et al., 2000, Melmoth & Grant, 2006, Watt & Bradshaw, 2000).

For example, Watt and Bradshaw (2000) found that with monocular viewing peak grip apertures were larger and reached earlier in the movement than during binocular viewing, and more time was spent in the terminal portion of the reach when the hand closed in on the object to be grasped. This suggests that a dissociation exists between the information used in control of the reach and grasp components of prehension, or that similar information is used, but that this information is weighted differently for the reach and grasp components. The pattern of results found by Watt and Bradshaw and others is consistent with the brain adopting a 'conservative strategy' in the face of increased uncertainty, whereby the hand is opened wider sooner and more slowly closes in on the object to be grasped. This would build in an increased margin of error so as to avoid collision with the object or an unstable grasp (Melmoth & Grant, 2006, Watt & Bradshaw, 2000). The importance of disparity in the terminal portion of the reach could also indicate that the role of disparity is in providing relative information about the position of the digits of the hand and the object to be grasped (Bradshaw & Elliott, 2003, Brenner & Smeets, 2006, Morgan, 1989), rather than the recovery of metric shape information (Milner & Goodale, 1995).

5.1.2 Are misperceptions of 3-D shape manifest in grasping?

Given the proposed importance of disparity for the control of prehension movements (Servos et al., 1992) and the demonstration of systematic distortions in the perception of shape from disparity (Johnston, 1991), it is important to understand whether these distortions are evident in kinematic indices related to the prehensile movement. There are significant differences between the viewing situations in which systematic distortions in perceived shape have been demonstrated and those in which natural prehensile movements are made. Perceptual distortions of shape have typically been demonstrated using computer rendered stimuli presented at eye height, in contrast prehensile movements are made toward real objects that are typically viewed from above on a support surface. This provides improved information about distance and shape from height in the field (Gardner & Mon-Williams, 2001) and a view of an object's upper surface and contour. This could mean that distortions in the perception of 3-D shape would not be evident in the viewing situations typical of natural prehension. Contrary to this, previous research indicates that distortions of perceived shape from disparity are evident in prehensile responses to real objects presented at both eye height (Cuijpers et al., 2006) and when viewed from above (Hibbard & Bradshaw, 2003). The results from the experiment described in Chapter 4 also suggest that the improved distance and shape information available from height in the field, and from a view of an objects upper surface and contour, are not sufficient for completely veridical perception of shape from disparity in all observers.

Hibbard and Bradshaw (2003) investigated whether binocular disparity alone was sufficient for the control of prehensile movements. Observers reached and grasped elliptical cylinders presented at either 30 or 50cm, across both their width and depth dimensions. These cylinders were either real (cylindrical wooden blocks) or disparity-defined (random dot stereograms). Real objects were viewed through a semi-silvered mirror on a surface 17cm below eye level, which was lit by a desk lamp. Disparity defined objects were viewed on a monitor screen via the semi-silvered mirror in a dark room and rendered such that they appeared at the same position as the cylinders in the real object condition. During presentation of the disparity defined stimuli real objects of the same dimensions were placed on the table surface at the appropriate distance from the observer. These were not visible to the observer but provided appropriate haptic feedback regarding shape and distance. All reaches were open loop, observers had a two second viewing period before the desk light (real objects) or screen (stereograms) was extinguished.

There were significant differences between reaches performed to real objects and those performed to disparity-defined objects. Peak wrist velocity for real and disparity-defined objects was approximately the same at 30cm, but at 50cm peak wrist velocity was significantly slower for the simulated objects. Maximum grip aperture for disparity-defined cylinders was greater than that for real objects at the near distance and approximately the same at the far distance, this was true when observers grasped either the depth or width dimensions. These differences could be due to a number of factors, firstly disparity in conjunction with vergence and height in the field may not be able to provide metric 3-D shape information for prehension. This is akin to the distortions of shape seen in perceptual studies (Johnston, 1991). Alternatively, the differences could simply reflect the level of uncertainty the visual system has about the shape and distance of the object to be grasped. In the disparity-defined object condition many visual cues are absent, this might result in increased uncertainty, which could in turn cause the system to adopt a conservative strategy of slower reaches with larger grip apertures (Watt & Bradshaw, 2000).

Cuijpers et al. (2006, 2004) have investigated reaching and grasping of elliptical cylinders with a range of aspect ratios, positioned at different distances from the observer, viewed either at eye level or from above. Observers were instructed to naturally reach and grasp objects with a precision grip (first finger and thumb only), but otherwise no specific instructions were given as to where observers should grasp the objects. Cylinders tended to be grasped across either their major or minor axes, which represent the most stable grasp points to apply grip force (Cuijpers et al., 2006). The distribution of grasp points was more variable for cylinders viewed at eye height and shifted linearly with changing cylinder orientation, but with a gain less than one. The increased variability of grasp points indicates that for eye height stimuli observers were less certain of the cylinder's shape. The gain less than one indicates that although observers shifted the orientation of their grasp with the orientation of the cylinder they did so less than would be required for them to grasp the cylinders across their major or minor axes (Cuijpers et al., 2006).

When grasping cylinders at eye height maximum grip apertures were uniformly increased relative to those found when grasping cylinders viewed from above, although the same relationship was found between maximum grip aperture and the cylinder's aspect ratio in both instances. This was interpreted as an increase in uncertainty about the shape of the object when presented at eye height. If maximum grip aperture is an index of perceived size, it might be expected to correlate well with the grip aperture in the final part of the movement before the hand contacts the object (Cuijpers et al., 2006). For grasping cylinders when viewed from above this was true, grip aperture in the last 1% of the traversed distance of the reach showed a correlation coefficient of 0.8 with maximum grip aperture, but this was decreased to 0.6 when grasping cylinders at eye height. This was interpreted in terms of observers misperceiving object size when cylinders were viewed at eye height (Cuijpers et al., 2006, Cuijpers et al., 2004).

Within the prehension literature it is difficult to distinguish between effects on grasping that can be attributed to systematic distortions in perceived shape, as demonstrated in perceptual tasks (Johnston, 1991, Tittle et al., 1995, Todd et al., 1995) and those due to the overall level of certainty an observer has regarding perceived shape (Cuijpers et al., 2006, Hibbard & Bradshaw, 2003, Melmoth & Grant, 2006, Watt & Bradshaw, 2000). Manipulating the availability of visual cues has so far invariably meant manipulating the reliability of information available about the object to be grasped. For example, removing binocular vision typically involves removing all input to one eye (Servos et al., 1992), this in itself could disrupt natural prehensile movements beyond any effect the removal of binocular information will have (Bradshaw et al., 2004). It is also possible that a change in the reliability of signals in and of itself introduces bias into estimates of 3-D shape (Hogervorst & Eagle, 1998).

5.1.3 Combining information from across the scene in the perception of three-dimensional shape

If metric information about shape is required for the programming of prehension, and this information is provided predominantly by disparity, the visual system will need to scale disparity with some estimate of object distance. Many potential sources of information that could be used to recover information about distance and shape are present in extended, spatially separate, regions of the scene. For example, it has been demonstrated that vertical disparity information, which can be combined with horizontal disparity to estimate geometrically correct 3-D shape (Mayhew & Longuet-Higgins, 1982), only improves perceived 3-D shape if the stimuli subtend 20 degrees or more (Bradshaw et al., 1996). This is attributable to the fact that vertical disparities increase with eccentricity, peaking at plus or minus 45 degrees (Rogers & Bradshaw, 1993). Utilising vertical disparities in natural scenes is therefore likely to require integrating information across spatially separate surfaces and objects. Recent evidence has shown that the visual system can indeed integrate vertical disparity information from spatially separated surfaces in the estimation of both size and shape (O'Kane & Hibbard, 2002). The potential value of information from other objects and surfaces for perception of 3-D shape and layout contrasts with much of the research in this area, which has tended to present isolated sparsely defined objects (Johnston, 1991, Johnston et al., 1994, Todd et al., 1995).

Another source of information that might be useful for the estimation of metric shape is object motion. The conjunction of stereo and motion information in an object theoretically allows its 3-D shape and therefore distance to be recovered (Richards, 1985). This is due to the different way in which object disparity and motion scale with distance. The differential scaling means that there is only one object distance at which both motion and disparity indicate the same object shape, this is the true distance of the object (Brenner & Landy, 1999). As was reviewed in Chapter 1, the evidence for the combination of motion and stereo information is equivocal. Some studies have found that the addition of motion to a stereo defined object improves judgements about its 3-D shape (Johnston et al., 1994), whereas others have not (Todd et al., 1995). If a better estimate of object shape and distance is obtained from stereo-motion combination, this distance estimate might be used to better estimate the 3-D shape of other objects in the scene.

Econopouly (1996) investigated this possibility. Observers were presented with two horizontally orientated cylinders defined by stereo and texture positioned one above the other on a textured background surface. The lower cylinder always defined a circular cylinder, whereas the upper cylinder varied in depth. In some conditions the lower cylinder was static and in others it rotated centrally around its vertical axis. The cylinders were viewed at either 37 or 200cm. The task was to indicate whether the upper cylinder was squashed or stretched in depth relative to a circular cylinder. When viewed next to a static lower cylinder, the upper cylinder's depth was misperceived such that its depth was overestimated at the near distance and underestimated at the far distance, consistent with previous results (Johnston, 1991). When viewed next to a rotating lower cylinder, the upper cylinder's depth was perceived more (but not completely) veridically at both the near and far distances, as was the depth of the rotating cylinder itself. This pattern of results is consistent with stereo and motion information in the rotating cylinder being used to better estimate object shape and distance (Johnston et al., 1994, Richards, 1985), and this improved distance estimate being used to rescale the disparities present in the second static cylinder.

Brenner and Landy (1999) further investigated the circumstances under which stereo-motion combination might be used to better estimate object distance and shape. Two texture and disparity defined ellipses were presented side by side. The observer's task was to set the size and depth of the left hand ellipse to match that of a tennis ball (present as a reference before each session). The observer also set the distance, size and depth of the right hand ellipse, but the exact settings made differed between experiments. In experiment one its shape and size were set to match that of a tennis ball and its distance was set to be half that of the left hand ellipse, which in some trials was static and in others trials rotated about a horizontal axis. In experiment two its size and shape were again set to match that of a tennis ball, but this time its distance was set to be the same as that of the left hand ellipse, which now always rotated. In a final experiment the distance of the right hand ellipse was set to match the distance of the left hand ellipse, which again always rotated, but now its size and depth were set to be double that of a tennis ball.

Improvements in perceived shape were highly restricted, shape settings were improved in the left hand ellipse when it rotated, but were only improved in the right hand static ellipse when it was set to be the same distance and size of the left hand ellipse. These improvements were also restricted to set shape, as set size and distance were not improved (Brenner & Landy, 1999). These data question the extent to which stereo and motion information are combined to estimate 3-D shape and distance. It is possible that shape and distance estimates are improved, but only applied under very restricted circumstances; alternatively the visual system could adopt different processing strategies for the recovery of 3-D shape when stereo and motion cues are available (Landy & Brenner, 2001). For example, the depth of a horizontally-orientated cylinder rotating round its vertical axis (as used in Econopouly, 1996) can be geometrically specified from two views of the cylinder's right angle end cut contour if one of these is fronto-parallel to the line of sight (Todd & Norman, 2003). Similarly, the contour of a non-spherical rotating ellipse changes as it rotates, spherical settings could therefore be achieved by nulling distortions in the rotating ellipses contour (Brenner & Landy, 1999, Todd & Norman, 2003). Local changes in depth and orientation could also be used to provide useful information about shape (Brenner & Landy, 1999).

In a control experiment Brenner and Landy (1999) sought to isolate the role of the occluding contour by removing the motion parallax cue through using a limited lifetime texture below the threshold for detecting structure from motion. In this control experiment observers set the shape of a single rotating ellipse. Settings for the limited lifetime texture ellipse were more spherical than for a static ellipse, but less spherical than when motion parallax was present. This suggests that both cues could have been used in the original shape judgement tasks. In a separate control experiment the shape of a monocularly viewed rotating ellipse was set to match that of a tennis ball. Paradoxically settings for this ellipse were comparable to that found with a binocularly presented rotating ellipse (Brenner & Landy, 1999). It therefore remains unclear what cues underlie the improved performance found in these studies.

It is possible that the perceived shape of both the rotating and static objects in these studies was improved through the use of alternative processing strategy rather than true stereo-motion combination (Landy & Brenner, 2001). It is also apparent that although the shape of the two objects was perceived *more* veridically, shape constancy was not achieved. The strategy which observers adopt to complete the task, whether stereo-motion combination or some alternative processing strategy, therefore results in partial shape constancy at best, despite both theoretically allowing veridical shape settings to be made (Richards, 1985, Todd & Norman, 2003). Interestingly, Econopouly (1996) demonstrated that the rotating cylinder did not need to be circular for perceived shape to be improved in the static cylinder. This suggests that methodological differences could underlie the more restricted effect found by Brenner and Landy (1999).

As Landy and Brenner (2001) point out, the objects in Econopouly (1996) were presented embedded in a surface so they were clearly at the same distance, whereas those of Brenner and Landy (1999) were not. In this latter study observers were also required to make size, distance and shape settings, which was not the case in Econopouly (1996). It is possible that observers confounded set size with set distance when making their settings in Brenner and Landy (1999), which might account to some extent for the restricted effects observed in this study (see also Brenner & van Damme, 1999). Some of the reduced constancy in both studies could also be due to the presence of conflicting cues to distance and shape from the monitor display surface (Watt et al., 2005a). However, regardless of the processing that underlies performance in these tasks, clear interactions occur between multiple objects in the scene when observers make estimates of 3-D shape. Similar effects might also be observed in visuomotor tasks such as prehension.

5.1.4 The use of environmental and contextual information for the control of prehension

We rarely reach and pick up objects in an otherwise empty scene, typically many objects and surfaces are present during natural prehensile movements. Despite this fact few studies have investigated the potential benefits additional objects and surfaces in the scene could have for the estimation of distance and shape. Instead, the role of visual cues in prehension has typically been studied in sparse single object scenes (Servos et al., 1992, Watt & Bradshaw, 2000), as has been the case with perceptual studies on the estimation of 3-D shape. Whilst it would seem reasonable to assume that visuomotor systems would benefit from the additional information available from other objects and surfaces in the scene, some authors have suggested that there are fundamental differences in the way in which 'perception' and 'action' systems process contextual information (Milner & Goodale, 1995). Specifically, it has been suggested that processing for the control of motor acts such as prehension could essentially ignore the information available from other objects and surfaces in the scene, focusing instead on the information available from the object to be grasped alone.

Pictorial visual illusions such as the Ebbinghaus illusion (Figure 5.1) have been used as a tool to investigate this proposed disassociation. Size contrast effects between the central target and the surrounding flankers are thought to underlie the illusory percept. It has been argued that whilst visual processing is clearly influenced by such effects, motor acts such as prehension would remain immune as these require metric information about distance, size and shape (Aglioti, Desouza & Goodale, 1995). In discussing the effect of size contrast illusions on perceptual and motor systems Milner and Goodale (1995) state "Mechanisms such as these, in which the relations between object[s] in the visual array play a crucial role in scene interpretation, are clearly central to perception… In contrast, the execution of a goal-directed act such as prehension depends on size computations that can be *restricted to the target itself*" (pp. 169-170, italics added).

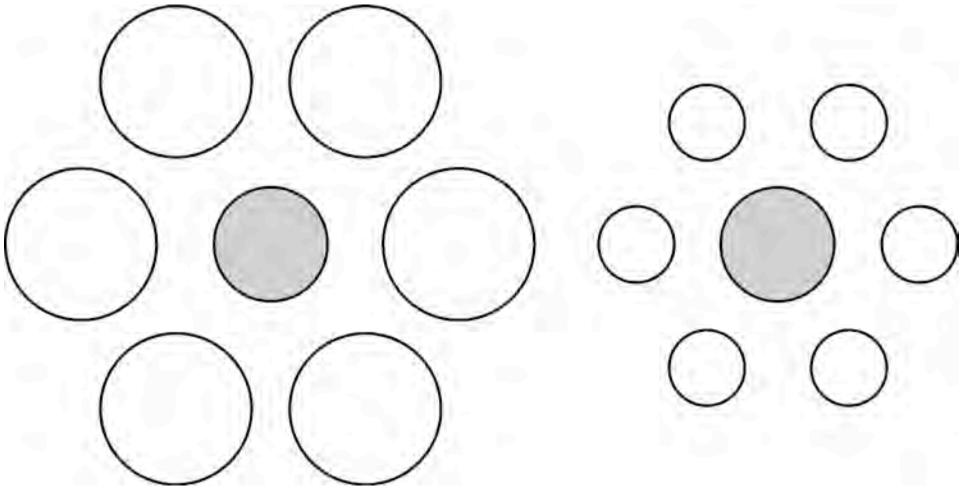

Figure 5.1: Diagram showing the Ebbinghaus illusion. Two target circles are shown shaded in grey, these circles are of identical size to one another. The left target circle is surrounded by larger flanking circles, whereas the right target circle is surrounded by smaller flanking circles. The effect of this is to cause the perceived size of the two target circles to differ. The target circle on the left is perceived to be smaller than the target circle on the right, despite both being identical in size.

Evidence purporting to show a disassociation perception and action systems in normal observers remains controversial (Aglioti et al., 1995, de Grave, Brenner & Smeets, 2004, Franz, 2001, Franz, 2003, Franz, Bulthoff & Fahle, 2003, Franz, Gegenfurtner, Bulthoff & Fahle, 2000, Haffenden & Goodale, 1998, Haffenden & Goodale, 2000, Haffenden, Schiff & Goodale, 2001, Smeets et al., 2002). Whilst this disassociation is one computational strategy that the visual system may have adopted, or evolved, contextual information from other objects and surfaces could provide valuable information about distance and shape for perception and visuomotor control (Bradshaw et al., 1996, O'Kane & Hibbard, 2002). It would seem likely that this informational benefit would outweigh the likelihood with which object configurations producing size contrast illusions will be encountered in everyday life.

A number of studies have shown clear effects of the presence of additional objects in the scene on prehensile movements in more naturalistic tasks. These effects encompass the timing of the reach (Watt & Bradshaw, 2002), the reach path (Tipper, Howard & Jackson, 1997), the reach velocity (Mon-Williams, Tresilian, Coppard & Carson, 2001, Watt & Bradshaw, 2002) and grip aperture (Churchill et al., 2000, Mon-Williams et al., 2001). For example, Churchill et al. (2000) found that the removal of environmental cues, produced by turning off the lights but providing feedback of the hand, resulted in larger grip apertures, slower peak velocities and more time spent in the slow phase of the reach. However, research on the role of other objects in the scene has tended to focus on more negative aspects. Additional objects in the scene are typically seen as obstacles to be avoided during the movement (Mon-Williams et al., 2001) or as distracting objects for which competing motor programs are produced (Tipper et al., 1997). This contrasts with the idea that information from additional objects in the scene could be particularly valuable for the programming and control of prehension.

5.1.5 Experimental rationale and predictions

The current experiment set out to investigate whether information from multiple objects in the scene is integrated in the programming and control of prehensile movements, as has been investigated in perceptual studies (Econopouly, 1996). Full methodological details are provided below, but briefly, observers reached and grasped a target object either presented on its own or in the presence of a flanking object. The target object was always static and viewed binocularly whereas the flanking object could either be static or rotating and could be viewed either binocularly or monocularly. The advantages of this experimental design are that the object to be grasped is always viewed binocularly; this means that the reliability of the information from this object is not altered. Another advantage is that it allows us to directly investigate disparity scaling in prehension, this has not been possible in previous studies which have contrasted binocular and monocular viewing (e.g. Servos & Goodale, 1994, Servos et al., 1992).

A prehension task is also beneficial as it allows concurrent measures of distance and shape to be gained and is a task for which disparity information has been considered of primary importance. Concurrent measurement of indices relating to perceived distance and shape avoids the possible confounds which might have been encountered in perceptual shape setting tasks (Brenner & Landy, 1999). A range of experiment predictions can be made regarding the presence of another object in the scene, these predictions depend on how the presence of the flanking object is interpreted. Discussion in this section will focus on the predictions that can be made regarding the flanking object as a source of additional information about distance and shape. Broader discussion of the role of flanking objects in the scene will be provided in the discussion, this will include discussion of factors such as the flanking object as a possible obstacle to the grasp (Mon-Williams et al., 2001, Tresilian, 1998) and the flanking object as a distracter object producing competing motor programs (Tipper et al., 1997).

The perceptual literature would suggest that the depth of a disparity defined object viewed at a near distance would be overestimated, as a consequence of its distance being overestimated (Johnston, 1991). If the same estimate of distance were used to scale angular size the width of the object should also be overestimated, but to a lesser extent than the object’s depth (see Chapter 1 and 3 for more details). Whilst the use of real objects in prehension studies generally prohibits the complete isolation of the disparity cue, observers reaching to grasp objects defined predominantly by disparity (see Watt & Bradshaw, 2000) would be expected to reach faster and with a wider grip aperture, than if they estimate depth correctly (Servos, 2000, Servos et al., 1992). This would demonstrate a relationship between perceived distance (maximum wrist velocity) and perceived size (maximum grip aperture when grasping depth or width), but this relationship would be unlikely to satisfy Equations 5.1 and 5.2 for the reasons discussed previously.

If information about object distance were improved, the estimated distance and depth of an object should decrease and become closer to their true values. As a consequence of this, observers would be expected to make slower reaches with smaller grip apertures (for both depth and width). Perceptual studies suggest that the overestimation of near distances decreases up to approximately 80cm after which distances are underestimated (Johnston, 1991). However, it is highly unlikely that this relationship would be evident in kinematic indices due to the restricted range of near distances used here, and the fact the relationship between maximum wrist velocity and distance, and maximum grip aperture and depth do not show slopes of one and intercepts of zero. Improved performance might therefore be characterised by a similar decrease in wrist velocity and grip aperture across the whole distance range.

The presence of a flanking object in the scene could provide the observer with improved information about distance and shape. In particular, the conjunction of stereo and motion information in the flanking object might be particularly beneficial as it offers the possibility of obtaining metric information about distance and shape, this improved distance estimate could be used to better estimate the shape of other objects in the scene (Brenner & Landy, 1999, Econopouly, 1996, Richards, 1985). If this were the case both maximum grip aperture (to depth and width) and maximum wrist velocity would be expected to decrease when grasping the target object. Alternatively, improvements could be restricted to object shape alone (Brenner & Landy, 1999, Brenner & van Damme, 1999, Landy & Brenner, 2001, Todd & Norman, 2003), in which case only maximum grip apertures would be affected, not maximum wrist velocity.

Brenner and Landy (1999) found that stereo-motion information in a flanking object only improved the perceived shape of a static target object, and only if the target and flanking object were set to be the same size and at the same distance. If this were the case maximum grip aperture to the target object would only improve when the target and flanker were of the same size. In contrast, Econopouly (1996) found that when the objects were clearly presented at the same distance they did not have to be the same depth as one another for the addition of motion in the flanking object to improve the perceived shape of the target object. As has been discussed above, it is possible that the nature of the task led to such restricted effects in Brenner and Landy (1999), so improvements in perceived shape and distance could be evident for all target object shapes.

Discussion has so far focused on the improvements that could be gained by having a binocularly viewed rotating flanking object in the scene, but it is also possible that the remaining flanking object conditions could provide useful information for the perception of distance and shape. It is possible that a binocularly viewed flanking object, whether static or rotating, might provide useful relative disparity information, an improved stimulus for vergence and an improved source of vertical disparity information. Binocular and monocular flanking objects might also provide information about the orientation of the support surface, which could be used to better estimate object distance, from for example height in the field (Gardner & Mon-

Williams, 2001, Watt & Bradshaw, 2002). This improved distance estimate could in turn improve estimates of object shape, and might be evident in kinematic indices.

5.2 Method

5.2.1 Observers

Five observers took part in the experiment. Observers PBH, PS and SJW were experienced psychophysical observers and aware of the purposes of the experiment. Observers AW and CT had limited familiarity with psychophysics and were naïve to the purposes of the experiment. All observers were right-handed except for observer SJW who was left-handed; handedness was assessed by verbal report. All observers reached and grasped with their right hand, observer SJW who was left-handed found this completely natural and comfortable to do.

5.2.2 Stimuli and apparatus

The stimuli were wooden cylinders 10cm in height; each was painted matt black, and painted with randomly positioned blobs of light chargeable luminous paint on their surface and top. The density of the blobs was $\approx$0.4 blobs/cm and each blob was 3-4mm in diameter. The objects were light charged and viewed in the dark, the rationale behind this type of stimulus is to try and isolate the role of binocular information as much as possible (c.f. a random dot stereogram) whilst using real world objects (Watt & Bradshaw, 2000). Observers reached and grasped a target cylinder presented either on its own or in the presence of a flanking cylinder. Target cylinders had a fixed dimension (5.8 cm) and a variable dimension (3.9, 4.8, 5.6, 6.6, or 7.8 cm), so ranged from being squashed to stretched ellipses, with one being approximately circular. The variable dimension allowed either the depth or width of the cylinders to be varied whilst holding the other dimension constant, by positioning the cylinders rotated through 90 degrees. The flanking cylinder was always circular (5.8cm diameter) and had the same height and dot density as the target cylinders.

The experimental set-up is shown in Figure 5.2. Observers sat at a table covered with matte black material, their head was stabilised using a chin rest at the edge of the table. Eye height was 46cm from the table surface. The cylinders sat on two custom built circular platforms; the left platform could rotate, whereas the right platform remained stationary. The two platforms had matte black surfaces and were housed in a matt black box; this also housed a LabJack U12, which was used to computer control the platform's rotation. The height of the platforms surface above the table surface was 9cm; this resulted in the base of the cylinders being 37cm below eye height. The target cylinder was positioned on the centre of the right hand platform 8cm to the right of the observers mid-line and was always viewed binocularly. The flanking cylinder, when present, was placed on the centre of the left hand platform 8cm to the left of the subject's mid-line at the same distance as the target cylinder. There was therefore a fixed distance of 16cm between the centres of the target and flanking cylinders, when both were present. The cylinders were viewed at 30, 40 or 50cm, as measured from the edge of the table to the centre of the platforms, and were always presented at the same distance as one another. A square raised 'start' pad was attached to the table surface directly in front of the cylinder to be grasped.

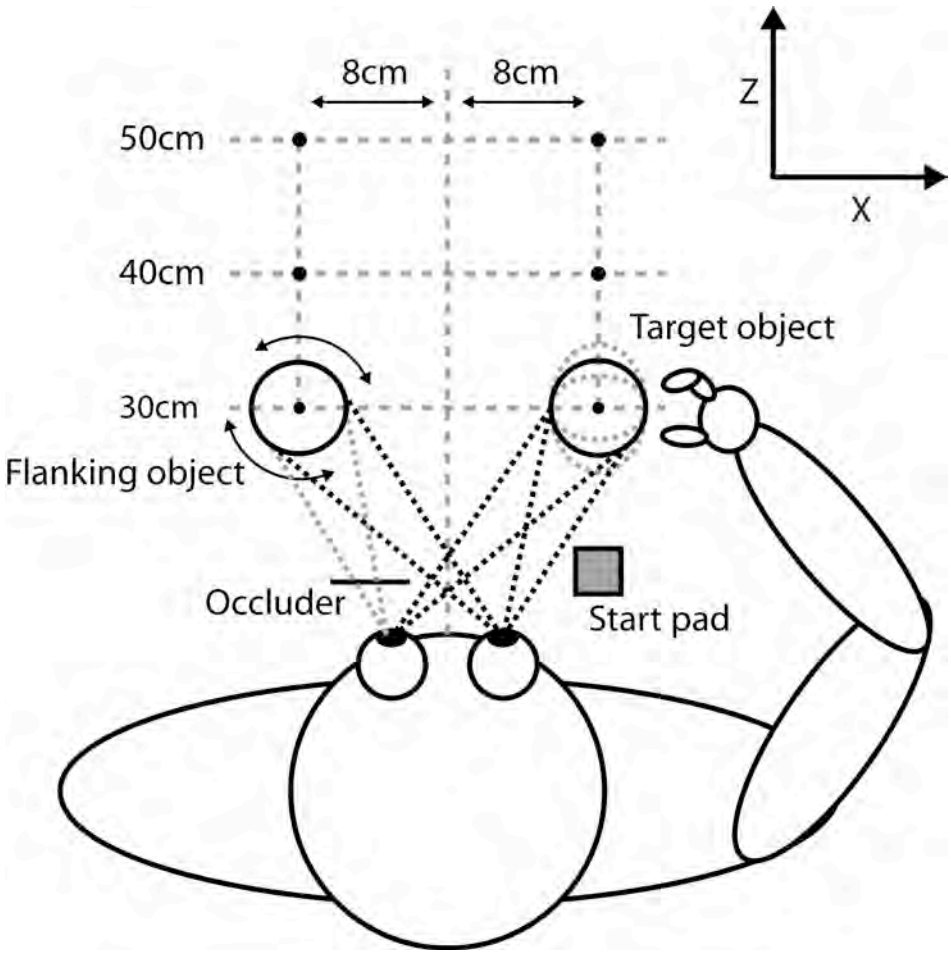

Figure 5.2: Diagram showing the experimental set-up. Observers reached and grasped a target object to the right of their midline in the presence of absence of a flanking object at one of three distances. The target object was always static and viewed binocularly, whereas the flanking object could be viewed either binocularly or monocularly (with the use of the occulder) and was either static or rotating. For further details see the accompanying text.

In total there were five viewing conditions, in one condition the target object was presented alone, whereas in the remaining four conditions the flanking object was also present. The target cylinder was always static and was always viewed binocularly, meaning that the information available from the object to be grasped remained constant across all viewing conditions. The flanking object, when present, was viewed either monocularly or binocularly and was either static or rotating. The five viewing conditions were therefore: (1) ***Target object only***: in this condition only the binocularly viewed static target object was present, (2) ***Monocular static flanker***: in addition to the binocularly viewed static target object, a monocularly viewed static flanking object was present (3) ***Monocular rotating flanker***: in addition to the binocularly viewed static target object, a monocularly viewed rotating flanker was present, (4) ***Binocular static flanker***: in addition to the binocularly viewed static target object, a binocularly viewed static flanker was present and (5) ***Binocular rotating flanker***: in addition to the binocularly viewed static target object, a binocularly viewed rotating flanker was present.

For each of the five viewing conditions observers grasped either the depth or width of the target cylinders in separate blocks of trials, depth was defined as the extent of the cylinder in the z- dimension, width the extent of the cylinder in the x- dimension (Figure 5.2). Therefore, in total there were 10 blocks of trials (2 grasp orientations x 5 viewing conditions). Each block of trials consisted of five reaches to each of the 15 object distance combinations (3 distances x 5 objects) i.e. 75 reaches in total per block. The ten blocks of trials were completed in a randomised order over a period of 1-2 weeks and within each block of trials object and distance conditions were presented in a randomised order. Observers AW, CT, PS and SJW completed all ten blocks of trials, observer PBH completed only the 5 blocks in which object depth was grasped.

The stimuli were light charged prior to starting a block of trials and topped up throughout the experiment. The exact duration of the light charging was not controlled, but the objects were clearly viewable and did not give off sufficient light to illuminate the platform surfaces. Light breaks were taken as needed for comfort and to avoid dark adaptation. When the flanking object rotated it did so at 0.25Hz, randomly either clockwise or counter clockwise. This rotation speed was chosen to be concordant with previous research demonstrating information integration across objects (Brenner & Landy, 1999). Concurrent monocular viewing of the flanking cylinder and binocular viewing of the target cylinder was achieved by positioning a matt black occluder in front of the left eye (Figure 5.2). Positioning of the occluder was carried out prior to the appropriate block of trials. The experimenter adjusted the position of the occluder until the observer reported being able to see the left hand cylinder monocularly (with the right eye) and the right hand cylinder binocularly, at each of the three viewing distances. Observers were therefore aware of the placement of the occluder on a given block of trials. However, subjectively, the conditions in which the occluder was present felt natural, this was likely due to both eyes being open and monocular and binocular regions of the scene being gaze contingent.

5.2.3 Movement recording

Prehensile movements were recorded using a two camera ProReflex motion tracking system running at 240Hz. This system was calibrated using a wand and frame calibration provided with the ProReflex prior to each block of trials. Three passive markers were attached to the right hand, one on the top of the thumbnail, one on the top of the first fingernail and one on the wrist. Another marker was positioned on the top of the object and one as a reference on the table surface to the left of the apparatus. Cameras were positioned so as to get a sufficient field of view of the scene for accurate recording and so as to minimise the risk of a hand marker being occluded by the flanking object when it was present.

5.2.4 Procedure

The experiment was conducted in the dark, the observer sat at a table with their eyes closed and their head supported in a chin rest with their right hand positioned on the start pad. A small desk lamp was used whilst the experimenter placed the target object in the appropriate orientation on the right hand platform and the flanking object on the left hand platform, in those blocks of trials where it was required. Accurate placement of the objects on the platforms was aided by alignment marks on each platform and object. Similarly, markers on the table surface allowed accurate positioning of the box containing the two object platforms at each of the three viewing distances. The markers on the objects, platforms and table were not visible to the observer during the experiment.

Once the desk lamp was extinguished the start of the trial was triggered and the ProReflex began recording. The observer heard three ‘clicking’ sounds, spaced 1.5 seconds apart, these signalled a ‘3’-‘2’-‘1’ countdown until they could commence the reach and grasp movement. In those blocks of trials when the flanking object rotated it began doing so 1 second before the first ‘click’. Prior to starting the experiment observers were instructed that upon hearing the first click they should open their eyes and “look across the scene”, beyond this no specific instructions were given to observers as to where they should look or fixate once they opened their eyes. Similarly, observers were told that upon hearing the third click they should reach and grasp the target object.

There was therefore a pre-movement viewing time of ≈ 3 seconds depending on when observers initiated their movement. Piloting showed this to be a comfortable time in which observers could look across the scene without feeling rushed before having to make the reach and grasp movement. On the third click observers reached out and picked up the target object using a precision grip (first finger and thumb only) across either its depth or width dimension and put it down in the vicinity of the starting position. Once they had done this they repositioned their hand on the start pad and closed their eyes whilst the experimenter set up the next object and distance. On some trials observers reported pre-empting the third click and initiating movement early. Similarly, in some instances the observer grasped the object across the wrong dimension or collided with the object. These trials were labelled void and repeated at the end of the session.

Observers were aware before commencing the experiment which block of trials would contain a flanking object and whether the flanking object would be static and or rotating. The nature of the precision grip was demonstrated to the observer prior to starting the experiment, so that they were aware of the dimension of the cylinder they would be grasping. Observers are known to prefer grasping across the depth and width of cylindrical objects when given no specific instructions as to how they should pick an object up (Cuijpers et al., 2006). No instructions were given as to the height of the object at which observers should grasp. Piloting had shown that observers tended to grasp near to the top of the cylinders, likely due to this providing a stable grasp by which to apply grip force and lift the objects. This was beneficial as it reduced the chance of finger and thumb markers being occluded.

5.2.5 Data analysis

Data analysis was carried out in two stages. Firstly, the 2D camera data were tracked into 3D x, y and z coordinates using the Qualysis Track Manager software. The x, y, and z coordinates were then exported and dependent measures obtained using custom written code running in the Matlab programming environment. The principle indices of interest were maximum grip aperture (MGA) and maximum wrist velocity (MAXV). Both MGA and MAXV occur in between the start of the movement and the point of object contact. When grasping an object grip aperture increases as the prehensile movement progresses reaching a maximum about 70% of the way through the movement, after this point the grip aperture decreases as the digits come into contact with the object. It has been shown that maximum grip aperture is linearly related to the grasped dimension of an object (Jeannerod, 1988). Maximum wrist velocity increases as the hand is accelerated from a stationary start position and reaches a maximum velocity before the point of maximum grip aperture, the hand then decelerates as it comes into close proximity of the object to be grasped (see Figure 5.3 in the results section for example grip aperture and velocity profiles).

Grip aperture was defined as the 3-D distance between the finger and thumb markers. Wrist velocity was calculated using a sliding window of two consecutive frames for the duration of the recording. In some instances the view of the wrist, finger or thumb marker was lost to one or both of the cameras, this caused gaps in the grip aperture and wrist velocity traces. No attempt was made to interpolate over these gaps, as it was unclear what type of gap filling would be appropriate. If gaps in the recording were filled it is possible that some of the derived indices such as maximum grip aperture (MGA) and maximum wrist velocity (MAXV) would fall within the gap filled region. These indices would then be based on interpolated data, which might not represent the true values of MGA or MAXV. Each section of the grip aperture and velocity traces was filtered with a dual-pass zero-lag Butterworth filter with an 8Hz cut-off.

The dependent measures of MGA and MAXV were obtained from the filtered wrist velocity and grip aperture profiles. Peaks in the grip aperture and wrist velocity profiles were identified as zero crossings in the second derivative of each profile. The largest peak value in each profile was identified as the initial MGA and MAXV respectively. The grip aperture and velocity profiles for all trials were visually checked to make sure that the correct peak for the MGA and MAXV had been identified. A number of factors could cause the largest peak in the profile not to be the MGA or MAXV. Firstly, if the observer reached and grasped the object, but then re-adjusted their grasp before lifting the object, the grip re-adjustment might give a larger peak grip aperture then for the true MGA index. Secondly, because gaps in the data were not filled / interpolated each data section had to be filtered separately. The Butterworth filter added false peaks when it reached the end of each data section. Therefore, if a gap were near the true position of the MGA or MAXV it was possible for one of these false peaks to be incorrectly identified as one of these indices.

Finally, false peaks in the data could occur when markers temporarily merged, this happened when one of the cameras views of a marker could not be disambiguated from that of another marker. Those instances where merged markers, or gaps, affected the region of the profile from which MGA or MAXV was obtained, the trial was discarded for that kinematic index. The final value for the MGA was corrected for the width of the finger and thumb. This correction was carried out by first obtaining the terminal grip aperture (grip aperture when the object is grasped) for trials with the circular target cylinder. The diameter of the cylinder was then subtracted from this value to give the width of the finger and thumb. Only the circular target cylinder was used as the terminal grip aperture with this object would be independent of the orientation with which the object was grasped, this would not be the case with the elliptical objects.

In some instances the observer knocked into the object when attempting to grasp it. In these instances the trial was discarded and repeated at the end of the session. An observer colliding with an object could be could be taken evidence for observer misestimating the distance of the object (e.g. Bradshaw et al., 2004). However, in the present study object collision occurred very rarely. Across all conditions and observers there were only 12 instances of object collision, this represented 0.64% of the total number of trials. Incidence of object collision ranged from 0 to 1.3% of trials across viewing conditions. Interestingly all twelve instances occurred when the observer grasped the depth of the object. Given the very low incidence of object collision this index was not analysed any further.

5.3 Results

5.3.1 General Results and analysis

Example grip aperture and velocity profiles can be seen in Figure 5.3. The grip aperture profile is characterised by a rapid opening of the hand up to the point of maximum grip aperture and a more gradual closing of the hand around the object (Figure 5.3a). The wrist velocity profile is characterised by a rapid increase in velocity to its maximum, which is achieved before the point of maximum grip aperture, wrist velocity then decreases and plateaus as the hand nears and comes into contact with the object (Figure 5.2b). During this plateau there are only small movements of the wrist before its velocity increases as the object is picked up. Figure 5.4 shows an example of the scaling of grip aperture with object size and wrist velocity with object distance. As would be expected grip apertures linearly increased with object size (Figure 5.4a) and wrist velocities linearly increased with object distance (Figure 5.4b), this represents the normal scaling found with these parameters (Jeannerod, 1988).

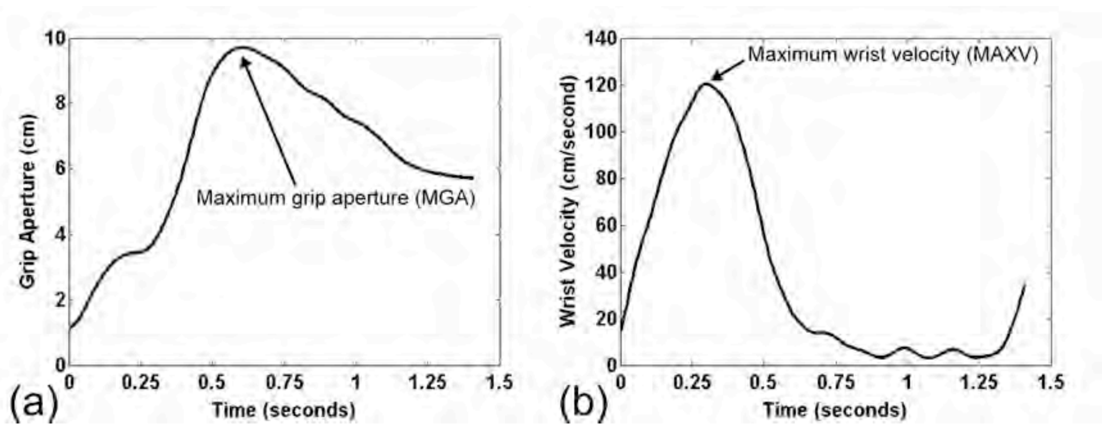

Figure 5.3: Example Grip aperture (a) and wrist velocity (b) profiles taken from observer PS when grasping across the depth of the target cylinder in the presence of a binocularly viewed

static flanker. Time on the ordinate axis represents time from the beginning of the prehensile movement. Maximum grip aperture (MGA) and maximum wrist velocity (MAXV) are indicated on the plots.

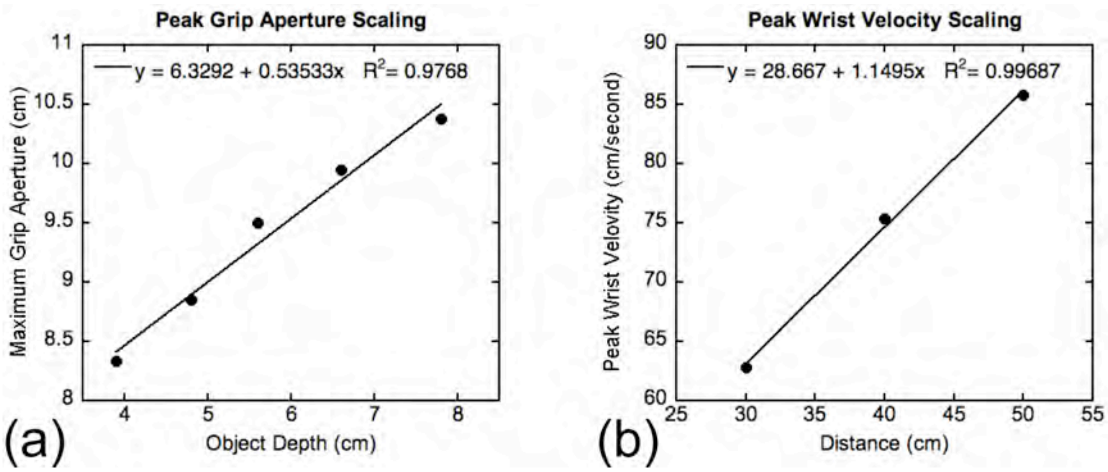

Figure 5.4: Graphs showing grip aperture scaling with object size (a) and wrist velocity scaling with object distance (b). Linear fits to the data are plotted and the equation of this fit shown. As can be seen grip aperture linearly increases with object depth with a gain of around 0.5 and peak wrist velocity increases linearly with object distance with a gain near to 1. Data are taken from observer CT when grasping the depth of an object in the presence of a binocularly viewed rotating flanker.

Individual plots showed that there were large differences between observers as to the effect of viewing condition on MGA and MAXV. Large individual differences are typically the case in 3-D shape judgement tasks (Champion et al., 2004, Glennerster et al., 1996, Oruc et al., 2003, Todd & Norman, 2003). Significant differences between observers were also found by Brenner and Landy (1999), which suggests that differences may exist in the extent to which observers use or weight the information available from other objects in the scene. Individual observers' data were therefore analysed separately with univariate ANOVA with object, distance and viewing condition as fixed factors. Where necessary main effects and interactions were investigated further with pairwise comparisons using simple effects analysis with Sidak corrections.

5.3.2 *Reaching and grasping object depth*

The data for grasping the depth of an object for each observer are plotted in Figure 5.5; the statistics for the analysis of MGA are shown in Table 5.1 and for the analysis of MAXV in Table 5.2. All observers showed a main effect of distance on MAXV and object size on MGA (Table 5.1). Paired comparisons showed that MGA increased with object depth and maximum wrist velocity increased with object distance, as would be expected. All observers also showed a significant effect of condition on MGA and MAXV, except for observer PBH who only showed an effect of condition on MAXV. The effect of condition differed greatly between individuals (Figure 5.5). Paired comparisons showed that for observers CT and PS the addition of another object in the scene resulted in decreased grip apertures compared to those found with no flanking object. This was more prominent for binocularly viewed objects (stationary or moving) for CT, whereas PS showed no clear effect of the type of information available from the flanking object e.g. binocular versus monocular, or static versus rotating.

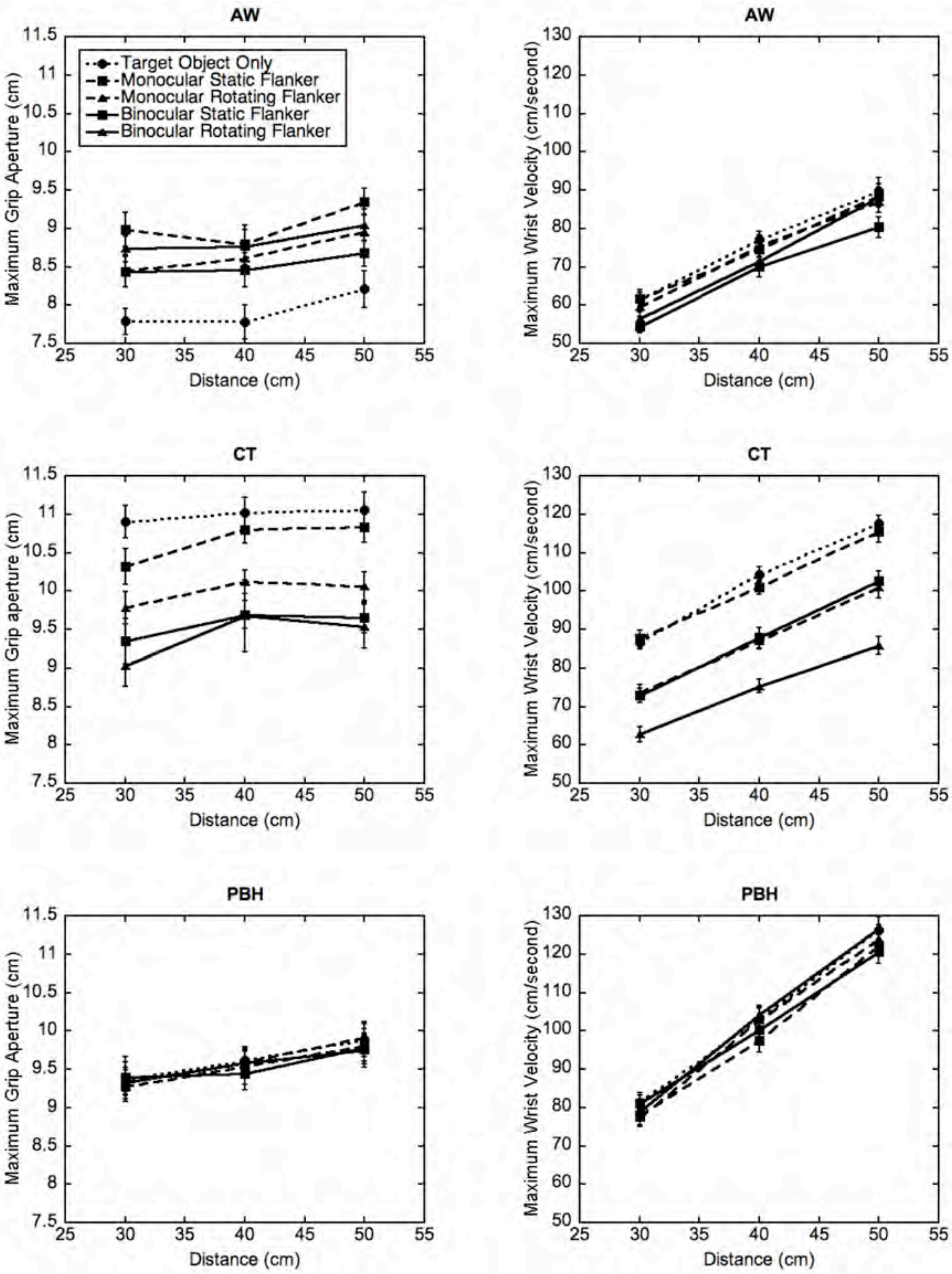

Figure 5.5: Showing average maximum grip aperture (left column) and maximum wrist velocity (right column) across objects when grasping an object's depth dimension (for observers AW, CT and PBH). Error bars show standard error of the mean.

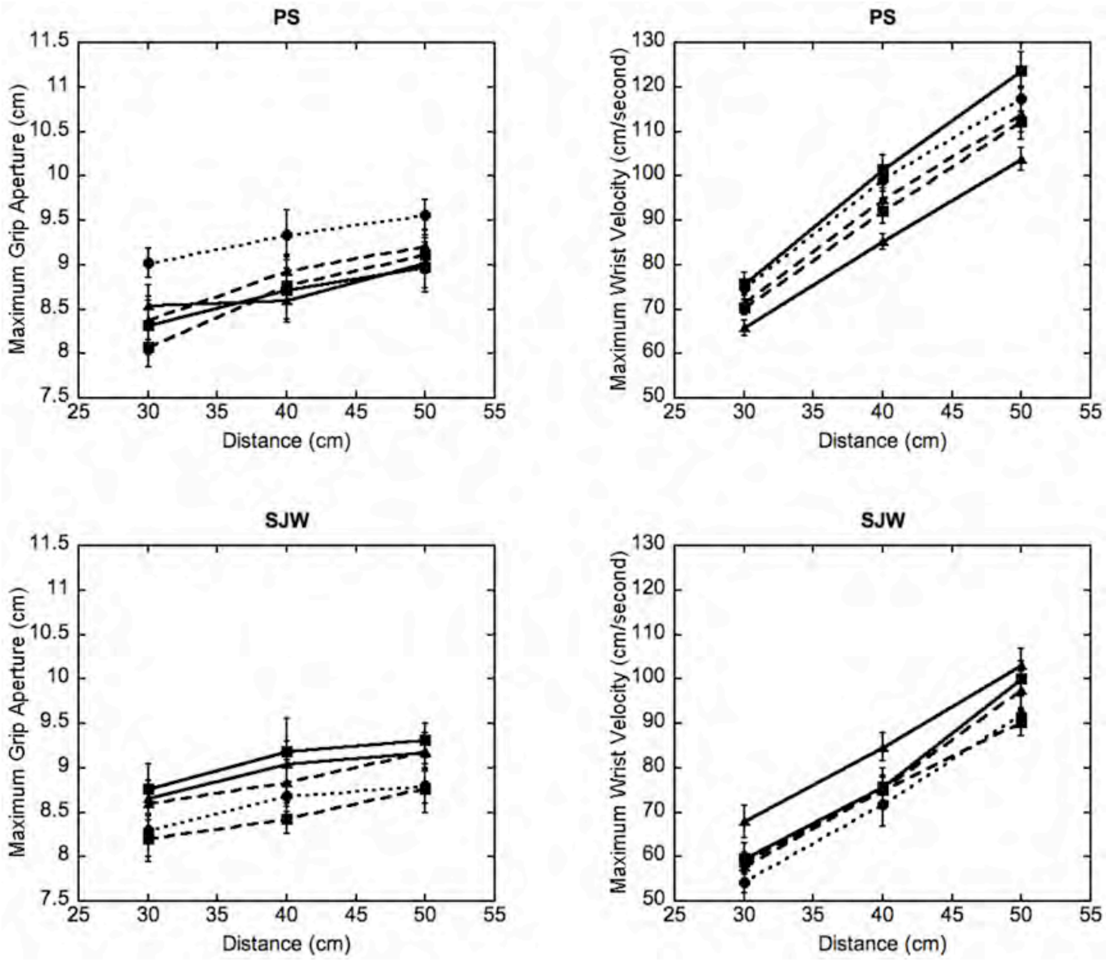

Figure 5.5 (continued): Showing average maximum grip aperture (left column) and maximum wrist velocity (right column) across objects when grasping an objects depth dimension (for observers PS and SJW). Error bars show standard error of the mean.

A similar, but not identical, pattern of results was found for MAXV for both observers, this was broadly consistent with a distance scaling explanation i.e. decreased grip apertures coupled with slower reaches. For CT the presence of a flanking object in the scene resulted in slower reaches compared to those with no flanking object in all conditions except for the monocularly viewed static flanker condition. The slowest wrist velocities were found with a binocularly viewed rotating flanking object, whereas those found with a binocularly viewed static flanker did not differ from those found with a monocularly viewed rotating flanker. For PS the presence of a flanking object resulted in slower reaches, the slowest being produced in the binocular rotating flanker condition. The fastest reaches by contrast were found with a binocularly viewed static flanking object. Overall, for these observers the addition of a flanking object, on average, resulted in slower reaches with smaller grip apertures, but there was no consistent effect regarding the type of information available from the flanking object.

Observer SJW showed the opposite pattern of results, on average the presence of a flanking object resulted in larger grip apertures and faster wrist velocities. Paired comparisons showed that the presence of a binocularly viewed flanker, whether rotating or static, resulted in significantly larger MGA's compared to those with the target object alone. By comparison MGA's with a monocularly viewed flanking object did not differ from those with the target object alone. Maximum wrist velocities in all conditions bar that of a monocularly viewed static flanker were greater than those found with the target object alone, with the fastest reaches being produced in the presence of a binocularly viewed rotating flanker. Although this pattern of results is also consistent with a distance scaling explanation, it is not in the direction predicted. The perceptual literature would suggest that near distances and depths would be overestimated (Johnston, 1991), so additional information about distance and shape would be expected to results in decreased grip apertures and wrist velocities (Servos et al., 1992).

For AW the presence of a flanking object resulted in increased grip apertures, with no clear differentiation regarding the type of information available from the flanking object. Maximum wrist velocities decreased with the presence of a flanking object compared to the target object alone, but only in the presence of a binocularly viewed flanking object. This pattern of results in MGA and MAXV is not consistent with a distance scaling explanation. Observer PBH showed the least effect of condition, no effect was found for grip apertures and only a restricted effect for wrist velocities. Wrist velocities to a target object with a monocularly viewed static flanker were smaller than those to both target objects alone and those with a binocularly viewed rotating flanker. No other conditions differed from one another. The highly restricted effect of condition in this observer is inconclusive regarding any distance scaling explanation.

All observers also showed a main effect of distance on MGA, paired comparisons showed that this was consistent with an increase in grip apertures with increasing distance. All observers bar AW also showed an effect of object on MAXV, paired comparisons showed that this resulted from observers' faster reaches to the largest 7.8cm object compared to one or more of the other objects. It is not clear why these main effects should be found; it could possibly result from a lack of distance scaling. In a number of instances the main effects of distance, object depth and viewing condition observed for both MGA and MAXV were further qualified by interactions. Observers AW, CT, PS and SJW showed interactions for MGA and observers CT, PBH and PS showed interactions for MAXV.

For the MGA index, observers AW and SJW showed a condition by object interaction, PS a distance by object interaction and CT a three-way condition by distance by object interaction. These interactions are not interpretable in terms of theory or consistent across observers. For example, both AW and SJW showed a condition by object interaction, for SJW this resulted from no effect of distance for the object with a 6.6cm depth, but significant effects of condition for all other objects. In contrast AW showed effects of condition for all objects, but these effects were clearest for the 4.8 and 5.6cm deep objects. PS showed a distance by object interaction, the most evident cause of this interaction was that there was no effect of distance on MGA for the 7.8cm deep object, but significant effects between some distances for all other objects. No other observers showed this interaction.

The interactions for the MAXV were generally not the same as those for MGA; similar interactions would be expected if MAXV and MGA could be interpreted as direct indices of perceived distance and shape. Observer CT showed both a condition by object interaction and a three-way condition by distance by object interaction, PBH a distance by object interaction and PS a condition by distance interaction. Again these interactions do not seem interpretable from any theoretical standpoint and are not consistent across observers. For example, for PS there seemed to be a clearer effect of condition on MAXV for the 40 and 50cm distances, but this was not true for any other observer. For PBH paired comparisons showed that MAXV increased with distance for all objects, but the magnitude of this effect differed. The only interaction that was found for both MGA and MAXV for any observer was the three-way distance by condition by object interaction for observer CT. However, the nature of this interaction differed for MGA and MAXV and was not theoretically interpretable in either case.

	Condition	**Distance**	**Object**	**Condition x Distance**	**Condition x Object**	**Distance x Object**	**Condition x Distance x Object**
AW	F(4,296)=55.64 p<0.001	F(2,296)=23.36 p<0.001	F(4,296)=317.35 p<0.001	F(8,296)=0.76 ns	F(16,296)=3.3 p<0.001	F(8,296)=1.24 ns	F(32,296)=0.74 ns
CT	F(4,296)=102.69 p<0.001	F(2,296)=16.81 p<0.001	F(4,296)=97.30 p<0.01	F(8,296)=0.87 ns	F(16,296)=0.89 ns	F(8,296)=0.90 ns	F(32,296)=1.76 0.01
PBH	F(4,280)=0.58 ns	F(2,280)=31.71 p<0.001	F(4,280)=143.08 p<0.001	F(8,280)=0.36 ns	F(16,280)=1.58 ns	F(8,280)=1.61 ns	F(32,280)=1.35 ns
PS	F(4,293)=16.42 p<0.001	F(2,293)=47.60 p<0.001	F(4,293)=259.88 p<0.001	F(8,293)=1.46 ns	F(16,293)=1.38 ns	F(8,293)=2.08 p<0.05	F(32,293)=1.32 ns
SJW	F(4,212)=12.06 p<0.001	F(2,212)=22.40 p<0.001	F(4,212)=145.69 p<0.001	F(8,212)=0.28 ns	F(16,212)=2.66 p<0.001	F(8,212)=0.52 ns	F(32,212)=0.78 ns

Table 5.1: Showing the main effects and interactions for maximum grip aperture (MGA) when grasping across the depth of an object. Cell shading denotes the significance level, dark grey indicates at least p<0.01, light grey indicates p<0.05 and clear cells indicate non-significance.

	Condition	Distance	Object	Condition x Distance	Condition x Object	Distance x Object	Condition x Distance x Object
AW	F(4,296)=19.19 p<0.001	F(2,296)=666.16 p<0.001	F(4,296)=1.55 ns	F(8,296)=1.40 ns	F(16,296)=0.88 ns	F(8,296)=0.87 ns	F(32,296)=1.35 ns
CT	F(4,296)=379.48 p<0.001	F(2,296)=912.76 p<0.001	F(4,296)=3.28 p<0.05	F(8,296)=2.69 p<0.01	F(16,296)=1.74 p<0.05	F(8,296)=1.24 ns	F(32,296)=1.52 p<0.05
PBH	F(4,294)=5.04 <0.001	F(2,294)=1289.52 p<0.001	F(4,294)=4.90 p<0.001	F(8,294)=1.89 ns	F(16,294)=1.45 ns	F(8,294)=2.52 p<0.01	F(32,294)=1.37 ns
PS	F(4,299)=60.57 p<0.001	F(2,299)=1335.99 p<0.001	F(4,299)=7.84 p<0.001	F(8,299)=2.14 p<0.05	F(16,299)=0.79 ns	F(8,299)=0.31 ns	F(32,299)=0.96 ns
SJW	F(4,293)=25.31 p<0.001	F(2,293)=602.14 p<0.001	F(4,293)=6.59 p<0.001	F(8,293)=1.63 ns	F(16,293)=0.72 ns	F(8,293)=1.30 ns	F(32,293)=0.92 ns

Table 5.2: Showing the main effects and interactions for maximum wrist velocity (MAXV) when grasping across the depth of an object. Cell shading denotes the significance level, dark grey indicates at least p<0.01, light grey indicates p<0.05 and clear cells indicate non-significance.

5.3.3 Reaching and grasping object width

The data for grasping across the width of an object for each observer are shown in Figure 5.6, the statistics for MGA are shown in Table 5.3 and for MAXV in Table 5.4. Note that observer PBH did not complete the trials in which object width was grasped, there is also no wrist velocity data shown for observer AW, this is because across all conditions tracking of the wrist marker was lost for large sections of the reach encompassing the point of MAXV in most trials. This made valid comparisons across object, distance and condition impossible. As was the case when grasping object depth, main effects of width on maximum grip aperture (Table 5.3) and distance on maximum wrist velocity (Table 5.4) were found for all observers. Paired comparisons confirmed that MGA increased with object width and MAXV increased with object distance for all observers. All observers also showed a significant effect of condition on MGA (Table 5.3) and MAXV (Table 5.4). Broadly speaking the presence of another object in the scene tended to result in increased grip apertures and decreased wrist velocities, but again there was large variation between observers.

For observer CT increased MGA's were found with the presence of a flanking object, compared to a target object alone in all conditions except for with a binocularly viewed rotating flanking object. There was some tendency for larger grip apertures to be found with a monocularly viewed flanking object, but grip apertures with a monocular static flanker and a binocular static flanker did not differ. MAXV's were slower with a monocularly viewed flanking object present compared to the target object alone, whether this flanker was static or rotating. For a binocularly viewed flanking object MAXV was greater compared to that with the target object alone, but only when the binocularly viewed flanker was static. As with observer CT, AW showed larger grip apertures in the presence of a flanking object compared to the target object alone, but there was no clear effect of the type of information available in this flanking object. No wrist velocity data were available for this observer.

PS also showed larger grip apertures in the presence of a flanking object but only when this flanking object was viewed monocularly. Wrist velocities were slower in the presence of a flanking object in all cases, with this effect being greater for a monocularly viewed flanking object, whether static or rotating. Overall observers CT, PS and AW tended to show slower wrist velocities and larger grip apertures when a flanking object was present regardless of the type of information available from the flanking object, this is consistent with observers adopting a conservative reaching strategy when there was another object in the scene. The data for observer SJW were less consistent with this explanation. For this observer MGA was smallest to target objects in the presence of a monocularly viewed rotating flanker, whereas the remaining conditions did not differ from one another. MAXV was slowest to targets in the presence of a monocularly viewed rotating flanking object compared to all conditions, except for the binocularly viewed static flanker, from which they did not differ. No other conditions differed from one another.

As was the case with grasping across the depth dimension, there was a significant main effect of distance on MGA and a significant main effect of object on MAXV, for some observers. Observers AW, CT, and SJW showed a main effect of distance on MGA, paired comparisons showed that this resulted from smaller MGA's for objects presented at the 30cm distance compared to the 40cm distance (CT and SJW) or both (AW). Only observer SJW showed a main effect of object on MAXV, this effect was of marginal significance (Table 5.1) and paired comparisons failed to locate differences in MAXV between objects. This is likely due to the marginal significance of the ANOVA main effect and the decreased power of paired comparisons with Sidak corrections to locate the effect. Compared to grasping depth these main effects seem less likely to be due to incomplete distance scaling.

There were also a number of interactions between distance, object width and condition, as can be seen from Tables 5.3 and 5.4 these were mainly related to MGA. As was the case with grasping across object depth these interactions show little consistency across observers. As an example, all observers showed a significant condition by object interaction for MGA. For observer AW this was predominantly because the three smallest objects showed the more prominent effect of condition. For CT the effect of condition is smaller with the largest object, whereas the rest of the objects showed similar effects of condition. SJW showed no effect of condition for the 3.9, 5.6 and 6.6cm width objects, and PS a less prominent effect of condition for the 5.6 and 6.6cm width objects. As another example, both AW and CT showed a distance by object interaction on MGA. For CT this was because an effect of distance was only seen for the smallest object, whereas for AW this arose because there was no effect of distance for largest width object. There was only one significant interaction across observers regarding MAXV, this was a condition by distance interaction for observer CT, and arose due to there being a larger effect of condition at the 40 and 50cm distances.

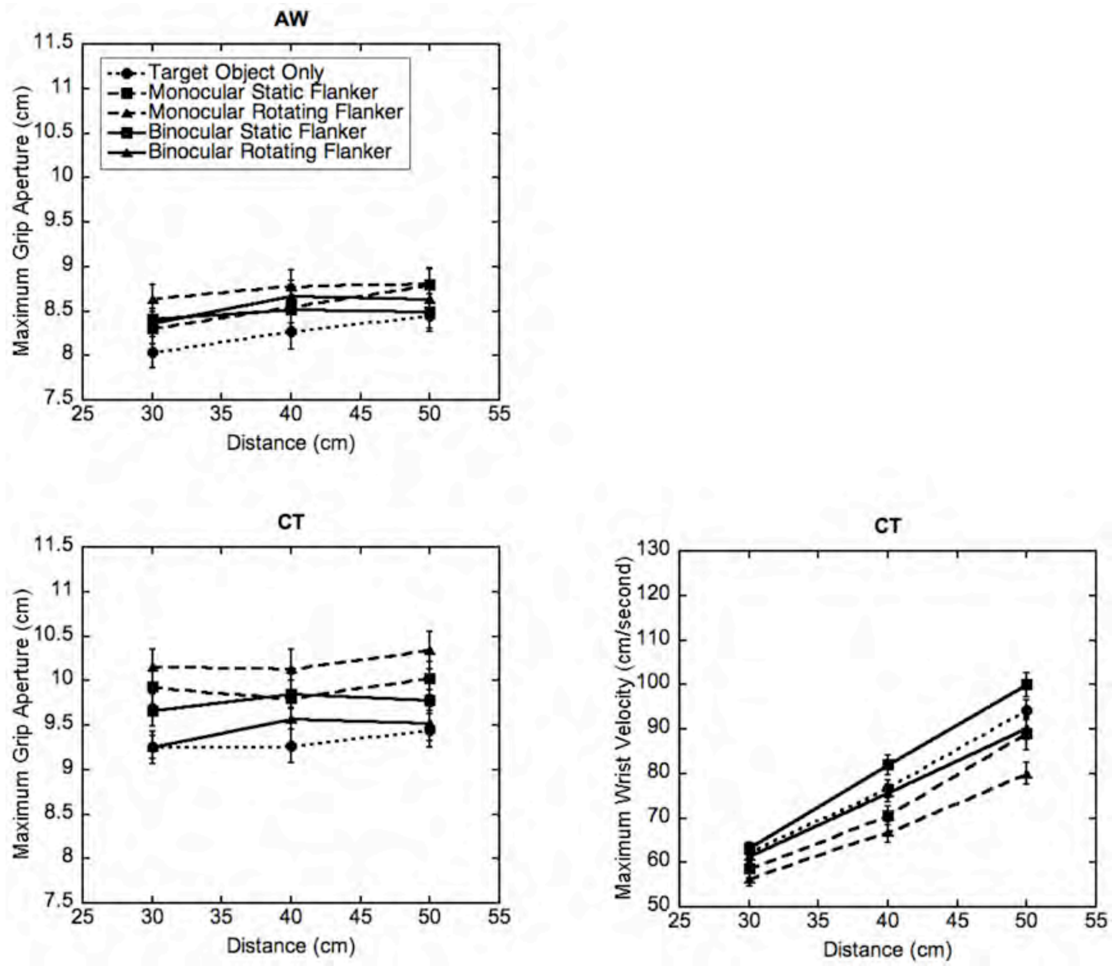

Figure 5.6: Showing average maximum grip aperture (left column) and maximum wrist velocity (right column) across objects when grasping an objects width dimension (for observers AW, and CT). Error bars show standard error of the mean. No wrist velocity plot is shown for observer AW due to the large amount of trials that needed to be discarded due to wrist marker tracking being lost over the point of peak wrist velocity. See accompanying text for more details.

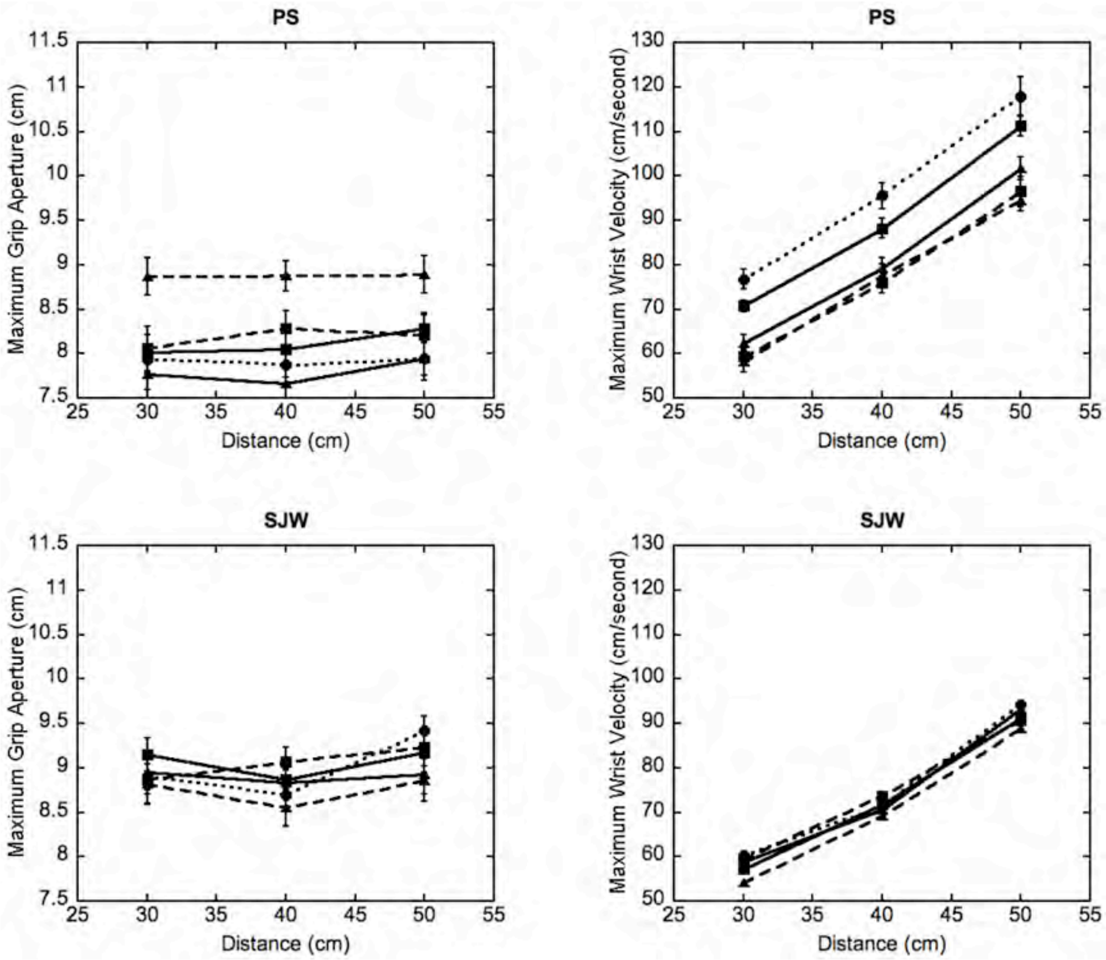

Figure 5.6 (continued): Showing average maximum grip aperture (left column) and maximum wrist velocity (right column) across objects when grasping an objects width dimension (for observers PS, and SJW). Error bars show standard error of the mean.

	Condition	Distance	Object	Condition x Distance	Condition x Object	Distance x Object	Condition x Distance x Object
AW	F(4,295)=12.73 p<0.001	F(2,295)=15.05 p<0.001	F(4,295)=476.99 p<0.001	F(8,295)=1.10 ns	F(16,295)=1.71 p<0.05	F(8,295)=2.48 p<0.05	F(32,295)=1.06 ns
CT	F(4,298)=48.85 p<0.001	F(2,298)=4.45 p<0.05	F(4,298)=257.81 p<0.001	F(8,298)=1.17 ns	F(16,298)=1.78 p<0.05	F(8,298)=2.39 p<0.05	F(32,298)=1.23 ns
PS	F(4,294)=63.48 p<0.001	F(2,294)=2.74 ns	F(4,294)=412.27 p<0.001	F(8,294)=1.01 ns	F(16,294)=3.85 p<0.001	F(8,294)=0.88 ns	F(32,294)=1.09 ns
SJW	F(4,299)=5.58 p<0.001	F(2,299)=13.03 p<0.001	F(4,299)=219.22 p<0.001	F(8,299)=2.65 p<0.01	F(16,299)=2.12 p<0.01	F(8,299)=0.81 ns	F(32,299)=1.18 ns

Table 5.3: Showing the main effects and interactions for maximum grip aperture (MGA) when grasping across the width of an object. Cell shading denotes the significance level, dark grey indicates at least p<0.01, light grey indicates p<0.05 and clear cells indicate non-significance.

	Condition	**Distance**	**Object**	**Condition x Distance**	**Condition x Object**	**Distance x Object**	**Condition x Distance x Object**
CT	F(4,299)=69.23 p<0.001	F(2,299)=955.09 p<0.001	F(4,299)=2.20 ns	F(8,299)=5.23 p<0.001	F(16,299)=1.43 ns	F(8,299)=1.19 ns	F(32,299)=1.47 ns
PS	F(4,296)=145.82 p<0.001	F(2,296)=1208.53 p<0.001	F(4,296)=0.19 ns	F(8,296)=1.29 ns	F(16,296)=0.70 ns	F(8,296)=0.50 ns	F(32,296)=1.17 ns
SJW	F(4,299)=4.72 p<0.001	F(2,299)=690.04 p<0.001	F(4,299)=2.44 p<0.05	F(8,299)=0.83 ns	F(16,299)=1.39 ns	F(8,299)=0.71 ns	F(32,299)=0.69 ns

Table 5.4: Showing the main effects and interactions for maximum wrist velocity (MAXV) when grasping across the width of an object. Cell shading denotes the significance level, dark grey indicates at least p<0.01, light grey indicates p<0.05 and clear cells indicate non-significance.

5.4 Discussion

5.4.1 General discussion

Beyond the typical scaling of MGA with object size (depth and width) and MAXV with object distance, there is a striking lack of consistency between observers as to the effect of viewing condition on reach and grasp parameters. There is also a lack of consistency between the effects of viewing condition when grasping the depth and width of an object for the same observer. When grasping the depth of an object observers CT and PS on average showed decreased grip apertures and decreased wrist velocities when a flanking object was present compared to when reaching to the target object alone. In contrast SJW tended to show increased grip apertures and increased wrist velocities in the presence of a flanking object compared to a target object alone. These three observer's results are broadly consistent with a distance scaling explanation for the changes in kinematic indices, i.e. a decrease in grip aperture coupled with a decrease in wrist velocity and vice versa (Servos & Goodale, 1994, Servos et al., 1992).

However, it is clear that not all flanking object conditions resulted in a similar change in kinematic indices, and there is no clear pattern as to those conditions where no effects of a flanking object are found. Previous research suggests that the conjunction of stereo and motion information in the flanking object might be expected to be particularly useful, given that this offers the possibility of recovering metric shape and distance information (Landy & Brenner, 2001, Richards, 1985). It is also possible that a binocularly viewed flanking object may have been more beneficial than a monocularly viewed flanking object by, for example, providing improved relative disparity information. Despite this no clear effects of the information available from the flanking object were seen in these observers' data. This was also true for the other observers when grasping depth, who in addition showed results that were inconsistent with a distance scaling explanation.

Observer PBH showed very little effect of condition on both MGA and MAXV. Whilst this is not inconsistent with a distance scaling explanation, it is also not clear evidence for such an explanation. On average observer AW exhibited decreased wrist velocities coupled with increased grip apertures with the presence of a flanking object, this pattern of results is inconsistent with a distance scaling explanation. Overall the results for grasping depth do not support the view that information in the flanking object is being used to better estimate distance and shape, and that this improved information is used in the control of prehension to the target object. There were clearly effects of the presence of the flanking object, but where these were observed they were consistent with more general effects of the presence of the flanking object, rather than the visual information available from this object.

More global effects of the presence of the flanking object were also evident when observers grasped the width of the target object. In this instance there was more consistency across observers but none showed a pattern consistent with a distance scaling explanation. Observers AW, CT and PS tended to show increased grip apertures and decreased wrist velocities when another object was present in the scene (note that no wrist velocity data were available for observer AW). Observer SJW showed a far more limited effect of condition, the effects of condition were limited to decreased wrist velocities and decreased grip apertures in the presence of a monocularly viewed rotating cylinder only.

If the changes in MAXV and MGA across conditions were due to changes in perceived distance and perceived size (depth or width), then, all other things being equal, two predictions can be made regarding the effects of viewing condition on perceived depth, width and distance. Firstly, viewing condition would be expected to have the same effect on MAXV (perceived distance) when grasping either depth or width, this is because the viewing condition manipulation was the same in both instances. Secondly, viewing condition would be expected to have the same pattern of effects on MGA when grasping both depth and width, but of greater magnitude for grasping depth. This is because both perceived depth and width scale with perceived distance, but whereas width scales directly with distance, depth scales with the square of distance (see Chapters 1 and 3 for more details). These predictions were not born out in the data.

5.4.2 *Comparing prehension to depth and width*

Observers AW, CT, PS and SJW completed blocks of trials for grasping both depth and width. For observers CT and PS the presence of a flanking object tended to result in decreased MAXV when grasping both depth and width, but decreased MGA when grasping depth and increased MGA when grasping width. SJW showed increased MGA and MAXV when grasping depth and but no clear effects when grasping width. AW showed increased MGA when grasping both depth and width and decreased MAXV when grasping depth; no wrist velocity data were available when grasping width. Considering all observers' data together when grasping both depth and width, the presence of a flanking object resulted in five instances of increased MGA, two decreases in MGA and two instances of no clear effect. For MAXV the presence of a flanking object resulted in five instances of decreased MAXV, one increase in MAXV and two instances of no clear effect. Therefore the predominant effect of a flanker, if any, seemed to be an increase in grip apertures and a decrease in wrist velocities, more so when grasping width.

This is clearly a gross simplification of the data and it must be kept in mind that numerous interactions were found in both the depth and width data. These interactions arose from differing effects of the various flanker conditions, including no effects for some conditions, and more complex interactions, all of which were inconsistent across observers and inconsistent for the same observer when grasping depth and width. A pattern of increased grip apertures and decreased wrist velocities is consistent with a conservative strategy, whereby an increased margin of error is built into reaches by opening the hand wider and moving more slowly in conditions of uncertainty (Melmoth & Grant, 2006, Watt & Bradshaw, 2000). If this strategy were adopted when a flanking object were present it is surprising, as all flanking object conditions served to provide additional information to that available from the binocularly viewed target object, all be this very limited additional information in some conditions.

A conservative strategy toward programming of the prehensile movements has typically been seen in reduced visual conditions, for example, during open loop movements or movements carried out with one eye occluded (Watt & Bradshaw, 2000).

5.4.3 Interpretation in alternative theoretical frameworks

The current study approached the investigation for how additional objects in the scene affect prehension from the standpoint that these objects could provide valuable information about distance and shape for the scaling of prehension. However, the data provided little evidence for an improvement in performance with the additional information provided by a flanking object. As was mentioned in the introduction, there are a number of alternative theoretical standpoints from which to interpret the influence of additional objects in the scene. Some authors have made a theoretical distinction between processing for 'perception' and 'action' (Milner & Goodale, 1995). This theory states that whilst relations between objects play a critical role in scene interpretation for 'perception', 'action' systems can restrict computations to the target object alone. The perception action distinction would therefore predict no effect of the presence of an additional object in the scene on prehension (Milner & Goodale, 1995).

The current data are inconsistent with this theoretical distinction; the presence of another object in the scene had clear effects on the prehensile movement across observers. This indicates that the 'action' system does not restrict computations to only the target object when programming and executing reach to grasp movements; information from other objects in the scene is also taken into account. This is relatively unsurprising as we rarely interact with single objects in an otherwise empty visual scene, objects are typically interacted with in cluttered visual environments containing many additional objects and surfaces. It would be counterintuitive for computations for visuo-motor control not to take these objects and surfaces into account. That additional object in the scene are taken into account for the programming and control of prehension is consistent with numerous studies in the literature (Churchill et al., 2000, Mon-Williams et al., 2001, Tipper et al., 1997, Watt & Bradshaw, 2002). This evidence has necessitated a more recent revision of the perception-action distinction to accept that additional objects in the scene are taken into account by the action system (Milner & Goodale, 2006).

Given that there is ample evidence suggesting that the visual system takes into account the presence of other objects in the scene during prehension, the question becomes how and why these objects affect prehension. The distracter interference framework has focused on the potential of additional objects in the scene to activate competing motor programs that need to be inhibited in order for the correct motor program to the target object to be selected (Tipper et al., 1997). Consider the case where an observer is reaching to grasp a target object in the presence a single flanking object. In the distracter interference framework it is proposed that the target object and the flanking object activate competing motor programs substantiated in partially overlapping neuronal population activity, and that the activity related to the distracter object is inhibited during action selection. This activity overlap and inhibition can predict distracter interference effects such as deviations of reach trajectories (Tipper et al., 1997).

It might be predicted that if interference were occurring between competing motor programs toward a target and flanking object, the reach might veer toward the flanking object and that compared to reaches to the target object alone, peak grip aperture would be larger in the presence of a flanker larger than the target and smaller in the presence of a flanker smaller than the target. However, typically more generalised effects of distracter objects are found that are unrelated to the distracter object's position or size/shape. For example, Kritikos et al. (2000) investigated prehensile movements toward a target cylinder in the presence or absence of a flanking cylinder placed either to the left or right. This flanking cylinder could be either larger or smaller in diameter than the target cylinder. Very restricted effects of the flanker were observed and these were restricted to the reach component alone. Increased wrist velocities were found in the presence of a differently sized flanker, whether smaller or larger, and peak velocity and peak vertical deviation were reached earlier in the movement when a flanker was present (Kritikos et al., 2000).

A number of other studies that have sought to find distracter interference have found very restricted effects. For example, Tipper et al. (1997) found effects of additional flanking objects on reach trajectories but only when the target object was determined by cueing on a trial by trial basis. Similarly, Castiello (1996) found interference effects but only when observers were engaged in a demanding secondary task involving the flanking object. The current experiment used a simple prehension task in which the object to be grasped was always presented on the right hand side of the secondary object. From previous evidence no distracter interference effects would be predicted from this experimental paradigm. This was confirmed by the current data, which showed no effect of the specific dimensions of the flanking 'distracter' object on indices of prehension when reaching to grasp the target object. It is also important to note that most experimental outcomes can be predicted by altering the way in which the distracter influence is modelled making it difficult to falsify the model (Tipper et al., 1997).

Tresilian (1998) has argued that the obstacle avoidance hypothesis provides a more parsimonious explanation for the observed effects of additional objects in the scene on prehensile movements. In this framework, if objects are perceived to be obstacles to the reach and grasp movement, even if they would not physically be obstacles to the movement production, the visuomotor system will program movements so as to avoid these objects, and slow the movement so as to maintain accuracy. A number of studies have provided evidence in favour of the obstacle avoidance hypothesis. Jackson et al. (1995) had observers reach and grasp a target object placed either 15cm to the left or right of their midline in fully lit conditions. Prehensile movements to the target object were carried out with the right hand, so reaches to the left were labelled contralateral and reaches to the right ipsilateral. Reaches were performed with either the target object alone or in the presence of an additional flanking object, this was placed either nearer to the midline than the target object or more peripheral than the target object. Observers reached either with full visual feedback or closed their eyes after object placement before reaching to eliminate all visual feedback.

With full visual feedback the presence of a flanking object resulted in decreased peak grip apertures, this effect did not interact with either the type for reach (contralateral/ipsilateral) or flanker position (midline/peripheral). No effect of the flanking object was seen for peak wrist velocity. The removal of visual feedback resulted in overall decreased grip apertures and decreased wrist velocities. In addition the presence of a flanking object again decreased peak grip apertures, with a non-significant trend for this effect to be greater for contralateral reaches with a midline flanker and ipsilateral reaches with a peripheral flanker. Wrist velocities were now also effected, peak wrist velocity decreased in the presence of a flanking object, and this effect was significantly greater for contralateral reaches with a midline flanker and ipsilateral reaches with a peripheral flanker (Jackson et al., 1995).

This pattern of results suggests that observers slowed down their movement and decreased grip apertures so as to maintain accuracy and avoid collision with the flanking object, especially so when visual information is not available. Tresilian (1998) suggests that the differential effect of flankers for contralateral and ipsilateral reaches could be due to grip aperture being modulated primarily by movement of the fingers rather than movement of the thumb (Wing & Fraser, 1983). Using a similar experimental setup Jackson et al. (1997) investigated the role of binocular vision in the control of prehension. The effects of removing binocular vision were predominantly found when a flanking object accompanied the target object, the removal of binocular vision resulted in significantly longer deceleration times and significantly larger peak grip apertures.

Flanker position interacted with the type of reach, in a similar way to that in Jackson et al. (1995). Decreased wrist velocities and grip apertures were found for contralateral reaches with a midline, but not peripheral, flanker. The reverse was true for ipsilateral reaches, decreased wrist velocities and grip apertures were found with a peripheral, but not midline, flanker. A strategy of decreased grip apertures and wrist velocities was also found by Mon-Williams et al. (2001) when investigating prehensile movements in multiple object scenes. This strategy was more pronounced when there were two objects in the scene compared to one. Overall this suggests that when flanking objects are perceived to be obstacles to the reach and grasp movement kinematic parameters are modulated so as to increase accuracy and avoid the risk off collision with the flanking objects.

However, not all experimental studies investigating the role of flanking objects in the scene are consistent with an obstacle explanation. Watt and Bradshaw (2002) investigated the presence of one to four flanking objects on prehension toward a target object in both lit or unlit conditions (in the unlit condition objects were self illuminated by the presence of blobs of luminous paint on their surfaces). Grip apertures were not significantly affected by the presence of flanking objects in the scene, in either the lit conditions or unlit conditions. In contrast effects of flanking objects were observed for the transport component of the reach. In unlit viewing conditions movements in the presence of four flankers reached higher peak velocities than those to single objects. Whereas, in lit viewing conditions peak wrist velocities were significantly slower in the presence of four flanking objects.

Watt and Bradshaw (2002) raised the possibility that the differential effect of the presence of flanking objects in lit and unlit conditions could be due to the flanking objects providing improved visual information about the scene in unlit conditions, such as information about the orientation of the support surface. This effect of the flanking objects would be masked in the lit conditions where there are many redundant sources of information. There were also no interactions with the viewing condition (monocular/binocular), which would be expected if binocular vision were particularly important movement control in multiple object scenes (Jackson et al., 1997).

In the context of the current experimental design the object avoidance hypothesis would predict that if the flanking object were interpreted as an obstacle decreased wrist velocities and grip apertures would be observed compared to prehensile movements made toward the target object alone (Jackson et al., 1995, Tresilian, 1998). This would act to decrease the possibility of the hand colliding with either object. Furthermore, if modulation of grip aperture is primarily achieved primarily by movement of the fingers (Wing & Fraser, 1983) these effects might only be observed if the flanking object were on the side nearest to where the fingers would contact the flanking object.

In the present experiment the flanking object was always presented on the left and observers reached and picked up the target object on the right with their right hand.

This meant that the flanking object was always on the side of the target object that the thumb came into contact with. It is therefore possible that the presence of the flanking object would have little or no effect on grasp parameters, as it would place no constraints on movement of the fingers (this is true when grasping either the depth or width of an object).

Overall the current data are too inconsistent to make firm conclusions regarding the object avoidance hypothesis. There were large differences in the affect a flanking object had on observers reach to grasp movements, and differences in the affect shown by individual observers when grasping across either the depth or width of the flanking object. The presence of a flanking object did however have effects on both wrist velocity and grip aperture across most observers, suggesting that under these experimental conditions no disassociation between the effect of the flanking object on the reach and grasp components of the prehensile movement were evident, as has been found previously (Watt & Bradshaw, 2002).

5.4.4 Conclusion: Interpreting kinematic indices

The experiment reported in this chapter was designed to investigate whether information from an additional object in the scene can be utilised by the visual systems for scaling disparity in a target object for the programming of prehension. The experiment was approached from the theoretical standpoint that additional objects in the scene could be beneficial for prehension by providing valuable information about distance and shape (Brenner & Landy, 1999). The results provided little evidence for a disparity scaling explanation of changes in kinematic indices across viewing conditions. When all the data were considered as a whole it was more consistent with observers slowing down their reach and producing larger grip apertures in the presence of another object in the scene. However, it must be emphasised that this pattern of results was not consistent across all observers, or consistent for individual observers grasping across an objects depth and width.

Slower reaches with wider grip apertures are indicative of a conservative strategy whereby an increased margin for error is built into the prehensile movement

(Melmoth & Grant, 2006, Watt & Bradshaw, 2000). If this pattern of results can characterise the data, it is surprising that it was found with the presence of a flanking object, because all flanking object conditions, to a greater or lesser degree, provided additional information about distance and shape compared to that available from the target object alone. In the previous section the data were considered in terms of alternative theoretical frameworks, specifically the perception-action distinction (Milner & Goodale, 1995), distracter interference (Tipper et al., 1997) and obstacle avoidance (Tresilian, 1998). The data were found to be inconsistent with the notion that the visuo-motor system can restrict processing to the target object alone, as proposed by the perception-action distinction (Milner & Goodale, 1995). The data also provided no evidence for distracter interference effects caused by competitive interference between alternative motor programs (Tipper et al., 1997).

Obstacle avoidance would therefore seem a more plausible candidate to account of the changes in kinematic indices found in the present study. The majority of studies which have investigated the effect of one or more flanking objects on prehensile movements have found that the presence of flanking objects results in the velocity of the reach and grip aperture being decreased (Mon-Williams et al., 2001). This is thought to be a reflection of the increased accuracy required to reach and grasp objects in multiple object scenes. Unfortunately the results observed in the present study were too variable to make firm conclusions regarding the object avoidance hypothesis. It is also true that decreases in grip aperture and wrist velocity are not always found when prehension is carried out in multiple object scenes (Watt & Bradshaw, 2002).

It therefore remains unclear what processes underlie the effects observed in the present experiment. It is clear that there was very little consistency between observers, which could suggest that multiple processes were at work. For example, different observers may weight the information available from flanking objects differently during prehension, or perceive the flanking objects as obstacles to differing degrees. Both of these factors would modulate the extent to which kinematic indices were affected. Given that multiple processes can influence kinematic indices of prehension, this questions the extent to which these indices can be simply interpreted in terms of environmental properties, such as distance and size, in more complex viewing situations. It also raises questions about the extent to which kinematic indices of distance and size can be compared to the results of perceptual studies.

Chapter Six

Discussion and summary

6 Discussion and summary

6.1 The utilisation of horizontal disparity by human observers

This thesis sought to examine the use of horizontal disparity in the estimation of three-dimensional shape for perception and visuomotor control. Horizontal disparity is a potentially useful source of visual information as once scaled by an estimate of the viewing distance it can geometrically specify the 3-D structure of the environment (Howard & Rogers, 2002). Despite this possibility, previous research has demonstrated systematic distortions in the perception of 3-D shape from disparity, such that the same disparity defined object has a different perceived shape when presented at different viewing distances. Objects appear stretched in depth extent at near distances and squashed in depth extent at far distances, this is consistent with observers scaling disparity with an overestimate of near distances and an underestimate of far distances (Bradshaw et al., 2000, Glennerster et al., 1996, Glennerster et al., 1998, Johnston, 1991).

Chapter one detailed an experiment that investigated the shape constancy of objects moving in depth. A simple prediction from the lack of shape constancy demonstrated in previous studies is that a disparity-defined object moving toward the observer should be seen to stretch in depth extent and a disparity-defined object moving away from the observer should be seen to squash in depth extent. Conversely, a perceptually constant cylinder should be one that squashes in depth extent when moving toward the observer and stretches in depth extent when retreating, thus countering the effects of incorrect distance scaling. Whilst this a direct prediction from previous demonstrations of the lack of shape constancy, observers are provided with additional information about 3-D shape when objects move, which could allow them to accurately judge an objects shape constancy as it moves in depth.

The conjunction of stereo and motion information theoretically allows the recovery of metric 3-D shape information (Landy & Brenner, 2001, Richards, 1985). Similarly, an alternative processing strategy was identified which could allow observers to accurately judge the rigidity of objects moving in depth, without the need to combine stereo and motion information (Landy & Brenner, 2001, Richards, 1985). In both instances an observer could be said to have a representation of metric 3-D shape gained from the combination of stereo and motion information, as they would be able to complete the task accurately, but with the alternative processing strategy this representation is task dependent (Bradshaw et al., 2000, Glennerster et al., 1996, Todd, 2004, Todd & Norman, 2003) and not generally applicable to other shape judgement tasks (Marr, 1982).

The results demonstrated that observers were unable to use the conjunction of stereo and motion information to identify whether objects remained constant in shape when moving in depth. To be perceived as constant in shape objects needed to squash in depth extent when moving towards the observer and expand in depth extent when moving away from the observer, countering the effects of incorrect distance scaling (Johnston, 1991). This highlights the striking consequences that can result from the failure of observers to obtain shape constancy by either combining stereo and motion information to recover a metric representation of shape (Richards, 1985) or by using some other processing strategy. If the visual system is unable to recover geometrically correct shape information a perceptually stable world is not necessarily a physically constant one.

This is consistent with recent research which has shown that in a virtual reality environment observers are unable to determine whether an object remains stationary whilst they move relative to it (Tcheang et al., 2005), or that a room is expanding four fold when they move across it (Glennerster et al., 2006b). Although these demonstrations of observers' inability to determine the stability of the world are found under very specific experimental circumstances (Glennerster, McKean & Gilson, 2006a, Tcheang et al., 2005), large failures of shape constancy have been demonstrated in full cue natural environments (Cuijpers et al., 2000, Wagner, 1985). It is possible that observers simply do not require metric information about distance and shape in most everyday circumstances, so therefore remain unaware of failures of shape constancy. It may take an experimental situation that forces an observer to make decisions about metric structure to bring these biases to light.

The aetiology of the distortions in the perception of shape from disparity and distance from vergence were investigated in Chapter three. Current research on cue combination has tended to model cue estimates of a world property as arising from unbiased but noisy perceptual estimation, cue estimates are modelled as Gaussian likelihood distributions centred on the true value of the world property (Hillis et al., 2002, Hillis et al., 2004). However, it is possible that noise in early visual signals could act to introduce bias into sensory estimates through the way in which a sensory signal is mapped onto a perceptual estimate. This would mean that perceptual estimates would no longer be well represented by Gaussian likelihood distributions. The likelihood distributions could be skewed and this could act to introduce bias into perceptual estimates (Fermuller & Malm, 2004, Hogervorst & Eagle, 1998).

In Chapter three Monte Carlo simulations were used to investigate the role of vergence, disparity and gaze angle noise in the perception of distance, depth, height and shape (depth to height ratio). The simulations showed that vergence noise predicted the progressive underestimation of distance, depth and height as the viewing distance increased. Furthermore, due to the different way in which height and depth scale with distance this resulted in depth to width ratios decreasing with increasing distance. This is consistent with the experimental literature in this area. Near distances are estimated well from vergence but far distances are progressively underestimated (Mon-Williams & Tresilian, 1999, Tresilian et al., 1999, Viguier et al., 2001) and in shape judgment tasks depth intervals are progressively underestimated relative to width/height intervals as the viewing distance increases (Glennerster et al., 1996, Glennerster et al., 1998, Todd et al., 1995).

The model does not however predict the overestimation of depth relative to width/height at near distances (Johnston, 1991, Todd et al., 1995). This together with the distance estimation literature, points to the fact that biases in the perception of shape at near distances are not caused by the misestimation of object distance. In contrast to the role of vergence noise, disparity and gaze angle noise predicted no bias in perceived 3-D shape. Overall the results of this chapter are important as they go some way to explaining why biases in the perception of shape are found in the experimental situation, without the need to propose cognitive or perceptual biases

(Mon-Williams & Tresilian, 1999, Poulton, 1989) or introduce learnt environmental distance priors (Yang & Purves, 2003).

The standard specific distance tendency explanation does not provide a satisfactory explanation of the experimental data relating to distance and shape estimation (Gogel & Tietz, 1973). Similarly, although prior distance information is an elegant way in which to explain distance misestimation, the inferred priors cannot currently provide a unitary explanation of the misestimation of 3-D shape in most experimental situations (Yang & Purves, 2003). This is not to say that prior distance and disparity information do not have any role to play in the misestimation of shape from disparity (Glennerster et al., 1998, Yang & Purves, 2003). It is difficult to assess whether the level of noise in the vergence signal needed to explain observers' perceptual bias is realistic, this is because it is difficult to empirically determine the level of noise present in the brain's estimate of the vergence state of the eyes. However, variation in the level of signal noise could go some way to explaining the large individual differences that are found in most 3-D shape judgement tasks (e.g. Champion et al., 2004, e.g. Oruc et al., 2003, Todd & Norman, 2003).

Whilst bias in the perception of distance and shape from binocular information allows the processes underlying perceptual estimation to be investigated, these biases are normally demonstrated under reduced cue conditions, such as viewing disparity-defined objects presented at eye height. This is not representative of the viewing situations in which observers could utilise binocular information for the estimation of distance and 3-D shape in natural everyday tasks. Some part of the bias in the perception of shape from disparity could therefore be due to the unnatural viewing situation that observers are subject to in the experimental situation. For example, we typically view objects from above when reaching to grasp them. With a more natural viewing situation there are a greater number of visual cues available to the observer, and these cues may provide more reliable information about distance and 3-D shape.

Chapter four investigated the role of height in the field (Gardner & Mon-Williams, 2001) and a view of an object's upper surface and contour in the perception of 3-D shape. Compared to a standard eye height shape judgement task, improved performance was found when observers viewed an object from above and had information available from the rendering of an objects upper surface and contour (although not for all observers). In contrast, no evidence was found for improved performance when height in the field was the only additional cue to shape. The lack of evidence for the use of height in the field information could have been due to the experimental viewing situation, in which a computer rendered object needed to be interpreted as resting on an inferred table surface. This highlights the lost ecological validity that typically has to be accepted when computer generated stimuli are used. Overall the data suggest that more naturalistic viewing situations can dramatically reduce biases in perceived 3-D shape.

Geometrically there is enough information in experimental stimuli in standard eye height shape judgement tasks to define metric shape, but this does not mean that the visual system is actually able to exploit this information. The use of vertical disparity information is a case to point (Bradshaw et al., 1996, Collett et al., 1991, O'Kane & Hibbard, 2002, Rogers & Bradshaw, 1993). This suggests that the extent of bias demonstrated in previous shape judgement tasks may in part be due to artificially reducing the amount and quality of cue information available to the observer. The results for Chapter 2 indicated that when vergence is the only cue to distance, noise in its measurement could result in perceptual bias in distance and shape estimates. These results, together with those of Chapter 4, suggest that when additional more reliable sources of information are available biases in perceived shape may be greatly ameliorated.

The final experimental chapter investigated the use of disparity information in the control of prehension. Prehension is a task in which binocular disparity has been seen to be of primary importance for the specification of 3-D shape. Previous studies have tended to examine performance by contrasting prehension when carried out under binocular and monocular viewing (Servos & Goodale, 1994, Servos et al., 1992). A problem with this type of methodology is that it removes all information from one eye; this could in and of itself disrupt prehensile movements beyond any effects that removal of binocular information will have (Bradshaw et al., 2004). Furthermore, the contrast between binocular and monocular viewing does not allow the scaling of disparity to be investigated in prehension, which has been the focus of perceptual studies relating to the use of disparity information (e.g. Glennerster et al., 1996).

The experiment investigated whether information available in a flanking object in the scene (binocular/monocular, static/rotating) was used to better scale the disparity and angular size of the object to be grasped, which was always binocularly viewed and static (Brenner & Landy, 1999). This meant that the information from the object to be grasped remained constant across all viewing conditions and allowed concurrent measures of distance and shape to be gained for the investigation of disparity scaling. The results were highly inconsistent across observers and inconsistent for a given observer when grasping across the depth or width of the target object. Overall there was little evidence to suggest that the visual system used information in the flanking object to better estimate the distance and shape of the target object.

More global effects relating to the presence or absence of a flanking object were found, rather than effects relating to the information available from this flanker. Previous research has tended to interpret the presence of other objects in the scene as obstacle to be avoided (Mon-Williams et al., 2001, Tresilian, 1998); it is possible that some observers interpreted the flanking object as an obstacle in the present study. However, observers did not show the pattern of kinematic indices typical of this interpretation i.e. decreased grip apertures and slower wrist velocities. If multiple processes affect kinematic indices such as distance scaling (Servos et al., 1992), obstacle avoidance effects (Tresilian, 1998) and the use of perceptual heuristics (Todd, 2004), this may make it difficult to interpret kinematic indices unambiguously in terms of perceived distance and shape.

6.2 Summary: do we need metric information about 3-D structure?

The thesis as a whole has demonstrated that the visual system's utilisation of horizontal disparity is limited by the extent to which the sensory apparatus can estimate properties of the world from binocular information. When estimating 3-D shape from binocular cues, such as vergence and horizontal disparity, biases in perceived 3-D shape could be the consequence of the visual system estimating the most likely state of the world to have produced the noisy visual signals. Theoretically it is possible for the visual system to overcome these perceptual biases by combining disparity information with other cues such as motion to more accurately estimate 3-D shape. However, when the visual system is unable to exploit this additional information a perceptually stable world is not necessarily a physically constant one. Disparity-defined objects moving toward the observer at eye height are perceived to stretch in depth extent and objects moving away contract in depth extent.

By contrast, when a disparity-defined object is viewed from above, such that additional information about distance and shape is provided by height in the field and a view of an objects upper surface and contour, shape constancy can approach 100%. This suggests that in a more naturalistic viewing situation biases in perceived shape can be much reduced. This highlights the fact that although human observers can be presented with experimental stimuli that can geometrically specify metric 3-D shape, this does not necessarily mean that the visual system will be able to utilise this information for the estimation of metric 3-D shape. The visual system has evolved to provide the organism with sufficient information for the adaptive control of behaviour. If sufficient information can be provided for the adaptive control of behaviour by other means, the visual system may simply not need to be able to estimate metric structure from binocular visual cues.

Humans interact in a highly skilled and adept way with their environment, despite striking failures in our ability to estimate the perceptual constancy and stability of the environment in experimental situations. During our natural interactions with objects in the world we can move so as to get the best information available for the estimation of object properties and the control of behaviour. This may allow us to overcome any deficiencies in the way in which our sensory apparatus has evolved to provide us with information about the world. It is also possible that in most natural situations we might rarely need metric information for the control of behaviour. This could explain the failure of studies, such as those of prehension, to find unambiguous evidence for the need of metric shape information. It seems unlikely that the “goal” of the visual system is to obtain metric real-time information about the world. The visual system simply needs enough information to get by.

Appendices

7 Appendices

7.1 Derivation of the disparity equation.

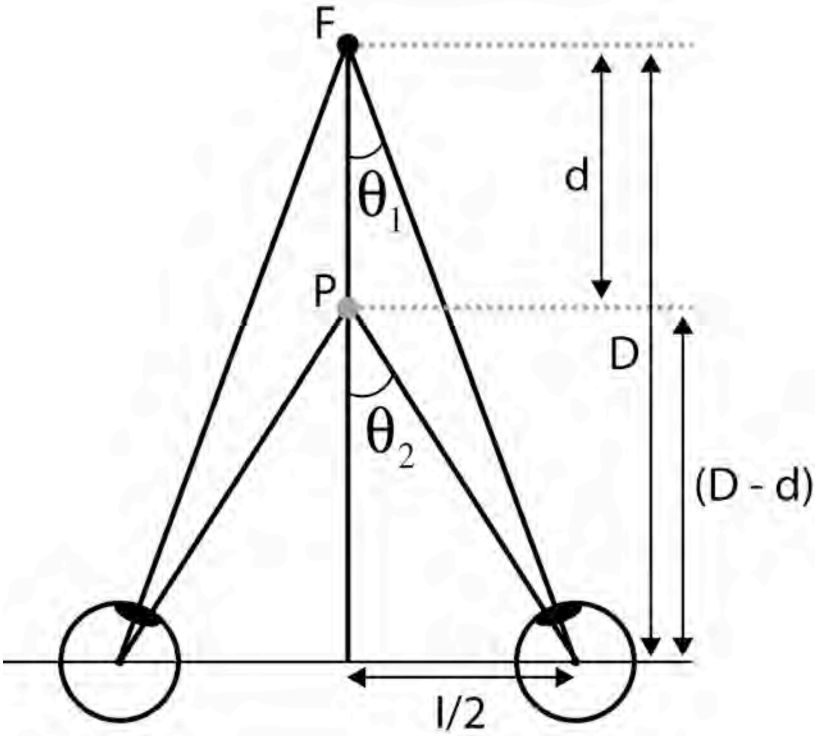

Figure A1.1: An observer with a half interocular distance of $I/2$ fixates a point F at a distance D along the median plane of the head. Point P is a distance d nearer to the observer along the median plane of the head. Point P has a disparity η with respect to F. This disparity is defined as $\eta = 2\times(\theta_2 - \theta_1)$.

From simple geometry,

$$\tan\theta_1 = \frac{I/2}{D}$$

(A1.1)

$$\tan\theta_2 = \frac{I/2}{(D-d)}$$

(A1.2)

Given the trigonometric identity, $\tan(\theta_2 - \theta_1) = \dfrac{\tan\theta_2 - \tan\theta_1}{1 + \tan\theta_2 \cdot \tan\theta_1}$, and substituting from equations (A1.1) and (A1.2),

$$\tan(\theta_2 - \theta_1) = \frac{\dfrac{I/2}{D-d} - \dfrac{I/2}{D}}{1 + \left(\dfrac{I/2}{D-d}\right) \cdot \left(\dfrac{I/2}{D}\right)}$$

(A1.3)

Simplifying gives,

$$\tan(\eta/2) = \frac{(I/2) \times d}{D(D-d) + (I/2)^2}$$

(A1.4)

The half-disparity, $\eta/2$, is therefore given by,

$$\eta/2 = \arctan\left(\frac{(I/2) \times d}{D(D-d) + (I/2)^2}\right)$$

(A1.5)

Given the small angle approximation, $\tan\theta \approx \theta$, and assuming that $D >> d$ and $D >> I$ equation (A1.4) can be simplified to,

$$\eta/2 \approx \frac{(I/2) \times d}{D^2}$$

(A1.6)

Therefore,

$$\eta \approx \frac{Id}{D^2}$$

(A1.7)

7.2 Stereo-motion heuristic

7.2.1 Derivation of the heuristic[9]

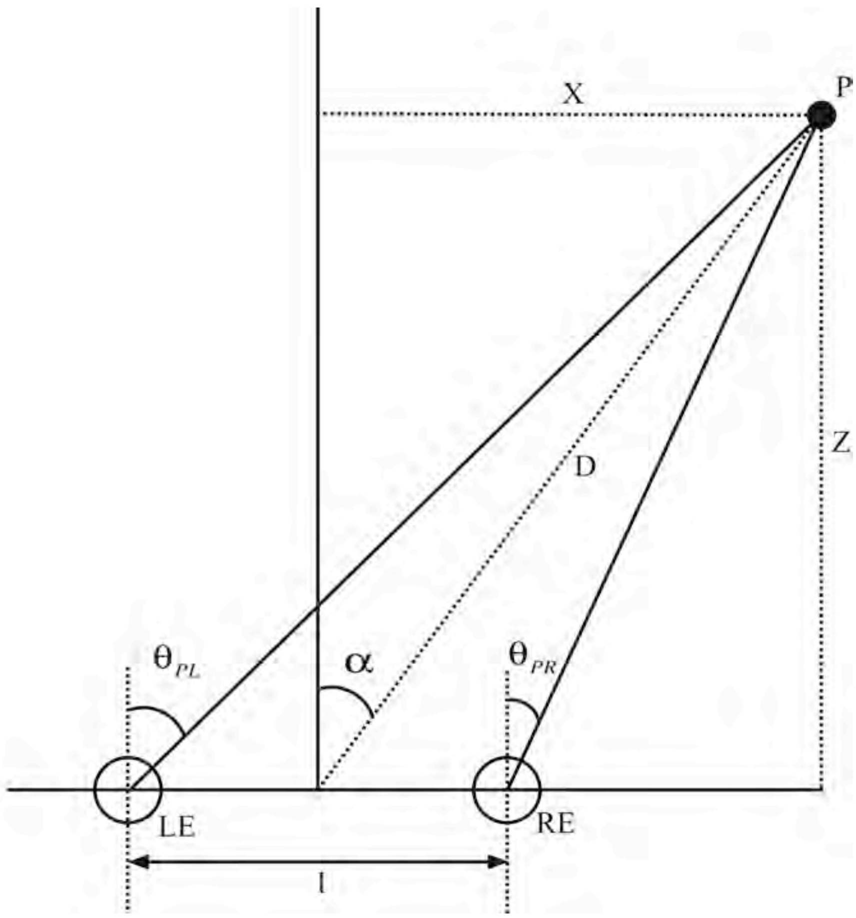

Figure A2.1: showing the geometry of a point P at a depth Z, and horizontal position X from the cyclopean eye. The visual direction of this point in the left eye (LE), right eye (RE) and cyclopean eye are given by θ_{PL}, θ_{PR} and α respectively. The inter-ocular distance is denoted by I, and the distance of P from the cyclopean eye by D.

Consider a point P at eye height at a depth Z and horizontal position X (Figure A2.1).

The visual directions of this point in the left eye and right eye are given by:

[9] Derivation of the heuristic performed by Dr. P. B. Hibbard, simulation testing the applicability of the heuristic performed by P. Scarfe

$$\tan\theta_{PL} = \frac{D\sin\alpha + I/2}{D\cos\alpha}$$

(A2.1)

and

$$\tan\theta_{PR} = \frac{D\sin\alpha - I/2}{D\cos\alpha}$$

(A2.2)

respectively.

Assuming that $D^2 >> I^2$, the optic array disparity of this point, $\gamma_P = \theta_{PL} - \theta_{PR}$, is given by:

$$\tan\gamma_P \approx \frac{-I\cos\alpha}{D}$$

(A2.3)

Now consider a second point Q, at a position $(X, Z + d)$, assuming that (i) $D >> I$ and (ii) $D >> d$, its optic array disparity is given by:

$$\tan\gamma_Q \approx \frac{-ID\cos\alpha + Id}{D^2}$$

(A2.4)

With fixation on point P, the disparity of point Q, γ, is given by:

$$\tan\gamma = \frac{-ID^2\cos\alpha + IDd + ID^2\cos\alpha}{D^3 + DI^2\cos^2\alpha - I^2 d\cos\alpha}$$

(A2.5)

Again, assuming (i) $D >> I$ and (ii) $D >> d$:

$$\tan\gamma \approx \frac{Id}{D^2}$$

(A2.6)

Now consider a point W, at position $(X + w, Z)$

Then

$$\tan\theta_{WL} = \frac{D\sin\alpha + I/2 + w}{D\cos\alpha}$$

(A2.7)

If $\phi = \theta_{WL} - \theta_{PL}$, and assuming $w << D$, then

$$\tan\phi \approx \frac{w\cos\alpha}{D}$$

(A2.8)

Let

$$S = \tan\gamma = \frac{Id}{D^2}$$

(A2.9)

If we assume that the current distance and direction of the point from the observer are given by D and α, that the (constant) radial and angular velocity are given by $\dot{D}$ and $\dot{\alpha}$, and that the location of the point at time t = 0 was D_0 and α_0, then

$$S = \frac{Id}{\left(D_0 + \dot{D}t\right)^2}$$

(A2.10)

and

$$\frac{\dot{S}}{S} = -\frac{2\dot{D}}{D}$$

(A2.11)

Let

$$T = \tan\phi = \frac{w\cos\alpha}{D} = \frac{w\cos(\alpha_0 + \dot{\alpha}t)}{D_0 + \dot{D}t}$$

(A2.12)

Then

$$\frac{\dot{T}}{T} = -\dot{\alpha}\tan\alpha - \frac{\dot{D}}{D}$$

(A2.13)

From this it is clear that if either $\dot{\alpha}$ or $\alpha = 0$:

$$\left(\frac{\dot{S}/S}{\dot{T}/T}\right) = 2$$

(A2.14)

In other words the ratio of the change in disparity divided by the disparity itself, to the change in retinal size divided by the retinal size itself, equals 2 for an object moving rigidly in depth at a constant speed along the cyclopean line of sight.

7.2.2 Testing the applicability of the heuristic

In order to test the stability of equation A2.14 and its applicability in the current experimental situation, simulations were carried out. Objects moving along the cyclopean line of sight where α and $\dot{\alpha}$ = 0 represent a restricted situation. For equation A2.14 to be a useful strategy the ratio should be equal to 2 for a range of α and $\dot{\alpha}$. Simulations were carried out using the parameters of the experimental setup to assess A2.14 for a range of depth path angles (θ in Figure A2.1). Over these different path angles both α and $\dot{\alpha}$ vary. The simulations showed that equation A2.14 gives a good approximation of a rigidly moving object even without the additional information of α and $\dot{\alpha}$. Over a range of depth path angles (0 to 40 degrees), error in A2.14 reached a maximum of 10.1% at 40 degrees, and decreased rapidly with decreasing path angle. For the path angle of the experiment, error in A2.14 was 5.28%. This suggests that equation A2.14 represents a useful strategy for observers to adopt. It should be noted that, even if observers did not adopt A2.14, α and $\dot{\alpha}$ should be readily available from the optic array, thus allowing the strategy to be applied more generally.

7.3 Monitor spatial calibration.

7.3.1 Why is monitor spatial calibration required?

When investigating the perception of 3-D shape from visual cues it is imperative that the stimuli are presented geometrically correctly. An experiment seeks to investigate the processing that occurs naturally when we perceive objects and scenes in the world, this will not be possible if the stimuli are somehow transformed by the way in which they are presented. When using real world stimuli this problem is avoided, as the stimuli need simply to be manufactured correctly and presented. However, the situation becomes more complex when using computer-generated stimuli presented on a monitor screen or by projector. This is typically the case in most psychophysical experiments, as precise control needs to be gained over the information available to the observer. However, when using a monitor there is an unknown mapping between the units in which the stimuli need to be presented e.g. centimetres or metres, and the pixel coordinates on the screen. Spatial calibration seeks to ascertain and control for this mapping, so that stimuli are presented geometrically correctly as intended.

7.3.2 Spatial calibration: method one

A transparent Perspex sheet with a $1cm^2$ grid machine scored into its surface was placed centrally in front of the monitor abutted to its casing. A $1cm^2$ grid of Gaussian blobs was displayed on the monitor screen. The pixel to centimetre conversion factor used to program the grid was calculated by counting the number of pixels in a known distance across the surface of the screen. The grid of Gaussian blobs was displayed on the monitor screen and the monitor geometry controls were used to adjust the screen geometry until the screen grid was is as close an alignment as possible to the Perspex grid. This calibration technique was used for the experiments reported in Chapters 2 and 3.

7.3.3 *Spatial calibration: method two*

For the experiment reported in Chapter 4, geometrically correct presentation of stimuli was achieved using a screen spatial calibration technique similar to that described by Backus, Banks, van Ee and Crowell (1999). The equipment and custom software needed to carry out this calibration was produced as part of these experiments[10]. The first stage of the calibration is to precisely position an observer so that their eyes are in a fixed position relative to the monitor on which the stimuli are to be presented. This was achieved by using custom built sighting device and bite bar. The bite bar has six degrees of freedom that can each be adjusted independently (Figure A3.1); these are movement along the x, y, and z-axis, pitch around the y-axis, roll around the x-axis, and rotation around the z-axis. In conjunction with the sighting device (Figure A3.2) this allows an observer's eyes to be precisely positioned in space.

[10] I am very grateful to Andrew Burnley of the University of St Andrews for taking sketches and photos of a sighting device and loom, and producing an expertly made working article. Setting up this calibration procedure simply would not have been possible without his continued help and advice. I am also grateful to Prof. Martin Banks of the University of California, Berkeley, USA, Prof. Julie Harris of the University of St Andrews, Scotland, UK and Dr. Simon Watt of the University of Bangor, Wales, UK, for help and advice about setting up the spatial calibration. In particular Dr. Simon Watt allowed the use of his bite bar design and got these manufactured, and has provided continuous feedback, help and advice throughout the process of getting the equipment produced, and during programming of the spatial calibration routine.

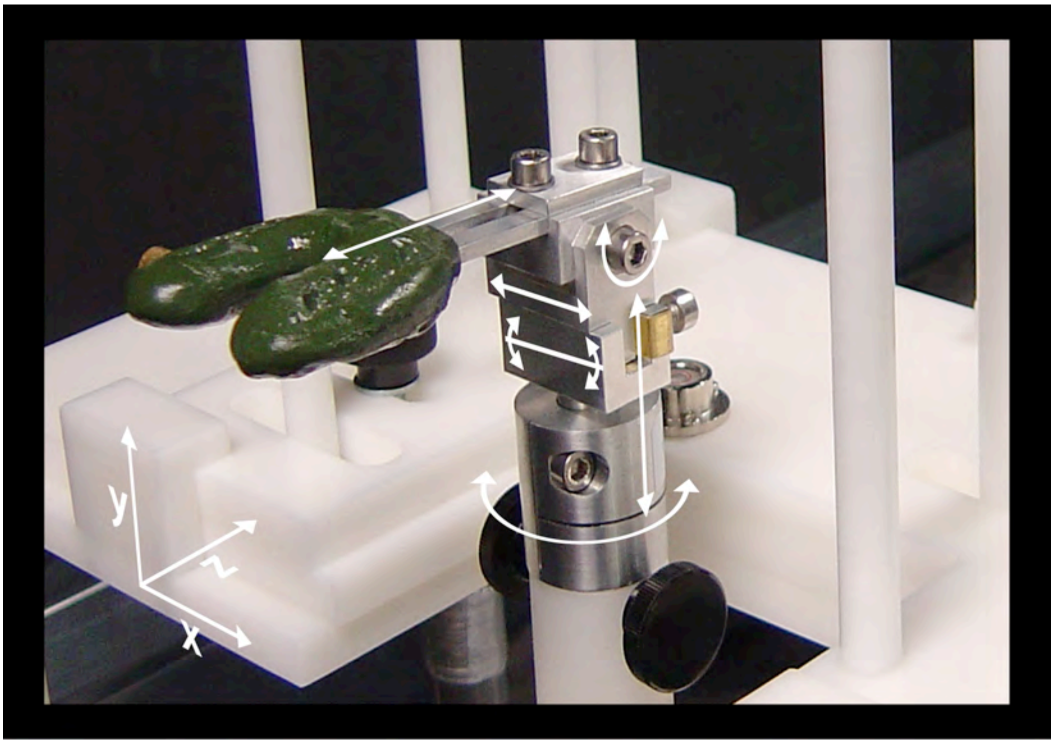

Figure A3.1: Photograph of the bite bar with arrows illustrating the directions of movement and rotation possible. The bite bar is used to precisely position the centre of rotation of an observer's eyes at a known distance and orientation relative to the monitor screen. See text for details.

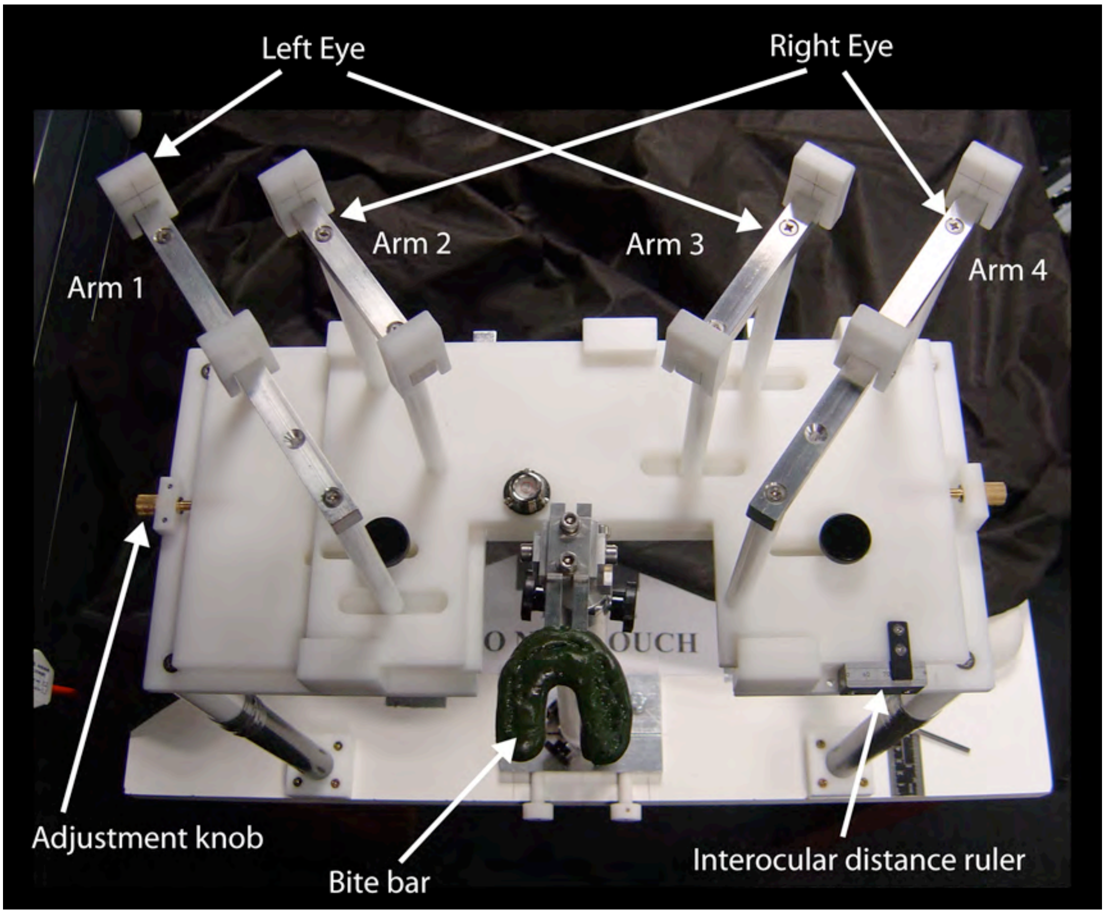

Figure A3.2: Photograph of the custom built sighting device that was used in conjunction with the bite bar to precisely position an observers eyes. The main features of the sighting device are labelled; see text for a more detailed description.

The sighting device has four sighting arms, labelled 1 to 4 in Figure A3.2; all four arms are at fixed height from the base of the sighting device, and therefore the support surface. Each arm consists of a target crosshair and a small sighting hole through which to view the crosshair (Figure A3.3). Figure A3.4 shows a schematic top down view of the sighting device. Arms 1 and 3 for the left eye are fixed relative to one another and define the triangle T_1. Similarly, arms 2 and 4 for the right eye are fixed relative to one another and define the triangle T_2. The lines of sight through arms 1 and 3, and arms 2 and 4 converge at points P_1 and P_2 respectively. The arm pairs forming triangles T_1 and T_2 can translate independently in the x-dimension. This allows the distance between P_1 and P_2, labelled I, to be adjusted.

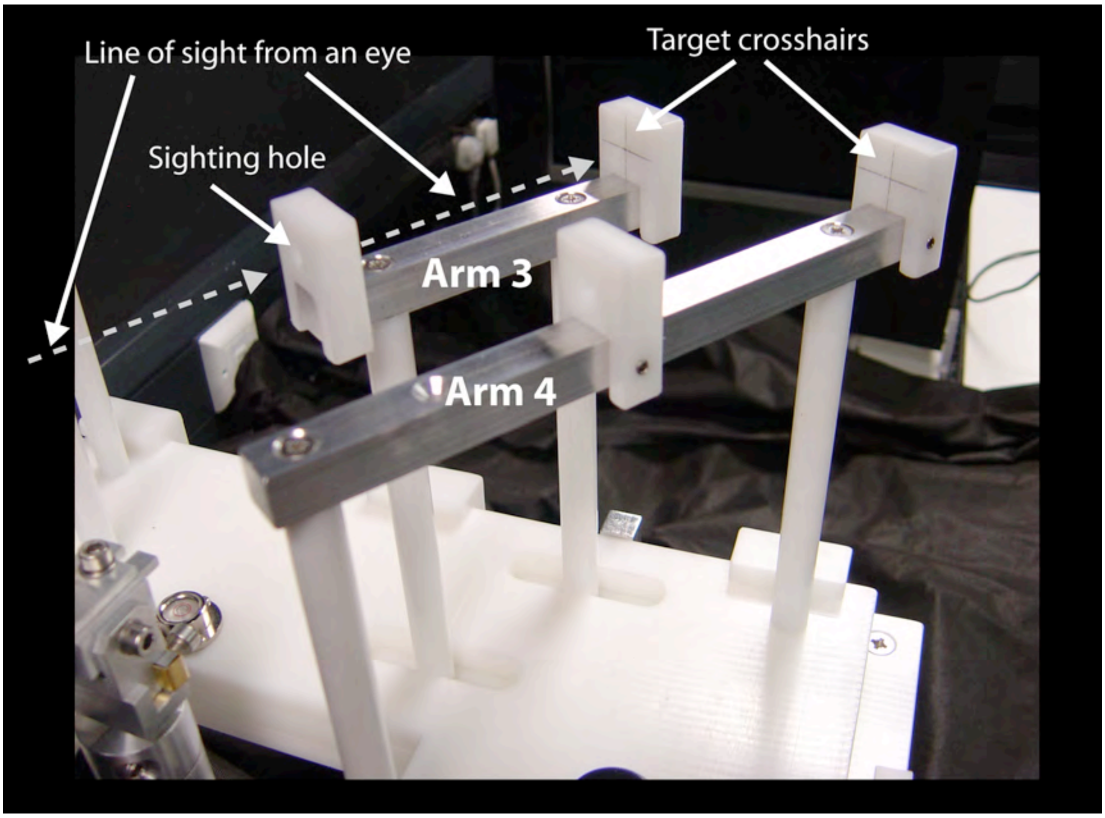

Figure A3.3: Photograph detail of two of the sighting devices sighting arms. Each arm consists of a sighting hole through which to view a target crosshair. When the eye is appropriately positioned the crosshair target is centred in the sighting hole, the grey dashed arrows indicate this line of sight. For further detail refer to the accompanying text.

Both T_1 for the left eye and T_2 for the right eye cover the same angle of azimuth, θ. When positioned appropriately, the centre of rotation for the left eye is positioned at P_1, and the centre of rotation of the right eye at P_2. This means that when the left eye is rotated in its orbit it can view crosshair of arm 1 (C1) through the sighting hole of arm 1 (S1), and the cross hair of arm 3 (C3) through the sighting hole of arm 3 (S3). Similarly, when the right eye is rotated in its orbit it can view the crosshair of arm 2 (C2) through the sighting hole of arm 2 (S2), and the crosshair of arm 4 (C4) through the sighting hole of arm 4 (S4). It also means that the distance I between points P_1 and P_2 is equal to the observer's interocular distance.

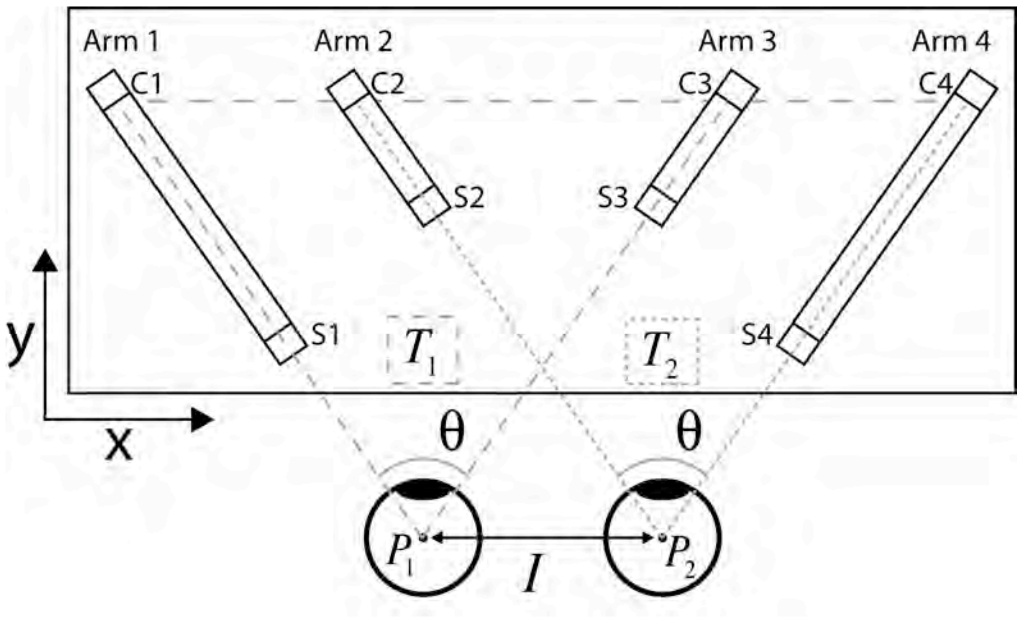

Figure A3.4: Schematic diagram showing a top down view of the sighting device. Each of the four sighting arms (Arm 1-4) has a crosshair target (labelled C1-C4) and a sighting hole (labelled S1-S4). Arm 1 and 3 are in fixed positions relative to one another and can be pictured as forming triangle T_1. Arms 2 and 4 are in fixed positions relative to one another and can be pictured as forming triangle T_2. Triangles T_1 and T_2 each cover the same angle of azimuth θ, and can be move relative to one another in the x dimension. This changes the x dimension distance between points P_1 and P_2, which is labelled I. Use of the bite bar and the sighting device allow the positioning of the centres of rotation of observers eyes at points P_1 and P_2, here I would be equal to the observer's interocular distance. For discussion refer to the accompanying text.

Measurement rulers on the sighting device indicate the distance each pair of arms has moved along the x-dimension, and the value of I. Equal and opposite movements of the left and right eye's arm pair along the x-dimension allows the sighting devices interocular distance, I, to be increased symmetrically about its mid point. In this way the sighting device and bite bar can be used to position observers of different interocular distances so that the centres of rotation of their eyes are fronto-parallel and at a known distance from the screen. The midpoint between their eyes is then also aligned with the midpoint of the monitor screen. The height of observers' eyes can also be adjusted by carrying out the alignment procedure with different length leg extensions for the sighting device, and a different length bite bar mount.

There is a slight difference between the centre of rotation of the eye, and the nodal point of the eye through which light is focused by the lens and projected to the retina, this is on the order of 5.7mm (Howard & Rogers, 2002). The eyes are also not rigid and will change shape when rotating in their sockets; this will change the relationship between the nodal point and centre of rotation of the eyes. This calibration method is unable to account for these effects, but their magnitude will be small, and within the positioning accuracy of the equipment.

Once the observer is appropriately positioned, the monitor screen is spatially calibrated relative to a 1cm square nylon filament loom grid. The nylon filaments of the loom are 0.234mm in diameter and spray painted black to be viewable against a dimly lit monitor screen. The central filament cross of the loom is marked in white. During calibration the loom is tightly fitted into housing fixed to the front of the monitor, so that the loom's surface is frontoparallel relative to the screen, and at a distance of 1cm from its surface. The monitor screen is viewed through the loom during calibration (Figure A3.5). The small distance between the surface of the loom and the monitor screen minimises the effects of conflicting focus cues, which are known to affect the perception of simulated 3D shape (Watt et al., 2005a). The plane of the loom surface serves as a 'virtual screen' of a known geometry upon which stimuli are displayed during experiments. The loom itself is only present for the calibration procedure.

7.3.4 Polynomial spatial calibration: procedure and software

The spatial calibration routine was written and presented using the Matlab programming environment with the Psychophysics toolbox extensions (Brainard, 1997, Pelli, 1997). For each observer calibration is carried out separately for each eye, and for each viewing distance and viewing height used. The goal of the calibration procedure is to provide a function that relates the distorted geometry of the monitor to the correct geometry of the loom. This gives a mapping from loom coordinates (X_L, Y_L) to pixel coordinates (X_P, Y_P). Stimuli can then be presented on the monitor screen to appear at known geometrically correct positions on the plane of the loom. In brief, the calibration procedure consists of presenting a square grid of Gaussian blobs on the monitor screen, the visual direction of each of these bobs is then adjusted, separately for each eye, so that the blobs are aligned at known intersections of the loom (Figure A3.5). Two-dimensional second-order polynomial equations are then used to give a function mapping the x and y coordinates of the loom (X_L, Y_L) to the x and y coordinates of the screen (X_P, Y_P). During stimulus generation these equations are used convert between centimetres on the loom (the coordinates the stimuli need to be presented in) and pixels on the screen (the coordinates the stimuli displayed on).

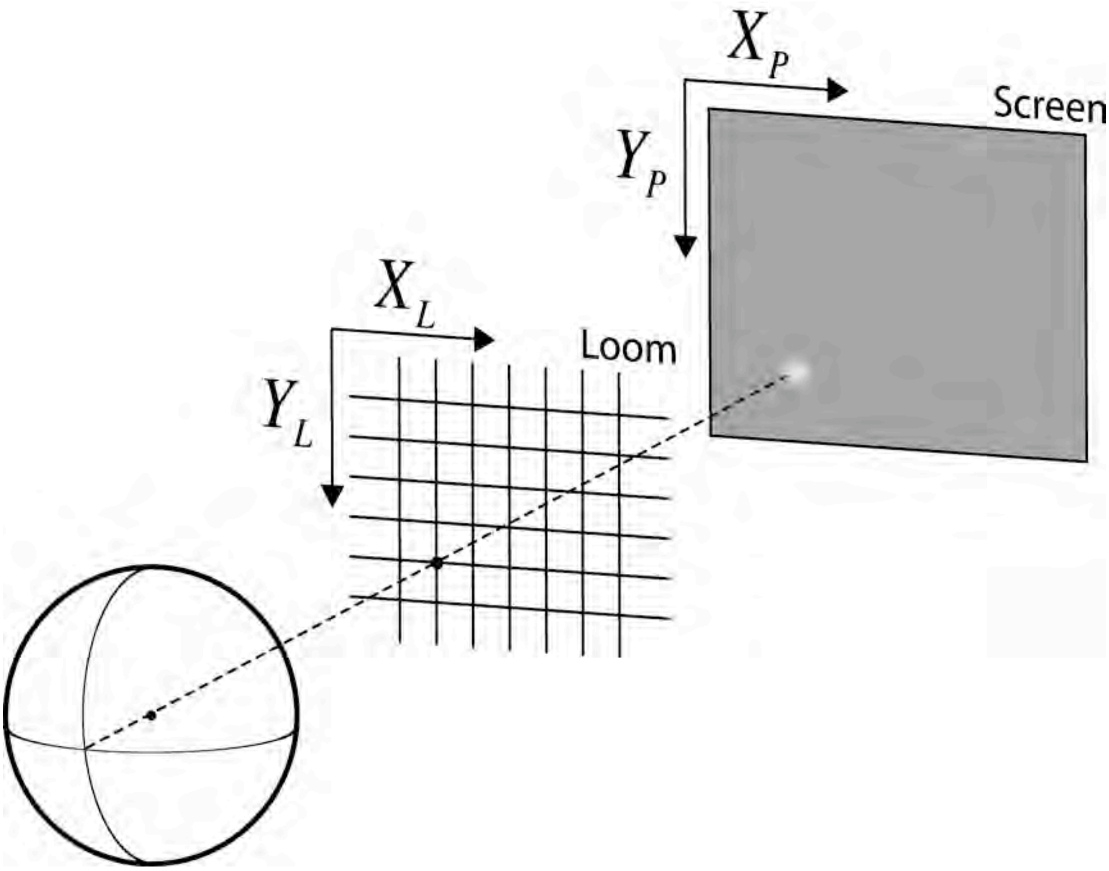

Figure A3.5: Schematic diagram of the loom calibration procedure. The observer positions Gaussian blobs that form a grid on the screen to be concordant with known intersections of the 1cm^2 loom (for clarity a single Gaussian blob is shown). This is carried out separately for each eye, viewing distance and eye height used. This allows a mapping between loom (X_L, Y_L) and pixel (X_P, Y_P) coordinates to be obtained. For further description see the accompanying text.

A central region of the loom was selected in which to present stimuli, this region was sufficient spatially for presentation of the stimuli. Distortions in screen geometry are also known to increase with distance from the centre of the screen, so it is beneficial to calibrate only a central region of the loom/screen for stimulus presentation. During calibration a grid of Gaussian blobs is presented on the screen to be lined up with intersections of the selected loom area. For example, if stimuli were going to be presented within a 4cm^2 area of the plane of the loom, and calibration was going to be carried out per centimetre; a grid of 16 Gaussian blobs would be presented on the screen i.e. one blob per intersection of the 4cm^2 area of the loom.

The positions of these blobs is then adjusted, separately for each eye, so that the visual direction of the grid of 16 blobs on the screen was concordant with the 16 intersections of the 4cm^2 region of the loom. This would result in a set of 16 x and y loom coordinates in centimetres (X_L, Y_L), and a corresponding set of 16 x and y screen coordinates in pixels (X_P, Y_P), for each eye. Two-dimensional second-order polynomial equations are then used to give a function relating the x and y dimensions of the loom to the x and y dimensions of the screen (Equations A3.1 and A3.2 respectively, X_P and Y_P refer to the screen coordinates, X_L and Y_L to the loom coordinates, the lower case letters a to l are coefficients).

$$X_P = aX_L + bY_L + cX_L{}^2 + dY_L{}^2 + eX_LY_L + f \tag{A3.1}$$

$$Y_P = gX_L + hY_L + iX_L{}^2 + jY_L{}^2 + kX_LY_L + l \tag{A3.2}$$

The polynomial equations used to map screen to loom coordinates give a fit up to second order terms, the more distorted outer regions of the screen may be fit more poorly by these equations, which would in turn influence the overall fit for the whole calibrated area. Calibration is therefore more accurate if only a central region of the loom/screen sufficient in which to present the experimental stimuli is calibrated. A higher order polynomial would better fit more complex screen nonlinearities, but by increasing the order of the polynomial, the fit moves away from smoothing out the effects of monitor distortion, which is wanted, toward fitting the polynomial equation to these distortions, which is clearly undesirable. A second order polynomial was chosen based on advice received and piloting of the calibration procedure. Piloting showed a second order polynomial to produce Gaussian blob positions in good concordance with the filaments of the loom.

Initially the grid of Gaussian blobs is presented using an approximate screen to loom conversion factor calculated from the measured dimensions of the visible screen in centimetres and the screen's pixel resolution. This allows the blobs to be presented in approximately correct positions pre-calibration. The background colour of the screen was dark grey, individual Gaussian blobs were either white to signal their position was to be adjusted or light grey. This allowed clear identification of the blob whose position was to be adjusted. When global adjustments of the size/position of the grid were being made, the central blob of the grid was white and the remaining blobs were light grey. This allowed the central position of the grid to be identified. Ideally the screen calibration would be carried out using red Gaussian blobs on a black screen, so as to match the rendering of the experimental stimuli exactly. However, this makes the loom invisible against the screen. A dark grey background was chosen to give sufficient backlighting to the loom over the range of distances calibrated, whilst being minimally bright. The Gaussian blobs were positioned and adjusted with sub-pixel accuracy, in the same way as the experimental stimuli were rendered.

The global position of the grid is first set so that the central Gaussian blob is aligned with the central intersection of the loom. The central intersection of the loom is identifiable as the filaments 1cm in each direction from the looms centre are white, whereas the rest of the loom filaments are black. Next, the grid is scaled globally in size so that the blob positions are as near as possible to the loom intersections within the calibration area. The program then cycles through the Gaussian blobs working from the loom's centre outwards. The observer positions each blob accurately at the appropriate intersection of the loom. At all stages adjustments are made from a course to fine scale, under control of the observer. The blob screen coordinates in pixels and loom intersection coordinates in centimetres/metres are then fit with the polynomial equations. A new twice as dense grid of Gaussian blobs is then displayed as calculated using the polynomial equations. If the equations provide a good mapping between loom and pixel coordinates all blobs should lay at loom filament intersections or along the loom filaments between intersections, this depending on how densely the calibration is carried out e.g. every 1cm or every 2cm.

The individual blob positioning section of the procedure is then repeated, except now the additional blobs that were generated for presentation in the denser grid remain as a 'background' to the previously adjusted blobs, whose position is now re-checked. The additional blobs remain on the screen and are presented in light grey, but otherwise the procedure is identical. The program cycles through each of the previously adjusted Gaussian blobs and the observer rechecks positioning relative to the loom. This positioning is that determined by the polynomial equations. If the blob is positioned well its position is not changed, else its position is readjusted. The polynomial equations are then refit to the new pixel coordinates and a new dense Gaussian blob grid is produced and displayed. This provides a final check as to the polynomial fit. The calibration procedure is then complete, the coefficients of equations (A3.1) and (A3.2) are saved and used during stimulus generation to convert stimulus coordinates in centimetres/metres to screen pixel coordinates.

7.4 Scaling distances for the apparently circular cylinder task (ACC) and the triangle task

7.4.1 Calculating scaling distance

The scaling distance D' is calculated using the equation A4.1, here θ represents the half height or width of the cylinder (height for horizontally orientated cylinders, width for vertically orientated cylinders), η represents the disparity generated by the maximum depth of the cylinder δ and I represents the interocular distance of the observer.

$$D' = \frac{\theta I}{\eta - \eta\theta}$$

(A4.1)

Following Johnston (1991) for the apparently circular cylinder task disparity η was calculated using equation 2.

$$\eta = \frac{\delta I}{D^2 - \delta D}$$

(A4.2)

In the case of the triangle task the equation A4.1 is used to calculate the scaling distances, but θ now represents the half height or width of the triangle (height for horizontally orientated triangles, width for vertically orientated triangles) and η now represents the disparity generated by half the set depth δ of the triangle stimulus, calculated using equation A4.3.

$$\eta = \frac{(\delta/2)I}{D^2 - (\delta/2)D}$$

(A4.3)

7.4.2 Graphing scaling distance error bars

It is possible to calculate error bars for a scaling distance corresponding to the error bars around the depth value that was used to derive the scaling distance, whether this depth value is the mean of a set of values for set depth or is estimated from the PSE of a psychometric function. Take as an example the case where an observer sets a mean depth D to match the width W in an apparently circular cylinder task and 95% confidence intervals are estimated for this value ($\pm CI_{95}$). These confidence intervals define a maximum and minimum depth around the mean, $D + CI_{95}$ and $D - CI_{95}$ to be plotted as error bars. In the same way that a scaling distance SD_D can be calculated from the set depth D, scaling distances can be calculated from $D \pm CI_{95}$, these values define the maximum and minimum values of the error bars around the scaling distance value. Importantly, the scaling distance calculated from $D + CI_{95}$ defines the *lower* error bars limit and the scaling distance calculated from $D - CI_{95}$ defines the *upper* error bars limit (Figure A4.1). This reversal is due to the way in which the scaling distance is related to the set depth.

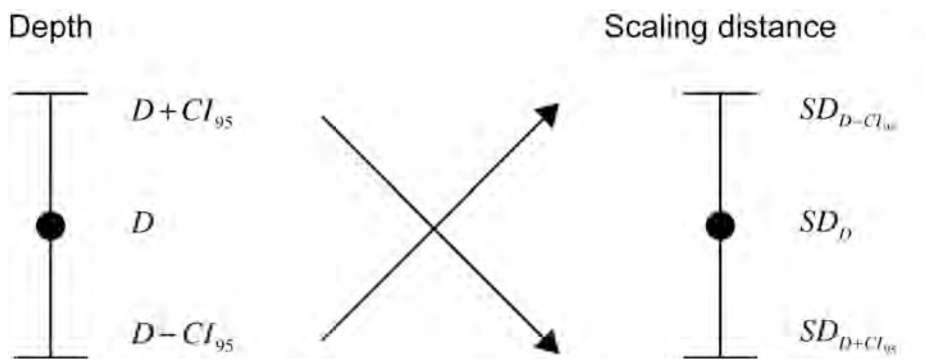

Figure A4.1: Showing variability around set depth in an ACC type task is converted into variability around a scaling distance value consistent with that depth.

The scaling distance is the same as the physical viewing distance when depth is set correctly, greater than the physical viewing distance when less depth is set than is needed and less than the physical viewing distance when more depth is set than needed. More generally, larger set depths result in smaller scaling distances and smaller set depths in larger scaling distances. As a consequence of this the scaling distance calculated from $D + CI_{95}$ is less than the scaling distance calculated from D ($SD_{D+CI_{95}} < SD_D$) and the scaling distance calculated from $D - CI_{95}$ is greater than the scaling distance calculated from D ($SD_{D-CI_{95}} > SD_D$). Therefore, the upper error bar of the depth setting defines the lower error bar of the scaling distance value and the lower error bar of the depth setting the upper error bar of the scaling distance value. This is not so important when the limits around a depth value are symmetrical about the mean because the upper and lower bar will be the same length. It is however important when the limits around the depth value are non-symmetrical, this is possible when confidence intervals are estimated around a PSE value derived from a psychometric function (Wichmann & Hill, 2001a, Wichmann & Hill, 2001b).

References

8 References

Adams, W.J., Graf, E.W., & Ernst, M.O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience, 7* (10), 1057-1058.
Aglioti, S., Desouza, J.F.X., & Goodale, M.A. (1995). Size-Contrast Illusions Deceive the Eye but Not the Hand. *Current Biology, 5* (6), 679-685.
Akeley, K., Watt, S.J., Girshick, A.R., & Banks, M.S. (2004). A stereo display prototype with multiple focal distances. *Acm Transactions on Graphics, 23* (3), 804-813.
Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14* (3), 257-262.
Assad, C., Rasnow, B., & Stoddard, P.K. (1999). Electric organ discharges and electric images during electrolocation. *Journal of Experimental Biology, 202* (10), 1185-1193.
Atkins, J.E., Fiser, J., & Jacobs, R.A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research, 41* (4), 449-461.
Atkins, J.E., Jacobs, R.A., & Knill, D.C. (2003). Experience-dependent visual cue recalibration based on discrepancies between visual and haptic percepts. *Vision Research, 43* (25), 2603-2613.
Backus, B.T., Banks, M.S., van Ee, R., Crowell, J.A., & Crowell, D. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research, 39* (6), 1143-1170.
Backus, B.T., & Matza-Brown, D. (2003). The contribution of vergence change to the measurement of relative disparity. *Journal of Vision, 3* (11), 737-750.
Barlow, H.B., & Foldiak, P. (1989). Adaptation and decorrelation in the cortex. In: C. Miall, R.M. Durbin, & G.J. Mitchison (Eds.), *The Computing Neuron* (pp. 54-72). Wokingham, England: Addison-Wesley.
Bingham, G.P. (2005). Calibration of distance and size does not calibrate shape information: Comparison of dynamic monocular and static and dynamic binocular vision. *Ecological Psychology, 17* (2), 55-74.
Bingham, G.P., Bradley, A., Bailey, M., & Vinner, R. (2001). Accommodation, occlusion, and disparity matching are used to guide reaching: A comparison of actual versus virtual environments. *Journal of Experimental Psychology-Human Perception and Performance, 27* (6), 1314-1334.
Bradshaw, M.F., & Elliott, K.M. (2003). The role of binocular information in the 'on-line' control of prehension. *Spatial Vision, 16* (3-4), 295-309.
Bradshaw, M.F., Elliott, K.M., Watt, S.J., Hibbard, P.B., Davies, I.R.L., & Simpson, P.J. (2004). Binocular cues and the control of prehension. *Spatial Vision, 17* (1-2), 95-110.
Bradshaw, M.F., & Glennerster, A. (2006). Stereoscopic acuity and observation distance. *Spatial Vision, 19* (1), 21-36.
Bradshaw, M.F., Glennerster, A., & Rogers, B.J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Research, 36* (9), 1255-1264.

Bradshaw, M.F., Parton, A.D., & Eagle, R.A. (1998). The interaction of binocular disparity and motion parallax in determining perceived depth and perceived size. *Perception, 27* (11), 1317-1331.
Bradshaw, M.F., Parton, A.D., & Glennerster, A. (2000). The task-dependent use of binocular disparity and motion parallax information. *Vision Research, 40* (27), 3725-3734.
Bradshaw, M.F., & Rogers, B.J. (1999). Sensitivity to horizontal and vertical corrugations defined by binocular disparity. *Vision Research, 39* (18), 3049-3056.
Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision, 10* (4), 433-436.
Brenner, E., & Landy, M.S. (1999). Interaction between the perceived shape of two objects. *Vision Research, 39* (23), 3834-3848.
Brenner, E., & Smeets, J.B.J. (2000). Comparing extra-retinal information about distance and direction. *Vision Research, 40* (13), 1649-1651.
Brenner, E., & Smeets, J.B.J. (2001). We are better off without perfect perception. *Behavioral and Brain Sciences, 24* (2), 215-+.
Brenner, E., & Smeets, J.B.J. (2006). Two eyes in action. *Experimental Brain Research, 170* (3), 302-311.
Brenner, E., & van Damme, W.J.M. (1998). Judging distance from ocular convergence. *Vision Research, 38* (4), 493-498.
Brenner, E., & van Damme, W.J.M. (1999). Perceived distance, shape and size. *Vision Research, 39* (5), 975-986.
Brooks, K.R., & Stone, L.S. (2004). Stereomotion speed perception: Contributions from both changing disparity and interocular velocity difference over a range of relative disparities. *Journal of Vision, 4* (12), 1061-1079.
Buckley, D., & Frisby, J.P. (1993). Interaction of Stereo, Texture and Outline Cues in the Shape Perception of 3-Dimensional Ridges. *Vision Research, 33* (7), 919-933.
Burge, J., Peterson, M.A., & Palmer, S.E. (2005). Ordinal configural cues combine with metric disparity in depth perception. *Journal of Vision, 5* (6), 534-542.
Burt, P., & Julesz, B. (1980). Modifications of the classical notion of Panum's fusional area. *Perception, 9* (6), 671-682.
Castiello, U. (1996). Grasping a fruit: selection for action. *J Exp Psychol Hum Percept Perform, 22* (3), 582-603.
Champion, R.A., Simmons, D.R., & Mamassian, P. (2004). The influence of object size and surface shape on shape constancy from stereo. *Perception, 33* (2), 237-247.
Churchill, A., Hopkins, B., Ronnqvist, L., & Vogt, S. (2000). Vision of the hand and environmental context in human prehension. *Experimental Brain Research, 134* (1), 81-89.
Clark, J.J., & Yuille, A.L. (1990). Data Fusion for Sensory Information Processing Systems. (Boston: Kluwer Academic Publishers.
Collett, T.S., Schwarz, U., & Sobel, E.C. (1991). The interaction of oculomotor cues and stimulus size in stereoscopic death constancy. *Perception, 20* (6), 733-754.
Collewijn, H., & Erkelens, C.J. (1990). Binocular eye movements and the perception of depth. In: E. Kowler (Ed.) *Eye movement and their role in visual and cognitive processes* (pp. 213-261). Amsterdam: Elsevier.
Connolly, J.D., & Goodale, M.A. (1999). The role of visual feedback of hand position in the control of manual prehension. *Experimental Brain Research, 125* (3), 281-286.
Cuijpers, R.H., Brenner, E., & Smeets, J.B.J. (2006). Grasping reveals visual misjudgements of shape. *Experimental Brain Research, 175* (1), 32-44.
Cuijpers, R.H., Kappers, A.M.L., & Koenderink, J.J. (2000). Large systematic deviations in visual parallelism. *Perception, 29* (12), 1467-1482.

Cuijpers, R.H., Smeets, J.B.J., & Brenner, E. (2004). On the relation between object shape and grasping kinematics. *Journal of Neurophysiology, 91* (6), 2598-2606.
Cumming, B.G., Johnston, E.B., & Parker, a.J. (1991). Vertical Disparities and Perception of 3-Dimensional Shape. *Nature, 349* (6308), 411-413.
Cutting, J.E., & Vishton, P.M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In: W. Epstein, & S. Rogers (Eds.), *Perception of Space and Motion* (San Diego: Academic Press.
Dawkins, R. (1986). The Blind Watchmarker. (London: Penguin Books.
de Grave, D.D.J., Brenner, E., & Smeets, J.B.J. (2004). Illusions as a tool to study the coding of pointing movements. *Experimental Brain Research, 155* (1), 56-62.
Durgin, F.H., Proffitt, D.R., Olson, T.J., & Reinke, K.S. (1995). Comparing Depth from Motion with Depth from Binocular Disparity. *Journal of Experimental Psychology-Human Perception and Performance, 21* (3), 679-699.
Econopouly, J. (1996). Binocular Disparities and KDE are Combined to Rescale Binocular Disparities. *Center for Neural Science,* Doctor of Philosophy (New York: New York University.
Ernst, M.O. (2006). A Bayesian view on multimodal cue integration. In: G. Knoblich, I.M. Thornton, M. Grosjean, & M. Shiffrar (Eds.), *Human body perception from the inside out* (pp. 105-131). New York: Oxford University Press.
Ernst, M.O., & Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415* (6870), 429-433.
Ernst, M.O., & Bulthoff, H.H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8* (4), 162-169.
Fermuller, C., & Malm, H. (2004). Uncertainty in visual processes predicts geometrical optical illusions. *Vision Research, 44* (7), 727-749.
Foley, J.M. (1976). Successive stereo and vernier position discrimination as a function of dark interval duration. *Vision Res, 16* (11), 1269-1273.
Foley, J.M. (1980). Binocular Distance Perception. *Psychological Review, 87* (5), 411-434.
Foley, J.M. (1985). Binocular distance perception: egocentric distance tasks. *J Exp Psychol Hum Percept Perform, 11* (2), 133-149.
Foley, J.M., & Held, R. (1972). Visually Directed Pointing as a Function of Target Distance, Direction, and Available Cues. *Perception & Psychophysics, 12* (3), 263-268.
Foster, D.H., & Bischof, W.F. (1991). Thresholds from Psychometric Functions - Superiority of Bootstrap to Incremental and Probit Variance Estimators. *Psychological Bulletin, 109* (1), 152-159.
Franz, V.H. (2001). Action does not resist visual illusions. *Trends in Cognitive Sciences, 5* (11), 457-459.
Franz, V.H. (2003). Manual size estimation: a neuropsychological measure of perception? *Experimental Brain Research, 151* (4), 471-477.
Franz, V.H., Bulthoff, H.H., & Fahle, M. (2003). Grasp effects of the Ebbinghaus illusion: obstacle avoidance is not the explanation. *Experimental Brain Research, 149* (4), 470-477.
Franz, V.H., Gegenfurtner, K.R., Bulthoff, H.H., & Fahle, M. (2000). Grasping visual illusions: No evidence for a dissociation between perception and action. *Psychological Science, 11* (1), 20-25.
Frisby, J.P., Buckley, D., & Duke, P.A. (1996). Evidence for good recovery of lengths of real objects seen with natural stereo viewing. *Perception, 25* (2), 129-154.

Frisby, J.P., Buckley, D., & Horsman, J.M. (1995). Integration of Stereo, Texture, and Outline Cues during Pinhole Viewing of Real Ridge-Shaped Objects and Stereograms of Ridges. *Perception, 24* (2), 181-198.
Garding, J., Porrill, J., Mayhew, J.E.W., & Frisby, J.P. (1995). Stereopsis, Vertical Disparity and Relief Transformations. *Vision Research, 35* (5), 703-722.
Gardner, P.L., & Mon-Williams, M. (2001). Vertical gaze angle: absolute height-in-scene information for the programming of prehension. *Experimental Brain Research, 136* (3), 379-385.
Gepshtein, S., & Banks, M.S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology, 13* (6), 483-488.
Gibson, J.J. (1950). The Perception of the Visual World. (Boston: Houghton Mifflin.
Gillam, B., & Borsting, E. (1988). The Role of Monocular Regions in Stereoscopic Displays. *Perception, 17* (5), 603-608.
Gillam, B., Chambers, D., & Russo, T. (1988). Postfusional latency in stereoscopic slant perception and the primitives of stereopsis. *J Exp Psychol Hum Percept Perform, 14* (2), 163-175.
Gillam, B., Flagg, T., & Finlay, D. (1984). Evidence for disparity change as the primary stimulus for stereoscopic processing. *Percept Psychophys, 36* (6), 559-564.
Girshick, A.R., Akeley, K., Watt, S.J., & Banks, M.S. (2004). Improved stereoscopic performance with consistent vergence and accommodative cues in a novel 3-D display. *Perception, 33*, 42-42.
Glennerster, A., McKean, D., & Gilson, S.J. (2006a). Manipulating prior assumptions about 3D stability. *Journal of Vision, 6* (6), 739.
Glennerster, A., McKee, S.P., & Birch, M.D. (2002). Evidence for surface-based processing of binocular disparity. *Current Biology, 12* (10), 825-828.
Glennerster, A., Rogers, B.J., & Bradshaw, M.F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research, 36* (21), 3441-3456.
Glennerster, A., Rogers, B.J., & Bradshaw, M.F. (1998). Cues to viewing distance for stereoscopic depth constancy. *Perception, 27* (11), 1357-1365.
Glennerster, A., Tcheang, L., Gilson, S.J., Fitzgibbon, A.W., & Parker, A.J. (2006b). Humans ignore motion and stereo cues in favor of a fictional stable world. *Current Biology, 16* (4), 428-432.
Gogel, W.C. (1965). Equidistance Tendency And Its Consequences. *Psychol Bull, 64*, 153-163.
Gogel, W.C. (1969). The sensing of retinal size. *Vision Res, 9* (9), 1079-1094.
Gogel, W.C. (1977). Indirect Measure of Perceived Distance from Oculomotor Cues. *Perception & Psychophysics, 21* (1), 3-11.
Gogel, W.C., & Tietz, J.D. (1973). Absolute Motion Parallax and Specific Distance Tendency. *Perception & Psychophysics, 13* (2), 284-292.
Haffenden, A.M., & Goodale, M.a. (1998). The effect of pictorial illusion on prehension and perception. *Journal of Cognitive Neuroscience, 10* (1), 122-136.
Haffenden, A.M., & Goodale, M.a. (2000). Independent effects of pictorial displays on perception and action. *Vision Research, 40* (10-12), 1597-1607.
Haffenden, A.M., Schiff, K.C., & Goodale, M.A. (2001). The dissociation between perception and action in the Ebbinghaus illusion: Nonillusory effects of pictorial cues on grasp. *Current Biology, 11* (3), 177-181.
Haijiang, Q., Saunders, J.A., Stone, R.W., & Backus, B.T. (2006). Demonstration of cue recruitment: change in visual appearance by means of Pavlovian conditioning. *Proc Natl Acad Sci U S A, 103* (2), 483-488.

Harris, J.M. (2004). Binocular vision: moving closer to reality. *Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences, 362* (1825), 2721-2739.
Harris, J.M., & Dean, P.J.A. (2003). Accuracy and precision of binocular 3-D motion perception. *Journal of Experimental Psychology-Human Perception and Performance, 29* (5), 869-881.
Harris, J.M., & Watamaniuk, S.N. (1995). Speed discrimination of motion-in-depth using binocular cues. *Vision Res, 35* (7), 885-896.
Helmholtz, H. (1925). Helmholtz's Treatise on Physiological Optics. 3 (Thoemmes Press.
Hershenson, M.H. (1999). Visual space perception: a primer. (London: MIT Press.
Hibbard, P.B. (In Press). A Statistical of Binocular Disparity. *Visual Cognition,*
Hibbard, P.B., & Bradshaw, M.F. (2003). Reaching for virtual objects: binocular disparity and the control of prehension. *Experimental Brain Research, 148* (2), 196-201.
Hibbard, P.B., Bradshaw, M.F., Langley, K., & Rogers, B.J. (2002). The stereoscopic anisotropy: Individual differences and underlying mechanisms. *Journal of Experimental Psychology-Human Perception and Performance, 28* (2), 469-476.
Hillis, J.M., Ernst, M.O., Banks, M.S., & Landy, M.S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science, 298* (5598), 1627-1630.
Hillis, J.M., Watt, S.J., Landy, M.S., & Banks, M.S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4* (12), 967-992.
Hogervorst, M.A., & Eagle, R.A. (1998). Biases in three-dimensional structure-from-motion arise from noise in the early visual system. *Proceedings of the Royal Society of London Series B-Biological Sciences, 265* (1406), 1587-1593.
Howard, I.P. (2002). Seeing in Depth: Basic Mechanisms. 1 (Toronto: I Porteous.
Howard, I.P., & Rogers, B.J. (2002). Seeing in Depth: Depth Perception. 2 (Toronto: I Porteous.
Hung, G.K., Zhu, H., & Ciuffreda, K.J. (1997). Convergence and divergence exhibit different response characteristics to symmetric stimuli. *Vision Res, 37* (9), 1197-1205.
Ivanoff, A. (1955). Night binocular convergence and night myopia. *J Opt Soc Am, 45* (9), 769-770.
Jackson, S.R., Jackson, G.M., & Rosicky, J. (1995). Are Non-Relevant Objects Represented in Working-Memory - the Effect of Nontarget Objects on Reach and Grasp Kinematics. *Experimental Brain Research, 102* (3), 519-530.
Jackson, S.R., Jones, C.A., Newport, R., & Pritchard, C. (1997). A kinematic analysis of goal-directed prehension movements executed under binocular, monocular, and memory-guided viewing conditions. *Visual Cognition, 4* (2), 113-142.
Jacobs, R.A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences, 6* (8), 345-350.
Jeannerod, M. (1988). The Neural and Behavioural Organisation of Goal-Directed Movements. (Oxford: Clarendon Press.
Johnston, E.B. (1991). Systematic Distortions of Shape from Stereopsis. *Vision Research, 31* (7-8), 1351-1360.
Johnston, E.B., Cumming, B.G., & Landy, M.S. (1994). Integration of Stereopsis and Motion Shape Cues. *Vision Research, 34* (17), 2259-2275.
Johnston, E.B., Cumming, B.G., & Parker, a.J. (1993). Integration of Depth Modules - Stereopsis and Texture. *Vision Research, 33* (5-6), 813-826.

Julesz, B. (1971). Foundations of Cyclopean Perception. (Chicago: Chicago University Press.
Knill, D.C. (1998). Surface orientation from texture: ideal observers, generic observers and the information content of texture cues. *Vision Research, 38* (11), 1655-1682.
Knill, D.C. (2005). Reaching for visual cues to depth: The brain combines depth cues differently for motor control and perception. *Journal of Vision, 5* (2), 103-115.
Knill, D.C., & Richards, W. (1996). Perception as Bayesian Inference. (Cambridge: Cambridge University Press.
Koenderink, J. (2001). Multiple visual worlds. *Perception, 30* (1), 1-7.
Koenderink, J.J. (1998). Pictorial relief. *Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences, 356* (1740), 1071-1085.
Koenderink, J.J., van Doorn, A.J., Kappers, A.M.L., & Todd, J.T. (2002). Pappus in optical space. *Perception & Psychophysics, 64* (3), 380-391.
Koenderink, J.J., van Doorn, A.J., & Lappin, J.S. (2000). Direct measurement of the curvature of visual space. *Perception, 29* (1), 69-79.
Kritikos, A., Bennett, K.M., Dunai, J., & Castiello, U. (2000). Interference from distractors in reach-to-grasp movements. *Q J Exp Psychol A, 53* (1), 131-151.
Lages, M. (2006). Bayesian models of binocular 3-D motion perception. *Journal of Vision, 6* (4), 508-522.
Landy, M.S., & Brenner, E. (2001). Motion-disparity interaction and the scaling of stereoscopic disparity. In: L.R. Harris, & M.R.M. Jenkin (Eds.), *Vision and Attention* (pp. 129-151). New York: Springer Verlag.
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.
Loomis, J.M., & Philbeck, J.W. (1999). Is the anisotropy of perceived 3-D shape invariant across scale? *Perception & Psychophysics, 61* (3), 397-402.
Loomis, J.M., Philbeck, J.W., & Zahorik, P. (2002). Dissociation between location and shape in visual space. *Journal of Experimental Psychology-Human Perception and Performance, 28* (5), 1202-1212.
Mamassian, P., Landy, M.S., & Maloney, L.T. (2002). Bayesian Modelling of Visual Perception. In: R.P.N. Rao, B.A. Olshausen, & M.S. Lewicki (Eds.), *Probabilistic Models of the Brain: Perception and Neural Function* (pp. 13-36).
Marotta, J.J., & Goodale, M.a. (1998). The role of learned pictorial cues in the programming and control of grasping. *Experimental Brain Research, 121* (4), 465-470.
Marr, D. (1982). Vision. (San Francisco: W. H. Freeman.
Mayhew, J.E.W., & Longuet-Higgins, H.C. (1982). A Computational Model of Binocular Depth-Perception. *Nature, 297* (5865), 376-378.
Melmoth, D.R., & Grant, S. (2006). Advantages of binocular vision for the control of reaching and grasping. *Exp Brain Res, 171* (3), 371-388.
Meng, J.C., & Sedgwick, H.A. (2001). Distance perception mediated through nested contact relations among surfaces. *Perception & Psychophysics, 63* (1), 1-15.
Milner, A.D., & Goodale, M.A. (1995). The visual brain in action. (Oxford: Oxford University Press.
Milner, A.D., & Goodale, M.A. (2006). The visual brain in action: second edition. (Oxford: Oxford University Press.

Mon-Williams, M., & Tresilian, J.R. (1999). Some recent studies on the extraretinal contribution to distance perception. *Perception, 28* (2), 167-181.
Mon-Williams, M., Tresilian, J.R., Coppard, V.L., & Carson, R.G. (2001). The effect of obstacle position on reach-to-grasp movements. *Experimental Brain Research, 137* (3-4), 497-501.
Mon-Williams, M., Tresilian, J.R., & Roberts, A. (2000). Vergence provides veridical depth perception from horizontal retinal image disparities. *Experimental Brain Research, 133* (3), 407-413.
Morgan, M.J. (1989). Stereopsis - Vision of Solid Objects. *Nature, 339* (6220), 101-103.
Moss, C.F., & Sinha, S.R. (2003). Neurobiology of echolocation in bats. *Current Opinion in Neurobiology, 13* (6), 751-758.
Norman, J.F., Todd, J.T., Perotti, V.J., & Tittle, J.S. (1996). The visual perception of three-dimensional length. *Journal of Experimental Psychology-Human Perception and Performance, 22* (1), 173-186.
O'Kane, L.M., & Hibbard, P.B. (2002). Vertical-disparity pooling across spatially segregated surfaces. *Perception, 31*, 154-155.
Ooi, T.L., Wu, B., & He, Z.J.J. (2001). Distance determined by the angular declination below the horizon. *Nature, 414* (6860), 197-200.
Ooi, T.L., Wu, B., & He, Z.J.J. (2006). Perceptual space in the dark affected by the intrinsic bias of the visual system. *Perception, 35* (5), 605-624.
Oruc, I., Maloney, L.T., & Landy, M.S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43* (23), 2451-2468.
Owens, D.A., & Liebowitz, H.W. (1976). Oculomotor adjustments in darkness and the specific distance tendency. *Vision Res, 20* (1), 2-9.
Owens, D.A., & Liebowitz, H.W. (1980). Accommodation, convergence, and distance perception in low illumination. *Am J Optom Physiol Opt, 57* (9), 540-550.
Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10* (4), 437-442.
Poulton, E.C. (1989). Bias in Quantifying Judgments. (London: Lawrence Erlbaum Associates.
Ramachandran, V.S. (1985). The neurobiology of perception. *Perception, 14* (2), 97-103.
Richards, W. (1985). Structure from Stereo and Motion. *Journal of the Optical Society of America a-Optics Image Science and Vision, 2* (2), 343-349.
Rogers, B.J., & Bradshaw, M.F. (1993). Vertical Disparities, Differential Perspective and Binocular Stereopsis. *Nature, 361* (6409), 253-255.
Rogers, B.J., & Bradshaw, M.F. (1995). Disparity Scaling and the Perception of Frontoparallel Surfaces. *Perception, 24* (2), 155-179.
Rogers, B.J., & Collett, T.S. (1989). The Appearance of Surfaces Specified by Motion Parallax and Binocular Disparity. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 41* (4), 697-717.
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. *Journal of the Optical Society of America a-Optics Image Science and Vision, 22* (5), 801-809.
Saunders, J.A., & Knill, D.C. (2003). Humans use continuous visual feedback from the hand to control fast reaching movements. *Experimental Brain Research, 152* (3), 341-352.

Saunders, J.A., & Knill, D.C. (2004). Visual feedback control of hand movements. *Journal of Neuroscience, 24* (13), 3223-3234.
Saunders, J.A., & Knill, D.C. (2005). Humans use continuous visual feedback from the hand to control both the direction and distance of pointing movements. *Experimental Brain Research, 162* (4), 458-473.
Servos, P. (2000). Distance estimation in the visual and visuomotor systems. *Exp Brain Res, 130* (1), 35-47.
Servos, P., & Goodale, M.A. (1994). Binocular Vision and the Online Control of Human Prehension. *Experimental Brain Research, 98* (1), 119-127.
Servos, P., Goodale, M.A., & Jakobson, L.S. (1992). The Role of Binocular Vision in Prehension - a Kinematic Analysis. *Vision Research, 32* (8), 1513-1521.
Sinai, M.J., Ooi, T.L., & He, Z.J. (1998). Terrain influences the accurate judgement of distance. *Nature, 395* (6701), 497-500.
Smeets, J.B.J., Brenner, E., de Grave, D.D.J., & Cujpers, R.H. (2002). Illusions in action: consequences of inconsistent processing of spatial attributes. *Experimental Brain Research, 147* (2), 135-144.
Stoper, a.E., & Cohen, M.M. (1986). Judgments of Eye Level in Light and in Darkness. *Perception & Psychophysics, 40* (5), 311-316.
Swenson, H.A. (1932). The relative influence of accommodation and convergence in the judgment of distance. *Journal of General Psychology, 7* (360-380)
Tcheang, L., Gilson, S.J., & Glennerster, A. (2005). Systematic distortions of perceptual stability investigated using immersive virtual reality. *Vision Research, 45* (16), 2177-2189.
Tipper, S.P., Howard, L.A., & Jackson, S.R. (1997). Selective reaching to grasp: Evidence for distractor interference effects. *Visual Cognition, 4* (1), 1-38.
Tittle, J.S., & Braunstein, M.L. (1993). Recovery of 3-D Shape from Binocular Disparity and Structure-from-Motion. *Perception & Psychophysics, 54* (2), 157-169.
Tittle, J.S., Todd, J.T., Perotti, V.J., & Norman, J.F. (1995). Systematic Distortion of Perceived 3-Dimensional Structure-from-Motion and Binocular Stereopsis. *Journal of Experimental Psychology-Human Perception and Performance, 21* (3), 663-678.
Todd, J.T. (2004). The visual perception of 3D shape. *Trends in Cognitive Sciences, 8* (3), 115-121.
Todd, J.T., & Norman, J.F. (2003). The visual perception of 3-D shape from multiple cues: Are observers capable of perceiving metric structure? *Perception & Psychophysics, 65* (1), 31-47.
Todd, J.T., Tittle, J.S., & Norman, J.F. (1995). Distortions of 3-Dimensional Space in the Perceptual Analysis of Motion and Stereo. *Perception, 24* (1), 75-86.
Tresilian, J.R. (1998). Attention in action or obstruction of movement? A kinematic analysis of avoidance behavior in prehension. *Experimental Brain Research, 120* (3), 352-368.
Tresilian, J.R., & Mon-Williams, M. (1999). A curious illusion suggests complex cue interactions in distance perception. *Journal of Experimental Psychology-Human Perception and Performance, 25* (3), 677-687.
Tresilian, J.R., Mon-Williams, M., & Kelly, B.M. (1999). Increasing confidence in vergence as a cue to distance. *Proceedings of the Royal Society of London Series B-Biological Sciences, 266* (1414), 39-44.
Trommershauser, J., Maloney, L.T., & Landy, M.S. (2003). Statistical decision theory and the selection of rapid, goal-directed movements. *Journal of the Optical Society of America a-Optics Image Science and Vision, 20* (7), 1419-1433.

van Damme, W., & Brenner, E. (1997). The distance used for scaling disparities is the same as the one used for scaling retinal size. *Vision Research, 37* (6), 757-764.
van Ee, R., Banks, M.S., & Backus, B.T. (1999). An analysis of binocular slant contrast. *Perception, 28* (9), 1121-1145.
vanDamme, W., & Brenner, E. (1997). The distance used for scaling disparities is the same as the one used for scaling retinal size. *Vision Research, 37* (6), 757-764.
Viguier, a., Clement, G., & Trotter, Y. (2001). Distance perception within near visual space. *Perception, 30* (1), 115-124.
Wagner, M. (1985). The metric of visual space. *Percept Psychophys, 38* (6), 483-495.
Walker, M.M., Dennis, T.E., & Kirschvink, J.L. (2002). The magnetic sense and its use in long-distance navigation by animals. *Current Opinion in Neurobiology, 12* (6), 735-744.
Watt, S.J., Akeley, K., Ernst, M.O., & Banks, M.S. (2005a). Focus cues affect perceived depth. *Journal of Vision, 5* (10), 834-862.
Watt, S.J., Akeley, K., Girshick, A.R., & Banks, M.S. (2005b). Achieving near-correct focus cues in a 3-D display using multiple image planes. *Proceedings of SPIE: Human Vision and Electronic Imaging*, 5666-5673.
Watt, S.J., & Bradshaw, M.F. (2000). Binocular cues are important in controlling the grasp but not the reach in natural prehension movements. *Neuropsychologia, 38* (11), 1473-1481.
Watt, S.J., & Bradshaw, M.F. (2002). Binocular information in the control of prehensile movements in multiple-object scenes. *Spatial Vision, 15* (2), 141-155.
Watt, S.J., & Bradshaw, M.F. (2003). The visual control of reaching and grasping: Binocular disparity and motion parallax. *Journal of Experimental Psychology-Human Perception and Performance, 29* (2), 404-415.
Weiss, Y., Simoncelli, E.P., & Adelson, E.H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5* (6), 598-604.
Welchman, A.E., Deubelius, A., Conrad, V., Bulthoff, H.H., & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience, 8* (6), 820-827.
Welchman, A.E., Tuck, V.L., & Harris, J.M. (2004). Human observers are biased in judging the angular approach of a projectile. *Vision Research, 44* (17), 2027-2042.
Westheimer, G. (1979). Cooperative Neural Processes Involved in Stereoscopic Acuity. *Experimental Brain Research, 36* (3), 585-597.
Wichmann, F.A., & Hill, N.J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63* (8), 1293-1313.
Wichmann, F.A., & Hill, N.J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics, 63* (8), 1314-1329.
Wing, A.M., & Fraser, C. (1983). The contribution of the thumb to reaching movements. *Q J Exp Psychol A, 35* (Pt 2), 297-309.
Yang, Z.Y., & Purves, D. (2003). A statistical explanation of visual space. *Nature Neuroscience, 6* (6), 632-640.
Zabulis, X., & Backus, B.T. (2004). Starry night: a texture devoid of depth cues. *J Opt Soc Am A Opt Image Sci Vis, 21* (11), 2049-2060.
Zeki, S. (1993). A Vision of the Brain. (Oxford: Blackwell.