Does semantic size affect size constancy scaling using lexical stimuli?

4625347

MSc Psycholinguistics

The University of Edinburgh

2010

Declaration

I have read and understood the University of Edinburgh guidelines on plagiarism and declare that this written dissertation is all my own work, except where I indicate otherwise by proper use of quotes and references.

Abstract

Binocular disparity allows us to perceive the world in 3-dimensions through the process of stereopsis. In this study, we used binocular disparity to induce the size constancy illusion in lexical stimuli. 47 undergraduate and postgraduate students took part in a within-subjects, repeated measures design. Pairs of words were presented dichoptically using a mirror stereoscope. Results showed a significant interaction between sex, and whether an individual reported perceiving depth. Further analysis showed that in males, the size constancy effect was significantly stronger when the "further" word was presented to the upper visual field, and in females, the effect was significantly stronger when the "further" word was presented to the lower visual field. There was no effect of semantic size, nor of any other semantic variable (concreteness, imageability, semantic category) on the size constancy illusion.

Contents

1. Introduction

Binocular disparity is one of a range of depth cues that we use to perceive depth in every day life, and can be used to elicit the size constancy illusion in the absence of other, monocular depth cues using random-dot stereograms (Julesz, 1963, 1971). The present paper provides the first attempt to demonstrate that binocular disparity can be used to induce the size constancy illusion using lexical stimuli. In addition, the embodied cognition literature provides increasing evidence that understanding lexical stimuli involves mental simulations which can encode a great deal of perceptual information (Zwaan, 2004). Other research shows that the real-world size of an object can affect the way in which we process those objects' names (e.g. Sereno, O'Donnell, & Sereno, 2009), and that the visual system combines prior knowledge of objects with binocular disparity when making depth and distance judgements (Hartung, Schrater, Bülthoff, Kersten, & Franz, 2005). Therefore, this study also aims to investigate whether the size constancy illusion, elicited using concrete nouns as stimuli, is affected by knowledge about those objects' real world size. Finally, we will attempt to establish whether another depth cue (height in the visual field) affects the strength of the size constancy illusion: more specifically, are we more likely to experience the size constancy illusion in a more ecologically valid condition, in which the more "distant" stimuli is positioned in the upper visual field?

1.1. Binocular disparity and stereopsis

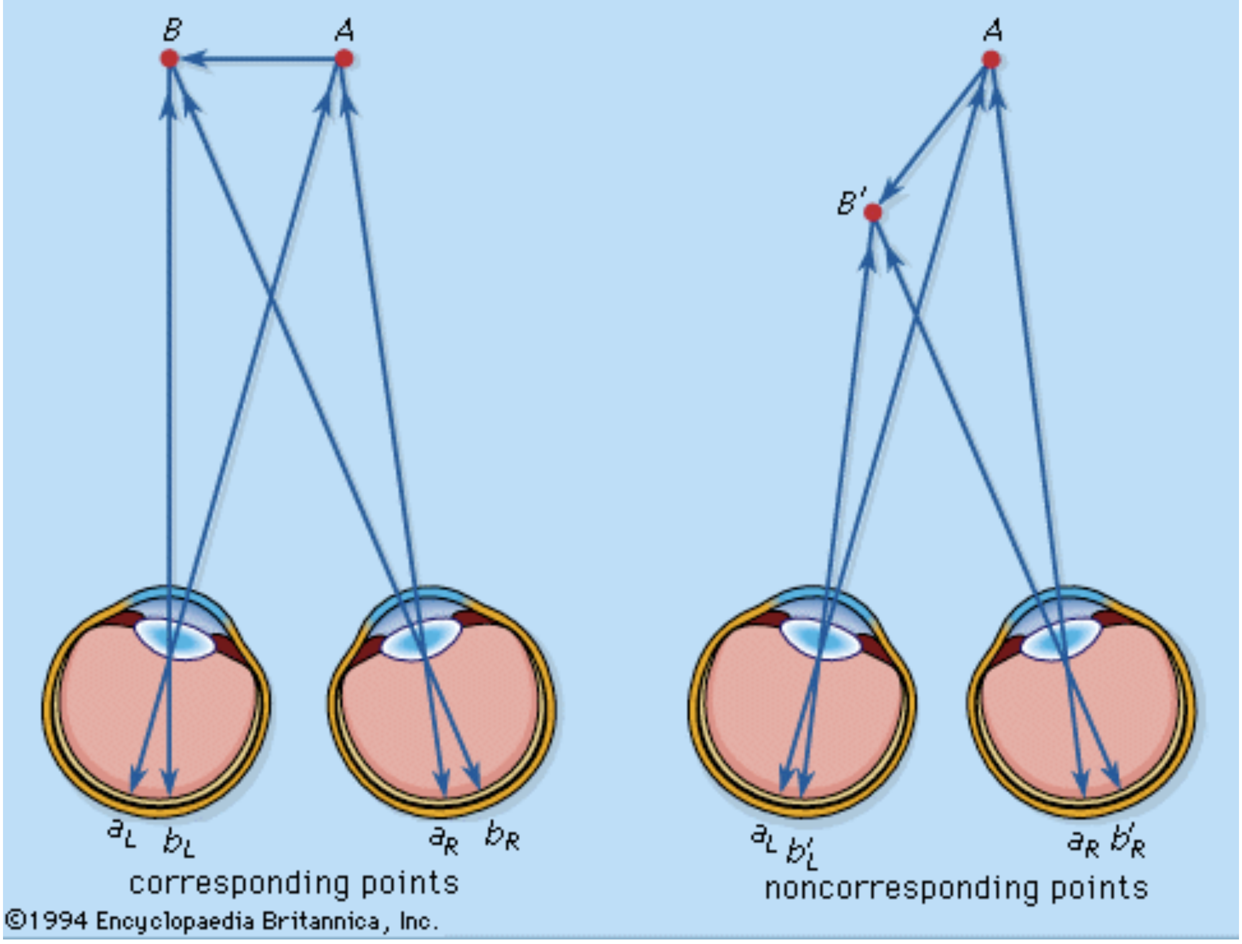
The fact that humans have two frontal eyes makes a non-trivial contribution to the way we view the world. Our eyes are positioned between 5.5 and 7.5 centimetres apart (Qian, 1997); each eye therefore provides a slightly different vantage point onto the world. The fusion of these two distinct, but overlapping images allows us to experience 3D vision through a process known as stereopsis (Wheatstone, 1938). This is the process which we shall use to elicit the size constancy illusion in lexical stimuli. Therefore, we begin by reviewing the basic processes involved in stereopsis, and some of the theoretical assumptions underpinning our use of binocular disparity as a research tool.

Stereopsis occurs when the visual system combines two slightly different retinal images of the same scene. The difference between these two images is called binocular disparity. Traditionally, research has focussed on horizontal disparities, assuming that vertical disparities have little or no role to play in stereopsis (e.g. Read & Cumming, 2006). While research now suggests that vertical disparities may, in fact be detectable by the human visual system, and used to infer depth perception (e.g. Bishop, 1994; Gårding, Porril, Mayhew, & Frisby, 2005; Matthews, Meng, Xu, & Qian, 2003), the present study will only manipulate horizontal disparity. As such, we will use the term binocular disparity to refer exclusively to horizontal disparities.

The binocular disparity of any given point can be defined as "the difference in retinal position of the left and right projections of the point" (Gårding *et al*., 2005, p.705,). By measuring this disparity, the visual system can estimate the relative depth of objects (Marr & Poggio, 1979). It should be noted at the outset that stereopsis is not synonymous with depth perception (Pollack 1955); although stereopsis provides a compelling 3D experience, other depth cues are available, and people who lack stereopsis are still able to perceive depth through these other cues (Mather, 2006). The relative importance of stereopsis and these cues will be discussed further in section 1.2, below.

In order to calculate the level of disparity, the visual system must first match an object-point in one retina to the corresponding object-point in the other retina (Marr & Poggio, 1979). This is known as the binocular correspondence problem; and, given the number of potential points in each eye, solving it is potentially very complex (Harris & Wilcox, 2009). Although several constraining factors, including similarity and continuity have been identified (see Mather, 2006, for a summary), the specifics of how the visual system solves this problem are still a matter of considerable debate (e.g. Hoffman & Banks, 2010; den Ouden, van Ee, & de Haan, 2008). If this matching of local features is not achieved, then interocular suppression occurs, and the input from one retina is inhibited (Baker & Graf, 2009). There is some debate as to whether interocular suppression occurs only when triggered by the failure of binocular fusion (e.g. Blake & Boothroyd, 1985), or whether both mechanisms operate independently from one another, and can therefore occur at the same time - the *co-existence hypothesis* (Su, He, & Ooi, 2009). The co-existence hypothesis is supported by evidence that participants are able to perceive depth in random dot stereograms even whilst experiencing binocular rivalry (Juselz & Miller, 1975). Binocular rivalry occurs when the input from one eye is inhibited, followed by the input from the second eye, in an alternating pattern (see Blake & Logothetis, 2002, for a review); if inhibition (and therefore binocular rivalry) only occur once fusion has failed, it should not be possible to achieve the fusion necessary to perceive depth in the stereogram. For current purposes however, it is sufficient to note that both binocular fusion and interocular suppression exist, and that fusion can provide a compelling impression of depth.

Once an object-point in one retina has been matched to the same object-point in the other retina, the disparity between those two object-points can be calculated. For any given retinal-point, there is a single corresponding point in the other retina. These pairs of points are known as corresponding points, and they can be defined in one of two ways: geometric corresponding points, and empirical corresponding points (Schreiber, Hillis, Filippini, Schor, & Banks, 2008). Geometric corresponding points are defined mathematically, in terms of their (physical) co-ordinates on the retinas in respect to the fovea (e.g. Howard & Rogers, 1995). Figure 1 shows examples of geometric corresponding and non-corresponding points.



©1994 Encyclopaedia Britannica, Inc.

Figure 1: Corresponding retinal points occupy corresponding locations in the left and right retinas, relative to the fovea. An image that falls on corresponding points has zero disparity. Non-corresponding points occupy different locations in the left and right retinas, relative to the fovea.

Empirical corresponding points are defined not by geometry, but by what the individual actually perceives: for example, while holding the eye's position fixed, a line is shown to one eye, and is then moved until the other eye perceives it in the same direction (Ogle, 1932). So far, research has failed to establish that empirical corresponding points coincide with geometric corresponding points (Helmholtz, 1925; Hillis & Banks, 2001; Ogle, 1950). This has led to some authors questioning the usefulness of geometric corresponding points (e.g. Glanville, 1933).

However, corresponding points remain an important concept in binocular disparity, because they allow us to posit an imaginary surface of zero disparity, known as the horopter. The horopter is the set of points in space that, for any given fixation (a point on which both eyes converge) will project images onto corresponding points in the retinas (Helmholtz, 1925). That is, any object placed on the horopter will appear in corresponding locations in the left and right retinas. Since an item that falls on corresponding points has, by definition, zero disparity, the horoptor constitutes a surface of zero disparity.

Just as there are two ways of defining corresponding points, there are also two possible ways of defining the horopter. The theoretical horopter (or Vieth-Müller circle) is the circle containing the eyes' fixation points and nodal points[1]. It is calculated by projecting rays from the pairs of geometric corresponding retinal points, and finding the intersections of these rays. The existence of binocular neurons which respond best when the two retinal images converge on corresponding points (Poggio & Talbot, 1981) has been posited as an argument for a neural basis for the horopter (Wolfe *et al*., 2008). Figure 2 shows the geometric corresponding points as they project onto the theoretical horoptor.

An image that lies outside the theoretical horopter will produce one of two types of disparity: An uncrossed (or far) disparity occurs when an object is further from the viewer than the horopter, so that the visual lines (caused by the eyes as they converge) intersect beyond the horopter. A crossed (or near) disparity occurs when an object is between the viewer and the horoptor, the visual lines intersect nearer than the horopter. The images that fall on the theoretical horopter have zero disparity, and are therefore seen in single vision, since they project to corresponding points. However, there is a region immediately around the horopter in which single vision is still possible, although the images do not project to corresponding points, and so horizontal disparity is not equal to zero. This region is known as Panum's fusional area (Panum, 1858). The region in which single vision is possible (i.e. the geometric horoptor, plus Panum's fusional area) is the empirical horoptor (Blakemore, 1970).

---

[1] Nodal points are the points in the eye through which a ray of light can enter (incident nodal point) and emerge from (emergent nodal point) without changing direction (Harris, 2010)

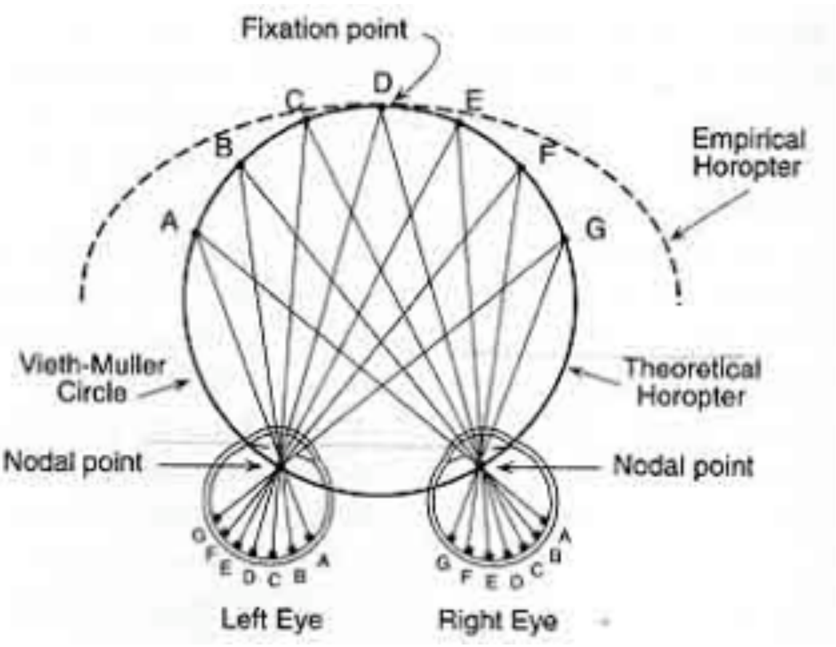


Figure 2: The theoretical horopter (Vieth-Muller circle) is the locus of all the points in space which project an image on to geometric corresponding points in the two retinas. The empirical horopter is less concave than the theoretical horopter, as it includes those disparities which project to non-corresponding points, but for which single vision is still possible. Source: *Jan Wassenaar.*

It should be noted that, while the overwhelming majority of research assumes that the mechanism behind binocular disparity is retinal disparity, recent work suggests that headcentric disparity can also produce a sensation of depth, and may even reverse the depth effects of retinal binocular disparity (Zhang, Cantor, & Schor, 2010). In headcentric binocular disparity, the brain combines the positions of the head and of the retinal images so that binocular disparity is defined in terms of differences between the visual directions of half images relative to the head. This depth system would rely on a different neural coding system than that posited by retinal disparity (Zhang *et al.*; van Ee & Erkelens, 2010). However, in the present paper we assume that retinal disparity alone forms the basis for binocular disparity.

Throughout the first half of the twentieth century, researchers believed that the fusion of the two retinal images needed for stereopsis arose from top-down cognitive processing (Howard & Rogers, 1995). However, the discovery of disparity sensitive neurons in the primary visual cortex (V1) of cats during the 1960s (Barlow, Blakemore, & Pettigrew,1967; Nikara, Bishop, & Pettigrew, 1968; Pettigrew, Nikara, & Bishop, 1968) indicated that the visual inputs are in fact combined very early in processing (Mather, 2006). Poggio (Poggio & Talbot, 1981; Poggio, Gonzalez, & Krause, 1988) claimed that these disparity sensitive neurons could be further divided, according to whether they responded best to uncrossed or to crossed disparities, although these results have since been disputed (e.g. Cumming & Parker, 2000). More than half of the cells in the human V1 are reported to be disparity sensitive, with the proportion increasing in higher visual areas (Poggio & Talbot, 1981). Such large amounts of resources seem necessary when we consider that absolute disparities are often smaller than the width of a single cone photoreceptor (Parker, 2007), and that humans are capable of detecting relative disparities of between 2 to 6 arcsec[2] (Howard, 1919). The neural basis for stereopsis is, however, beyond the scope of this essay; for a review of our current understanding see, for example, Backus, Fleet, Parker, & Heeger (2001); Orban, Janssen, & Vogels (2005); Parker (2007).

One final point to note is the distinction between absolute and relative binocular disparity (Blakemore, 1970). *Absolute disparity* refers to the disparity between projections of a single image point to the left and right retinas, and is therefore used when perceiving, for example, whether a single object is near or far relative to ourselves. *Relative disparity* refers to the difference in absolute disparities between two objects in the visual field, and is therefore used to judge depth relations between more than one object.

1.2. Depth cues and visual illusions

The present study will use (retinal) binocular disparity as a depth cue to elicit a visual illusion, known as the size constancy illusion. Below, we provide an overview of how binocular disparity can be used to create a size constancy illusion. We also note that size constancy can be elicited by depth cues other than binocular disparity, and that there is some debate about the relative contributions of binocular disparity versus other depth cues to our 3-dimensional perception of the world.

1.2.1. Depth cues

Depth perception in everyday life usually arises from the interaction of both monocular and binocular depth cues in a rich visual array (Berryhill & Olsen, 2009). Monocular depth cues are those that may be used by a single eye as well as by both eyes together. Binocular depth cues are only available when both eyes are used. Monocular depth cues include relative size (e.g. Ittelson, 1951); aerial perspective (e.g. O'Shea, Blackburn, & Ono, 1994); height in the visual field (e.g.

---

[2] 1 arcsec = 1/3600 of a degree

Ooi, Wu, & He, 2001); texture gradients (e.g. Bruce, Green, & Georgeson, 2003); image blur (e.g. Marshall, Burbeck, Ariely, Rollard, & Martin, 1996); shadows (e.g. Koenderink, van Doorn, & Kappers, 1996); linear perspective (e.g. Gregory, 1966); occlusion (e.g. Coren & Girgus, 1975); and motion parallax (e.g. Harris, 1994). The visual system can also use non-visual cues, which depend on the musculature of the eye to give information as to where the eye is fixating: this information can be monocular (accommodation), or binocular (vergence). Accommodation provides feedback about the shape the lens has taken in order to maintain focus, which depends on the distance between the viewer and the object in focus (e.g. Mon-Williams & Tresilian, 2000). Vergence provides feedback about the vergence angle of the eyes, which again depends on viewing distance (e.g. Viguier, Clement, & Trotter, 2001). For a summary of these cues and how they work see, for example, Mather (2006) or Wolfe et al., (2009).

There is some debate as to the relative importance binocular disparity compared to these other cues in normal vision. The existence of disparity sensitive cells in V1 (see section 1.1, above) indicates that it is extracted very early on in processing; this, coupled with our high sensitivity to changes in disparity, has lead to a general assumption that binocular disparity is the most important depth cue (Mather, 2006). There several reasons why we should be wary of such a claim. Firstly, it is possible that such high sensitivities to disparity have been detected because of the disproportionately large amount of research into binocular disparity compared with some of the other depth cues, and that were similar resources devoted to exploring less well-studied depth cues, similarly high levels of sensitivity might emerge. Secondly, the disparity sensitive neurons in V1 respond only to absolute disparities (Cumming & Parker, 1999), while relative disparities, which allow for more accurate depth judgements (e.g. Westheimer, 1979) are processed in higher visual areas (Thomas, Cumming, & Parker, 2002). Moreover, Cumming & Parker (1997) showed that disparity selective activity in V1 is not always correlated with depth perception. As Berryhill and Olsen (2009) note, it is as yet unclear to what extent different visual regions contribute to stereopsis, or even whether the same areas process both monocular and binocular depth cues. Therefore the existence of a large number of disparity sensitive neurons in V1 does not preclude other depth cues being equally, or even more, important in depth perception.

There is also a small but substantial body of behavioural research showing that in some cases, pictorial depth cues such as texture gradients can compete with, and even outweigh the depth information provided by binocular disparity (e.g. Stevens & Brookes, 1988; Allison & Howard, 2000). Richards (1977) found that stereopsis without monocular contours was significantly impaired compared with stereopsis with monocular contours; indicating a role for monocular depth cues in the stereopsis mechanism. Other authors have claimed that binocular disparities are too small to be detected and that the primary function of having two eyes is not stereopsis, but rather a more efficient visual system with reduced noise (e.g. Jones & lee, 1982). Harris and Wilcox (2009) note that this *binocular concordance* view may appear reasonable when considering long viewing distances, for which changes in retinal disparity are extremely small: for example, Mather (2006) reports that 90% of total variability in retinal disparity is used within 3 metres, given a fixation distance of 40 cm. This means that at longer distances, the visual system does tend to rely on monocular depth cues rather than binocular disparity (Berryhill & Olsen, 2009; but see Allison, Gillam, & Vecellio, 2009, for evidence of binocular disparity as a depth cue at distances of up to 18 metres). However, as Harris and Wilcox (2009) go on to point out, much of our need for depth perception involves interaction with objects at much shorter distances (< 2 metres), for which binocular disparity is a very efficient depth cue. Several authors have noted that stereopsis is most useful as depth cue when interacting with objects within grasping distance (Arsenault & Ware, 2004; McKee, Levi, & Browne, 1990; Morgan, 2003). For example, Servos and Goodale (1994) fitted participants with goggles to provide either monocular or binocular vision and asked them to grasp different objects; individuals “fitted” with monocular vision spent significantly longer time in contact with the object compared with participants fitted with binocular vision, and participants whose vision was switched from binocular to monocular part-way through the trial. This suggests that, although monocular vision still allowed participants to interact with the objects in front of them, binocular vision provided better information about the size and location of object, allowing for a better grasp (see also Servos, Goodale, & Jakobson, 1992). In addition to studies which emphasise the role of monocular depth cues, or question the assumed dominance of binocular disparity as a depth cue, we must be aware that the binocular visual field - the part of the visual field which is shared by both eyes - is only one part of the total visual field (Howard & Rogers, 1995). Those regions of the visual field which are only available to one eye are now recognised as playing an important role in depth perception (see Harris & Wilcox, 2009, for a review). However, the majority of research suggests that binocular disparity is, if not the dominant depth cue, at least one of the most important.

1.2.2. Size constancy

Size constancy can be defined as the mechanism by which the perceived size of an object remains constant, despite changes in viewing distance, which result in changes to the size of the object’s retinal image (e.g. Holway & Boring, 1941; Sedgwick, 1986). The basis for this mechanism is the absolute size of the retinal image projected by an object (Ross & Plug, 1998).

As the distance between viewer and object increases, the vergence angle of the eye decreases, leading to a decrease in the size of the retinal image. This can be summarised by the equation $R = S/D$, where R is the size of the retinal image, S is the physical size of the object, and D is the distance between object and viewer (Coren & Girgus, 1978). These reductions in retinal image size are progressively offset in the lateral geniculate nuclei[3], by corresponding increases in the sizes of what Ogle (1950), terms the *ocular* images - images that are processed later in the visual system (Bishop, 1994). Figure 3 shows an example of size constancy scaling in a visual illusion.

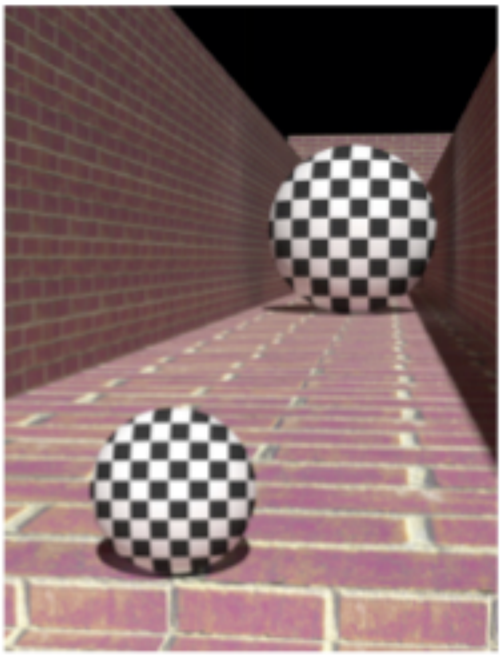


Figure 3: The corridor illusion relies on size constancy scaling. Both balls cast the same size retinal image, but Monocular depth cues mean that the ball in top of the picture is interpreted as being further away. The size constancy mechanism scales up the top ball's ocular image accordingly, and it is perceived to be the larger of the two balls. *Source: Kersten and Murray, 2010.*

Recent work has revealed a potential neural basis for the size constancy mechanism. Murray, Boyaci, and Kersten (2006), used functional magnetic resonance imaging (fMRI) to show that an object which appears more distant, and is scaled up using size constancy - for instance, the ball at in the top half of Figure 3 - activates a larger area of V1 compared with an object which appears closer and smaller; despite the fact that both objects create the same visual angle on the retina.

---

[3] The lateral geniculate nucleus, located in the thalamus, is the primary processing centre for information coming from the retina.

Several well known illusions, such as the Ponzo illusion (Ponzo, 1913; cited in Koehler & Wallach, 1944; see Figure 4) and corridor illusion (Fineman, 1981; see Figure 3) are based on the size constancy mechanism. The classical explanation of such illusions is that misleading depth cues make an object appear distant, so size constancy applies and the object's ocular image is scaled up, thus making it seem larger than another object of equal size (e.g. Coren & Girgus, 1978).

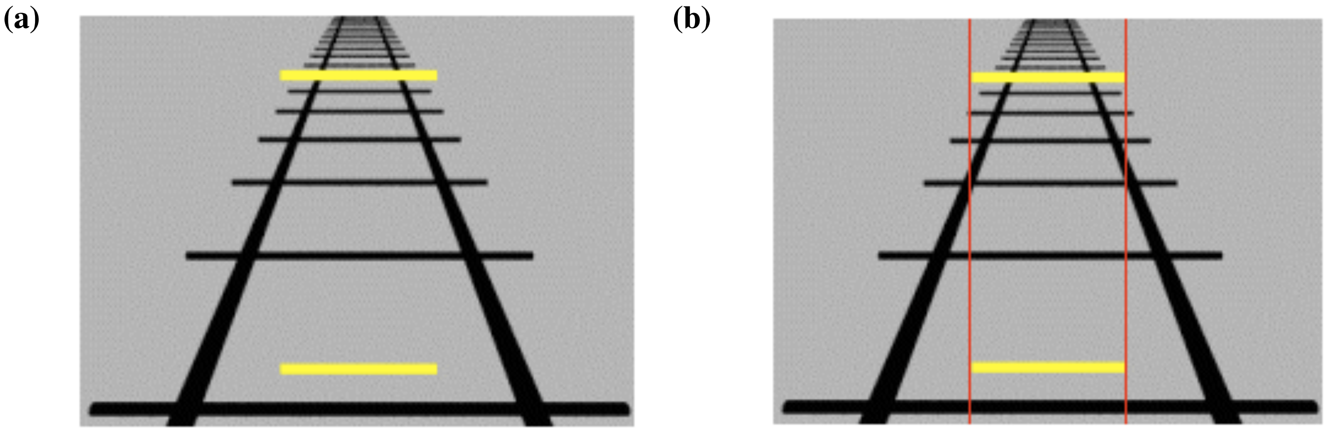


Figure 4: A variant of the Ponzo illusion. In (a), the top yellow line appears longer than the bottom yellow line but, as can be seen in (b), both are actually the same size. *Source: Dr. Tony Phillips.*

Given the debate about the pre-eminence of binocular disparity compared with other depth cues, it is perhaps not surprising that there is also some debate as to the relative importance of monocular and binocular depth cues involved size in constancy. Bishop (1994) argues that size constancy is a monocular phenomenon, and it is true that most visual illusions rely on monocular, pictorial depth cues: for example, the texture gradient in Figure 3, and the converging lines in Figure 4. Coren and Girgus (1978) argue that monocular depth cues allow for such compelling visual illusions because humans tend to interpret even the simplest and most stylised 2-dimensional array as depicting a 3-dimensional scene, so that we interpret the pictorial depth cues in such illusions very much as we would interpret monocular depth cues in the real world (e.g. Hudson, 1960, 1962). Greene and Gretner (2001) argued that, as visual illusions can also be elicited using rich scenes featuring a range of depth cues, it is not possible to attribute illusion mechanisms such as size constancy to a single type of cue; instead, illusions may be elicited by different cues according to the situation. Leibowitz, Shina and Hennessy (1972) proposed two separate mechanisms for size constancy in the real world, the choice of which depended on viewing distance: for distances less than 2 metres, she argued that non-visual depth cues, such as accommodation and convergence, were sufficient to allow depth perception; for distances over 2 metres, visual depth cues were more important.

Julesz (1971) elicited the Ponzo illusion in random dot stereograms, which included no monocular depth cues: the illusion was only visible when viewed with both eyes, by using binocular disparity to create the depth effect resulting in size constancy scaling. This implies that, even if binocular disparity is not always responsible for size constancy, it can at least be used to elicit it. However, Richards (1977) notes that such studies typically allowed free eye movements which, he argues, may also have played a role in creating the perception of depth. According to Richards, in order to fully isolate stereopsis, we can remove the effects of eye movements by presenting the stimuli in flashing stereo pairs. Studies which have used this method have shown no reduction in depth perception when monocular depth cues and contours are also present; however, when the stimuli are presented as a random dot stereogram (i.e. without monocular contours), stereopsis is universally impaired, and in some cased completely removed (Foley & Richards, 1972; Richards, 1971).

1.2.3. Variability in the perception of depth

Vision researchers have long known that there is considerable variation between individuals; and generalisations are made in the knowledge that they represent idealisations and abstractions. For example, in section 1.1., we assumed that the disparity of any given fixated point is zero. However, many people's eyes fail to converge entirely when fixating an object, leading to either a crossed or uncrossed disparity which none the less falls within Panum's fusional area, and therefore does not result in diplopia (see, for example, Collewijn & Erkelens, 1990, for a review). There is considerable variable as to the degree of fixation disparity between individuals, even in natural viewing conditions (e.g. Cornell, MacDougall, Predebon, & Curthoys, 2003). However, given that fusion is still possible in spite of such disparities (e.g. Jaschinksi, Jainta, & Kloke, 2010), we assume that fixation disparities will not effect participants' ability to read the stimuli, nor their ability to perceive depth.

A further point of variation is the timing necessary for stereopsis to occur. Lehmkuhle and Fox, (1980) found that participants could perceive depth in random dot stereograms presented for < 50 msec. Uttal, Davis, and Welke (1994) found that, provided that participants' eyes were pre-converged, such that disparity on the fixation point was approaching zero, depth could be perceived in random dot stereograms with stimulus durations of <1 msec. However, significant differences between individuals are also evident. Tam and Stelmach (1998) asked participants to judge which was the closer of two squares in a random-dot stereogram. Approximately half the participants were able to perform this task at 75% accuracy with display duration of 20 msec; the remaining participants required 1000 msec to perform to a similar level of accuracy. The task in the present study was therefore designed to be self-paced, so that as many participants as possible would be able to perceive the disparity-induced depth effect.

Richards (1970) claimed that up to 2.7% of the population are unable to perceive stereopsis. He also suggested that some people may be “blind” to crossed disparities, while able to perceive uncrossed disparities, and vice versa (Richards, 1970). However such claims are disputed by several authors, who argue that failure to perceive stereopsis in laboratory conditions is likely to be an artefact of the task, rather than an indicator of true stereo blindness (e.g. Patterson & Fox, 1974), and that the posited “one way” anomalies suggested by Richards are almost certainly strategic (Newhouse & Uttal, 1982).

1.3. Height in the visual field

In section 1.2.1, we referred to height in the visual field as a cue which helps us to perceive depth. More specifically, height in the visual field refers to the fact that items which are closer to us are more likely to appear in the lower visual field, while items that are further from us are more likely to appear in the upper visual field (e.g. Allison, *et al.*, 2009). In the present study, any given word will appear either above the fixation cross (in the upper visual field), or below the fixation cross (in the lower visual field). We expect that there may be an interaction between upper and lower visual fields, and the size constancy illusion, for the following reasons.

Firstly, Yang and Purves (2003) argue that the visual system takes account of those disparities which it expects to observe, and that these inform our perception of depth. Therefore, if in everyday vision we experience uncrossed disparities occurring more often in the upper visual field, and crossed disparities occurring more often in the lower visual field, participants may use this information to help them interpret what will, after all, be a visual display devoid of any real context.

Secondly, not only are far stimuli more likely to appear in the upper rather than lower visual field, but processing in the visual system appears to improve when this is the case. In one of the earlier studies to investigate the relationship between depth, and the upper and lower visual fields, Breitmeyer, Julesz, and Kropfl (1975) recorded the speed at which participants were able to detect depth in the upper and lower visual fields, using random dot stereograms. Participants were faster to detect the stereograms when "far" stimuli were located in the upper visual field, and "near" stimuli located in the lower visual field; compared to when "far" stimuli were located in the lower visual field, and "near" stimuli in the upper visual field. We therefore expect that, when the "far" stimulus is located in the upper visual field, this will lead to responses more consistently in line with the size constancy hypothesis, compared with the "far" stimuli located in the lower visual field.

1.4. Semantic size and its interaction with depth

In this paper we will investigate not only the size constancy illusion, but also the potential effect of semantic size on this size constancy illusion. Semantic size refers to the real world size of the object depicted by a given noun (Sereno et al., 2009). In particular, we wanted to investigate whether a concrete noun with "large" semantic size (e.g. *castle*) would strengthen the effect of the size constancy illusion, such that a "large" word in the distant position would be more likely to be judged larger, compared with a "small" or size neutral word in the same position. Below, we outline our reasons for thinking that such an effect might occur.

Firstly, there is evidence to suggest that semantic size can affect lexical processing. Sereno et al. (2009) found that bigger words were responded to faster in a standard lexical decision task, having controlled for frequency, imageability, and word length. They propose that this effect may be related to the concept of markedness (Greenberg, 1966; Jakobson & Halle, 1956), noting that, (in English) bigness is the unmarked form: questions such as *How tall is he?* and *How big is the house?* are much more natural than corresponding questions such as *How short is he?* and *How small is the house?* Alternatively, Sereno et al. speculate that larger items attract more attentional resources. This is based on Fischer's (2001) study into line bisection, which showed that when a line was presented with a digit on either side, participants bisected the line closer to the numerically larger of the two digits; and on Bruner and Goodman's (1947) finding of a size-value effect, where more valuable things were judged to be larger than less valuable things. Bruner and Goodman claim that a larger example of a category activates more neurons, attracts attention more easily, and may hold our attention for longer, which may help explain the results of Sereno et al.'s study.

Secondly, it appears that information about the size of an object is automatically activated by reading the object's name, even when this information is task irrelevant and impedes performance in the task. Rubinstein and Henik (2002) used a Stroop-like paradigm in which participants were asked to judge the relative sizes of animal names, in terms of font size: the results revealed both a facilitation effect - participants were faster in the congruent condition (e.g. elephant, ant) than the neutral conditions (e.g. elephant, ant / elephant, elephant) - and an interference effect - participants were slower in the incongruent condition (e.g. elephant, ant) compared with the neutral conditions. Setti, Caramelli, and Borghi (2009) noted that the activation of the animal's real world sizes could be a product of the task, which did, after all, require participants to make a judgement about relative sizes. They found a priming effect of semantic size, with semantically "large" primes facilitating semantically "small" targets and argued that this, unlike Rubinstein and Henik's results, could not be attributed to a product of the task. Convergent evidence that information about the real-world size of objects is elicited by nouns comes from studies using neuro-imaging techniques: the same regions are active when an object is presented visually as when the objects' names is read (see Martin, 2007, for a review). These results are consistent with findings from the embodied cognition literature, that reading an object's name activates perceptual information about that object, such as its shape or orientation (e.g. Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002); although note that while qualities such as shape are absolute properties of an object, its size is always relative (Setti et al., 2009).

Finally, although we are expecting the size constancy mechanism to operate at a relatively low-level of processing - studies show that it is also evident in baboons (Barbet & Faget, 2002); horses (Timney & Keil,1996; and pigeons (Fujita, Blough, & Blough, 1991, 1993) amongst other animals - several studies suggest that our perception of size can be mediated by higher-order processes. Haber and Levin (2001) found that participants were more accurate in judging the size of familiar objects when those objects tended not to deviate from a prototypical size (e.g. bowling ball), compared with objects which displayed a high degree of token variation in size (e.g. house plant). Haber and Levin attribute these results to participants relying on their memory of object size, rather than online information about the retinal image size and viewing distance of the object. Similarly, people tend to make larger errors in size judgements about familiar objects (e.g. key) compared with novel objects of similar dimensions; suggesting that participants only used information about the retinal size of an object when there was no size information available in memory (Wesp, Peckyno, McCall, &Peters, 2000).

2. The current study

This study aimed to elicit the size constancy illusion in lexical stimuli using binocular disparity as a depth cue: we are not aware of any published study in the literature which has shown this effect. The use of binocular disparity rather than monocular depth cues allowed us present dichoptic stimuli (i.e. to present distinct stimuli to the left and right eyes) using a custom built mirror stereoscope (see Figure 5).

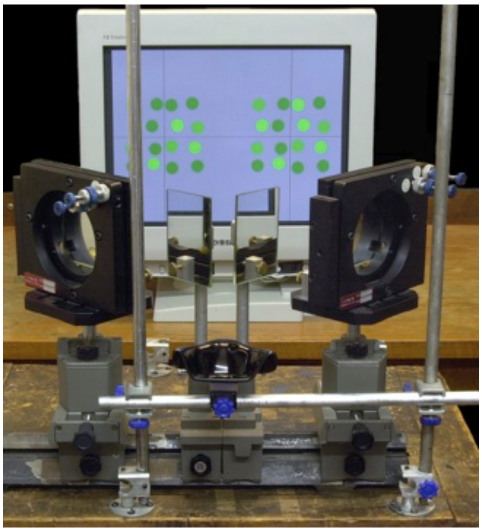


Figure 5: Example of a mirror stereoscope, which allows researchers to present distinct stimuli to the left and right eyes.

We presented the same pair of words to both the left and the right eyes separately; for example each eye would see *castle* above the fixation cross, and *pencil* below the fixation cross. By presenting a distinct stimuli to each eye, we were able to modify the position of each word as it appeared to the left eye, relative to the position of the same word as it appeared to the right eye. In this way, the word pairs were presented with a disparity of one character space which would result in stereopsis. For example, in the right eye's display, *castle* would be shifted one character space towards the left, relative to *pencil;* in the left eye's display, *castle* would be shifted one character space to the right, relative to *pencil*. This would produce a crossed disparity in *castle*, and an uncrossed disparity in *pencil*, such that, when the participant fused the pairs of words together, *pencil* appeared further away than *castle*. We hypothesised that *pencil* would then appear larger than *castle*, due to size constancy scaling (see Figure 6). The fixation crosses for the left and right eyes were presented at zero disparity, and were taken to represent the horopter.

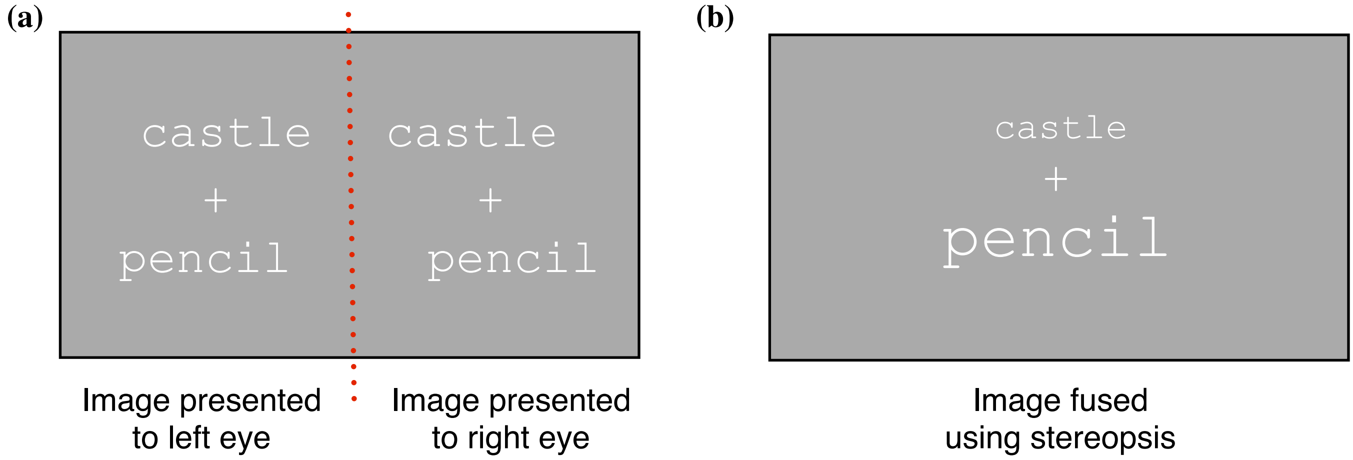


Figure 6: (a) Schematic of example stimuli presented with 1 character space of disparity, as viewed without the stereoscope. (b) Fused image of the same stimuli, as viewed with the stereoscope. The disparity has created a depth effect, via stereopsis, in which *pencil* appears further in the distance. Size constancy scaling is therefore applied to *pencil*, which appears larger, compared with the seemingly closer *castle*.

The level of relative disparity was limited to one character space because this is the maximum that can be fused in normal reading (Blythe, Liversedge, & Findlay, 2010); Kaufman (1964) showed that stereopsis is possible using one character disparity by created a stereogram of random alphabetic letters, in which a section of letters was displaced by one character space. Participants were asked to identify which of the words on their screen appears physically longer in absolute spatial terms. The term *longer* was used, rather than *larger*, because of the presence of “small” and “large” semantic size words. In this way, we hoped to distract participants from noticing that, amongst the semantic size items, one of the words always referred to an object that was physically larger than the object depicted by the second word. However, we expected that, if size constancy scaling was applied to a word, then that word would appear larger *overall*, as though presented in larger font, and not simply longer (i.e. we expected size constancy to produce a phenomenon similar to that in the corridor illusion, in which overall size is perceived to increase, as opposed to the Ponzo illusion, in which only length is perceived to increase.)

Participants' responses ("top" if the top word appeared longer; "bottom" if the bottom word appeared longer) were divided into three data sets: size neutral data, semantic size data, and combined data. This allowed us to check for effects which may only be present in one of the two data sub-sets (size neutral, semantic size), and to make comparisons between the two data sub-sets, and between each data sub-set and the combined data.

2.1. Hypotheses

We present three hypotheses to be tested.

i. Size constancy illusion :
Participants should report the word forming an uncrossed disparity relative to the fixation cross, as longer significantly more often than they report the word forming a crossed disparity relative to the central fixation cross.

ii. Height in the visual field :
Participants should report the size constancy effect significantly more often in the ecologically valid condition (when the "furthest" word is presented above the fixation cross), than when the "furthest" word is presented below the fixation cross.

iii. Semantic size :
Participants should report the size constancy effect significantly more often when the further word is a semantically "large" noun, compared with a semantically "small" noun, or size neutral word.

2.2. Methods

2.2.1. Participants

48 monolingual native English speakers (27 female and 21 male) were recruited using the University of Edinburgh's SAGE advertising service. Mean ages were 22 years and 9 months (*SD* = 4) for the males and 22 years and 7 months (*SD* = 3) for the females; 22 years and 8

months overall (*SD* = 3.5). Participants had normal or corrected-to-normal vision. All participants reported no history of reading difficulties. Age, sex, and handedness were self-reported. Dominant eye was checked using the aperture test (appendix 1). Full participant details are provided in Appendix A.

2.2.2. Materials

Semantic size data

A total of 93 six letter nouns judged to denote large objects, and 102 nouns judged to denote small objects, were selected as potential semantic-size stimuli. Large and small were defined operationally as size relative to an average human, as per Sereno et al. (2009). The semantic-size items were randomized and submitted to a pretest. 16 unpaid participants were sent a questionnaire, of whom 13 responded (7 females, 6 males). All pre-test participants were university graduates now in full-time employment, and were recruited by email. Participants were asked to indicate, for each item, whether it denoted a large object (e.g. *church*), or a small object (e.g. *button*), by writing either “large” or “small” respectively. Participants were instructed to write “unsure” if uncertain about an item: for example, if they were unsure whether the object it denoted was large or small, or if they thought it could refer to either a large or a small object, depending on context. Participants were also asked to indicate any words whose meaning they did not know. Potential answers were therefore “large”, “small”, “unsure” or “don’t know”.

In total, 27 “small” words, and 19 “large” words were rated as “unsure” or “don’t know” by at least one respondent, and were therefore excluded as potential stimuli. Full details of responses to the pre-test are provided in Appendix B. This left a total of 149 potential semantic size items: 75 potential “small” words, and 74 potential “large” words. These words were then grouped according to number of syllables. Our initial intention had been to create pairs of “small” and “large” words by matching both frequency and number of syllables, as per Sereno et al. (2009). However, the range of frequencies meant it was not possible to frequency match each “small” word to a “large” word of the same number of syllables. Pairs were therefore only matched for number of syllables, with each “small” word randomly assigned to a “large” word of equal number of syllables. Because one our hypotheses concerned the potential influence of position in the visual field (top or bottom), these pairs were then checked to ensure against any potential bias from iconic relationships (e.g. Zwaan & Yaxely, 2003): the words *kitten* and *settee,* which has been sorted together, were assigned to separate pairs. This resulted in a total of 74 quasirandom pairs. Due to uneven numbers, the word *rattle* was not assigned to any pair. 60 pairs were then chosen at random from the resulting 74 pairs.

Size neutral data

120 size-neutral words were selected from the BNC corpus. Because these items would act as a control when investigating the effect of semantic size, each size-neutral item was matched in both number of syllables and BNC frequency rating to one of the semantic-size items. Size-neutral items were selected according to the following criteria: they must not denote a concrete noun of any size; they must not be number terms (e.g. *twenty*); they must not be comparatives (e.g. *faster*); and they must not describe an action that could interpreted as "small" or "large" (e.g. *nudges* versus *shoved*). In this way, we hoped to avoid the potential confound of some form of general magnitude system (e.g. Walsh, 2003). These words were submitted to a short pre-test: 5 unpaid post-graduate students were asked to confirm whether any of the words gave an impression of size. No participant reported that any control word gave an impression of size. The size-neutral words were then sorted into pairs according to the semantic size word they had been syllable and frequency matched to.

There were therefore 60 semantic size items, each consisting of word "large" and one "small" item, and 60 size-neutral pairs. A further 8 pairs (4 semantic size pairs, and 4 size neutral pairs) were selected as practise items. All pairs of stimulus words, plus their frequency ratings (using the BNC lemmatised corpus) are provided in Appendix C.

2.2.3. Apparatus

All stimuli were viewed through the department's custom built mirror stereoscope. Stimuli were displayed on a 17 inch natural flat .25 pitch Vision Master Pro 413 IIYAMA monitor, which was viewed through two separate viewing tubes, to allow for the dichoptic presentation of stimuli. Throughout the experiment, these tubes were obscured with cardboard and plastic sheeting, and the viewing holes covered until testing to prevent participants realising that their left and right eyes would be viewing different stimuli. Viewing distance was fixed at 135 cm. All stimuli were shown in grey (RGB: 110, 110, 110) 24 point courier new font against a black background. Screen resolution was set to 1024 x 768. A duplicate desktop computer set up next to the stereoscope, in order for the experimenter to record participants' responses. Figure 7 shows the set up of experimental equipment. The experiment was designed and run using the E-prime experimental software suite (version 1.0).

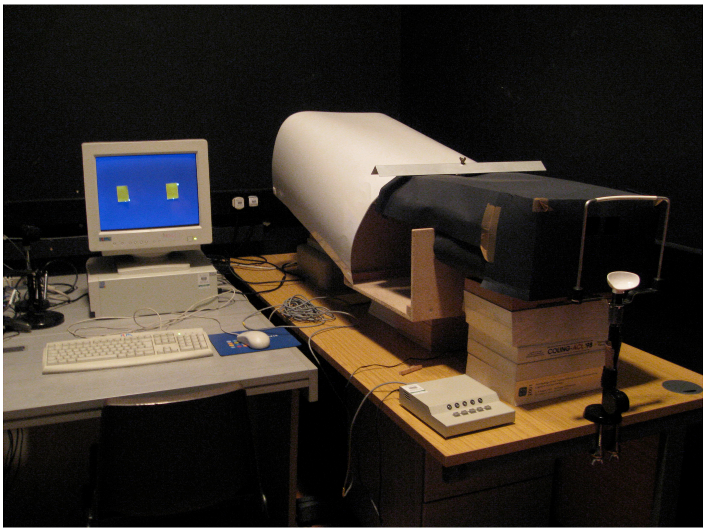


Figure 7: Set up of experimental apparatus. On the left, the mirror stereoscope, with separate viewing tubes obscured, chin rest, and serial response box; and the duplicate desktop computer on the right, on which participants' responses were recorded.

2.2.4. Design

The experiment was a repeated measures, within-subjects design: all participants viewed each word pair in only one condition. Word pairs were kept constant across all conditions and participants. All items were counter balanced across a Latin square design.

The two conditions were top-far condition, and top-near condition. In the top-far condition, the word positioned above the fixation cross appeared further away than the bottom word. In the top-near condition, the word above the fixation cross appeared closer than the bottom word. Within each of these conditions, the word pair could either be a semantic size pair (one "large" word, and one "small" word) or a size-neutral pair. For each pair, Word 1 could either appear above the fixation cross, or below the fixation cross. In the semantic size pairs, this meant that half the trials had a "large" word in the top position and half had a "small" word in the top position. This resulted in a 2 (top-far/ top-near) by 2 (size-neutral/ semantic size) by 2 (word 1 on top/ word 2 on top) design. Figure 7 provides examples of the different word types and positions across conditions, as viewed without a stereoscope. Figure 8 provides examples of the same word types and positions across conditions, as viewed with a stereoscope.

2.2.5. Procedure

Participants' eyesight was pre-screened prior to testing: only participants able to read size 10 courier new font from a distance of 1.35 metres (the same viewing distance used in the experiment proper) were accepted onto the experiment. Sex, age, handedness and dominant eye were recorded but were not counterbalanced across Latin square groups.

Participants were told they were participating in a visual word recognition study; neither semantic size nor stereoscopic viewing were mentioned. Participants were therefore naive as to the purpose of the study. Consent was obtained in accordance with British Psychological Society ethical guidelines. The project was approved by the University of Edinburgh's Psychology Research Ethics Committee.

At the beginning of each session, a prompt appeared on the experimenter's screen, asking participants to place their chin in the chin rest. This was presented as normal font in the centre on the screen, and was intended to give the impression that the participant would be viewing intact, normal text on their own display. The participant then placed her chin on the chin rest and adjusted the chair height as necessary. Participants were instructed not to adjust the chair height once the experiment was underway. Once comfortable, the participant took hold of a 5 button serial response box, placing one thumb over each of the outer buttons. Both hands were used to ensure that both cortical hemispheres were working equally throughout the task. Participants were requested to keep head movements to a minimum. The experimenter then switched off the lights, and removed the screen blocking the participant's view of the stereoscope display.

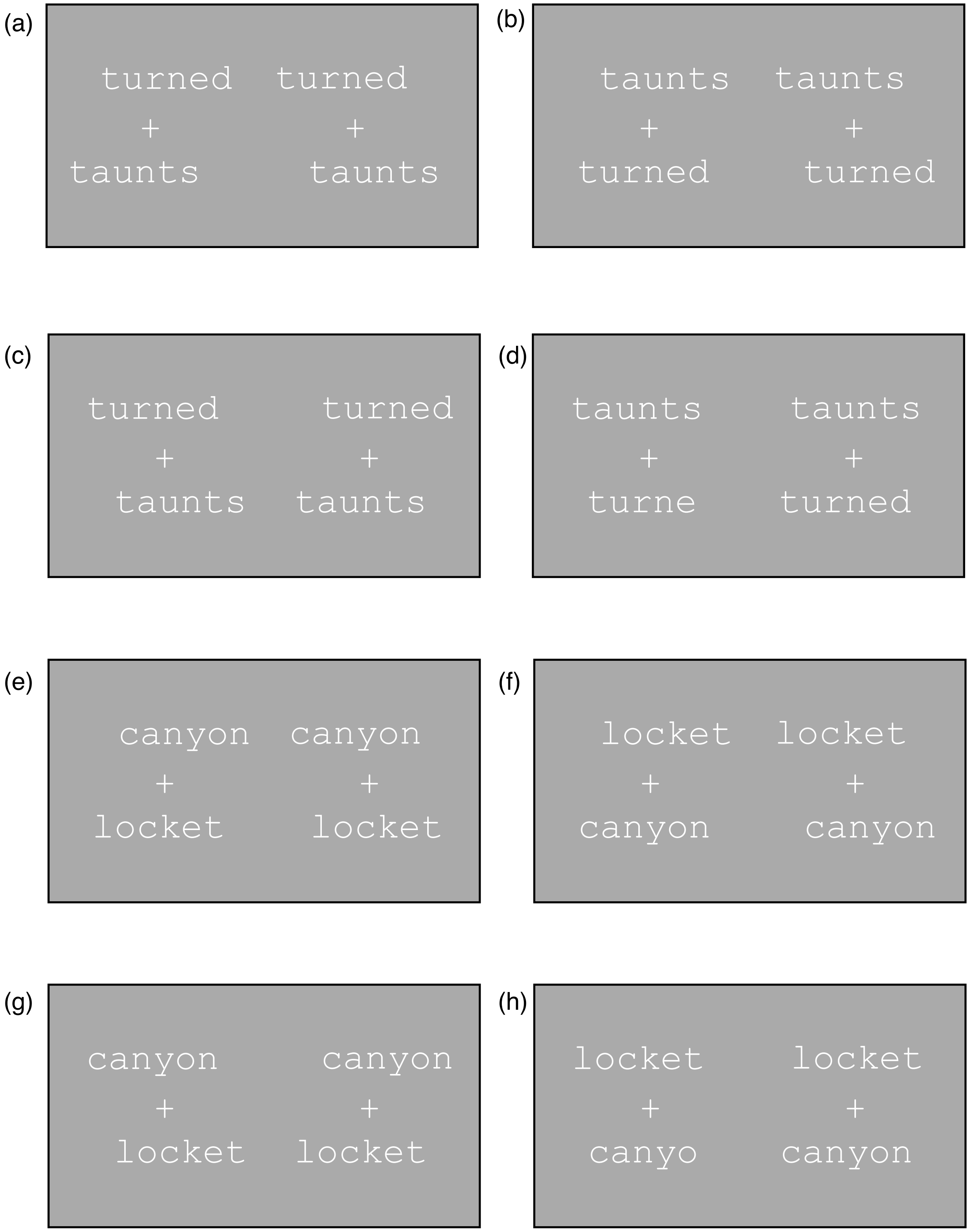


Figure 8: Example stimuli as viewed without the stereoscope. (a) Top-near condition, size-neutral, Word 1 on top. (b) Top-near condition, size-neutral, Word 2 on top. (c) Top-far condition, size-neutral, Word 1 on top. (d) Top-far condition, size-neutral, Word 2 on top. (e) Top-near condition, semantic size, Word 1 on top. (f) Top-near condition, semantic size, Word 2 on top. (g) Top-far condition, semantic size, Word 1 on top. (h) Top-far condition, semantic size, Word 2 on top.

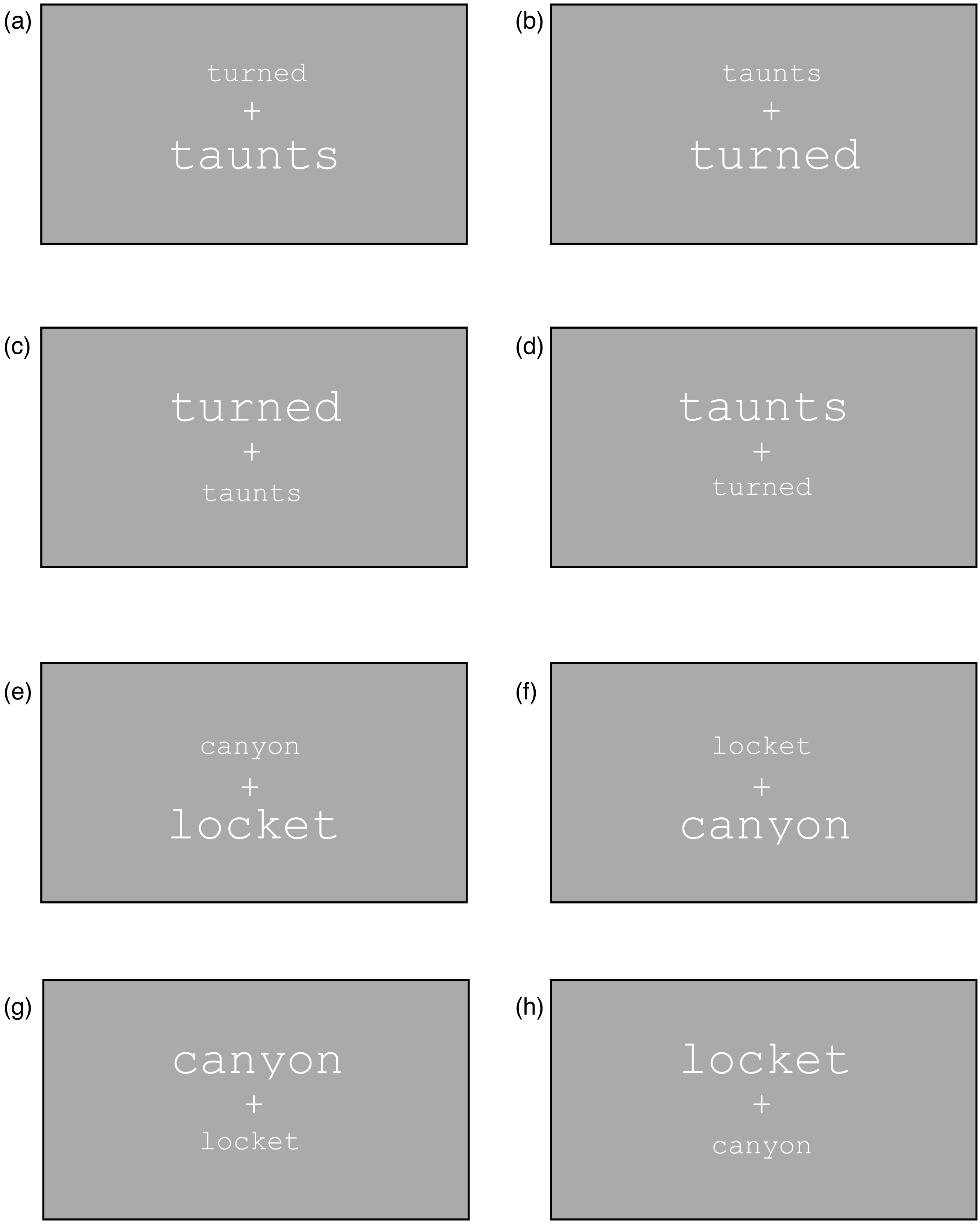
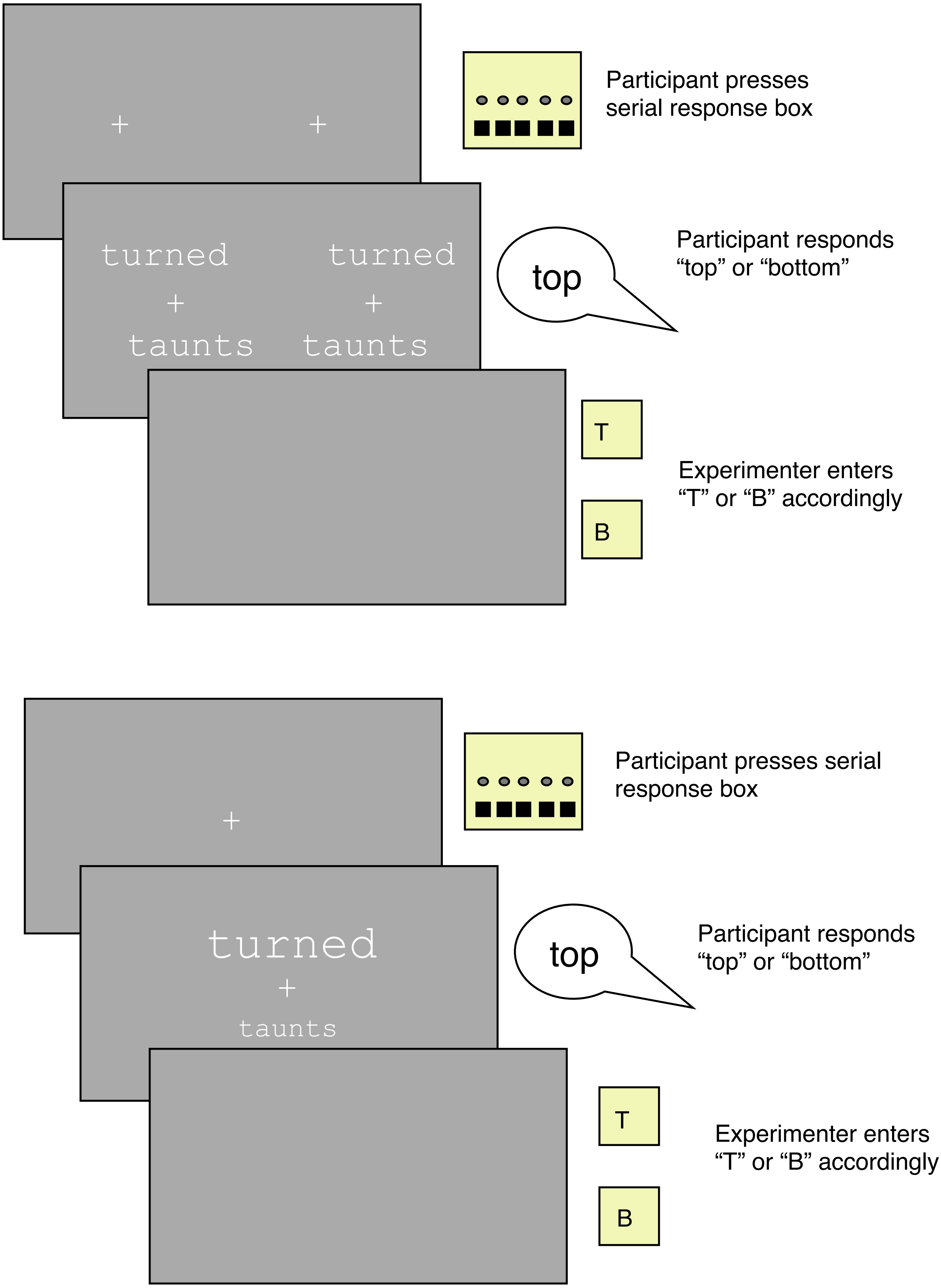


Figure 9: Example stimuli as viewed through the stereoscope. (a) Top-near condition, size-neutral, Word 1 on top. (b) Top-near condition, size-neutral, Word 2 on top. (c) Top-far condition, size-neutral, Word 1 on top. (d) Top-far condition, size-neutral, Word 2 on top. (e) Top-near condition, semantic size, Word 1 on top. (f) Top-near condition, semantic size, Word 2 on top. (g) Top-far condition, semantic size, Word 1 on top. (h) Top-far condition, semantic size, Word 2 on top.

The words “practice session” appeared on the participant’s screen: these were presented as dichoptic stimuli with zero disparity, so that they were fused into a single percept with no impression of depth. The experimenter checked that the participant could read this display with both eyes and with each eye individually. The experimenter also checked whether the stimuli had been fully fused, by checking whether the words appeared on the centre of the screen, or were skewed to the left or the right. When the experimenter was satisfied that the words has been fused, participants pressed both buttons on the serial response box, to begin testing. Each experimental session began with a practice session of eight trials. During this practice session the stimuli were presented with relative disparity of one character space, leading to an impression of depth. The experimenter reconfirmed that the participant could see the words clearly, and that they were not skewed to the left or right of the screen. The practice session was followed by 120 experimental trials. The order of both practice and experimental trials were randomised for each participant.

Each trial proceeded as follows: Two fixation crosses appeared on the experimenter’s screen, centred in the Y axis of the screen. These were fused into a single fixation cross in the centre of the participant’s screen: this fixation cross had zero disparity and therefore fell on the theoretical horoptor. The participant focussed on this cross and, when ready, pressed the two outer buttons on the serial response box. Two words appeared, one above, and one below the fixation cross. The words were presented 35 pixels above or below the fixation cross, with a relative disparity of one character space between the top and bottom words, in both conditions. In order to ensure that the participant read the words correctly, the participant was asked to read both words aloud before refocussing on the fixation cross. It was assumed that participants would read the words from top to bottom, however they were not explicitly instructed to do so. Whilst focussing on the fixation cross, the participant stated which of the words appeared physically longer on the screen, by stating either “top”, if the word above the fixation cross appeared longer, or “bottom” if the word below the fixation cross appeared longer. Participants were instructed to make a quick decision based on their first impression, however all trials were self-paced and reaction times were not recorded. Once the participant gave her answer, the experimenter entered either *t* (if the participant responded “top”) or *b* (if the participant responded “bottom”) into the duplicate desktop computer. The experimenter’s screen was covered throughout all trials, to eliminate experimenter bias. All participants were fully debriefed at the end of the experimental session as to the true nature of the study, in accordance with British Psychological Society ethical guidelines. Each session lasted approximately 30 minutes: 5 minutes instruction and screening, 15 minutes testing, and 10 minutes debriefing. Figure 10 shows a schematic representation of the procedure for each trial.



Participant presses serial response box

Participant responds "top" or "bottom"

Experimenter enters "T" or "B" accordingly

Participant presses serial response box

Participant responds "top" or "bottom"

Experimenter enters "T" or "B" accordingly

Figure 10: Schematic of the procedure for each trial. (a) As viewed without the stereoscope. (b) As viewed with the stereoscope.

3. Results

1 participant was unable to complete the session due to equipment failure; all data associated with this participant were removed from the analysis.

All analyses used Linear Mixed Effects (LME) models, implemented in the lme4 package (Bates, Maechler, & Dai, 2008) in R statistical software (R core development team, 2008). LME models allow the researcher to separate the manipulated independent variables ("fixed" effects) from noise inherent in subject or item selection ("random" effects). Using an LME model therefore allows for the high level of variation between different participants' visual systems (see section 1.2.3), by allowing the intercept to vary between subjects. It also eliminates the need for separate by subject and by item analyses, since the effect of individual items can be modelled as a random effect rather than as an F-ratio (Brysbaert, 2007). The error structure of the data was specified as binomial, and the model fit using the Laplace Approximation. *P*-values for the fixed effects were fitted using Markov Chain Monte Carlo Simulation.

3.1. Debriefing

During debriefing, participants were asked whether they had noted anything special about the way the words were presented. 28 participants reported seeing the depth effect throughout the experiment, and a further 10 participants reported seeing the depth effect in the experimental trials, but not the initial practise session. 9 participants reported seeing no depth effect at all. All participants, including those who did not report any depth effect, reported that the words appeared clear, intact, and centred throughout the session. Preliminary analysis revealed that males were significantly more likely to report no depth than females ($\chi^2 = 260$, p=2.2e-16).

Debriefing revealed that all participants remained naive as to the purpose of the experiment: in particular, no participant noticed that they were viewing dichoptic stimuli, nor that half of the stimuli were sorted into pairs of "big" and "small" words.

3.2. Effect of size constancy

The size neutral data were analysed to see whether the size constancy illusion had been elicited. This data would then form the baseline against which to measure any potential contributions from semantic features in the semantic size data. The dependent variable was whether or not a participant responded “top”. This was a binary variable, with “yes” or “no” as possible outcomes. Responding “top” in the top-far condition was in line with the size constancy hypothesis. Responding “top” in the top-near condition went against the size constancy hypothesis.

The random effects for the LME model were specified as subject. Word 1 and Word 2 were initially specified as random effects, however the variance associated with these random effects was zero, and they were removed from the model. The fixed effects that were tested were: condition (top-near or top-far), whether an individual reported depth perception (yes or no), age (continuous variable), dominant eye (left or right), dominant hand (left or right), sex (male or female), Latin square group (1, 2, 3, or 4), and trial number (continuous variable). Age (p=0.89), dominant eye (p=0.77), dominant hand (p=0.69), Latin square group (p=0.64), log frequencies for Word 1 (p=0.42) and Word 2 (p=0.83), and trial number (p=0.25) were non-significant, and were removed from the model. The fixed effects in the final, fitted model were a three-way interaction between condition, reported depth perception, and sex.

Overall, there was a significant effect of condition: participants were significantly more likely to respond “top” in the top-far condition, compared with the top-near condition (p=2e-16). Males responded “top” more often than females in both the top-far and top-near conditions; males were therefore more likely than females to perform in line with the size constancy hypothesis in the top-far condition, but less likely in the top-near condition. However, this trend for an interaction between sex and condition was not significant (p=0.29). There was a significant effect of reported depth: overall, participants who reported no depth were significantly less likely to respond in line with the size constancy illusion in both the top-far condition (p<0.00) and top-near condition (p=1.192-05). The interaction between these three variables (condition, sex, and reported depth) was significant, and significantly improved the fit of the model ($\chi^2 = 23, p < 0$).

Table 1: Summary of model goodness of fit

Fixed effects	AIC	BIC	deviance	p
Condition	3249	3267	3243	
Condition*Sex	3248	3277	3238	< 0.1
Condition*Reported.depth	3233	3263	3233	< 0
Condition*Reported.Depth*Sex	3218	3272	3200	< 0

In the top-far condition, males who reported no depth were significantly less likely to answer in line with the size-constancy hypothesis compared with males who reported depth (*p*<0.00); females who reported no depth were also less likely to answer in line with the size constancy hypothesis compared with females who reported depth, but not significantly so (*p*=0.87). In the top-near condition, men who reported no depth were significantly less likely to answer in line with the size constancy hypothesis (*p*=4.71e-05). Females who reported no depth were less likely to answer in line with the size constancy hypothesis, but this was not significant (*p*=0.43).

Table 2: Probability of responding "top" across conditions, sex, and reported depth.

	Top-far condition	Top-near condition
Males		
Reported depth	81%	26%
Reported no depth	58%	36%
Females		
Reported depth	71%	17%
Reported no depth	70%	23%

3.3. Effect of height in the visual field

Next, we checked whether the effect of the size constancy illusion was stronger in the top-far condition, compared to the top-near condition. The dependent variable was whether or not participant's response (Word 1 or Word 2) matched the response predicted by the size constancy illusion (Word 1 or Word 2). The response predicted by the size constancy illusion varied according to condition, and which word was on top. In the top-far condition, when Word 1 was on top (and therefore far), the predicted response was Word 1; when Word 2 was on top, the predicted response was Word 2. In the top-near condition, these responses were reversed: when Word 1 was on top (and therefore near), the predicted response was Word 2; when Word 2 was on top, the predicted response was Word 1. The random effect for this model was subject (Word 1 and Word 2 were associated with zero variance as random effects, and removed from the model). Initially, we modelled the data with condition (top-far or top-near) as the sole fixed effect. This was insignificant in all three data sets: size-neutral (*p*=0.27); semantic size (*p*=0.88), and combined (p=0.5). However, further analysis showed a significant interaction between condition and sex. The effect of size constancy in males was significantly stronger in the top-far condition than the top-near condition in all three data sets: size-neutral (p<0.00), semantic size (p=0.01), and overall (p=1.35e-05). The effect of size constancy in females was significantly *weaker* in the top-far condition in both the size neutral (p<0.00) and overall (p<0.00) data sets; it was also weaker in the top-far condition in the semantic size data set, but this trend did not reach significance (p=0.16). Adding this interaction with sex significantly improved the fit of the model, in all three data sets: size neutral (χ^2 =13; p< 0.00); semantic size (χ^2= 7; p=0.3); and combined (χ^2=19; p=7.761e-05).

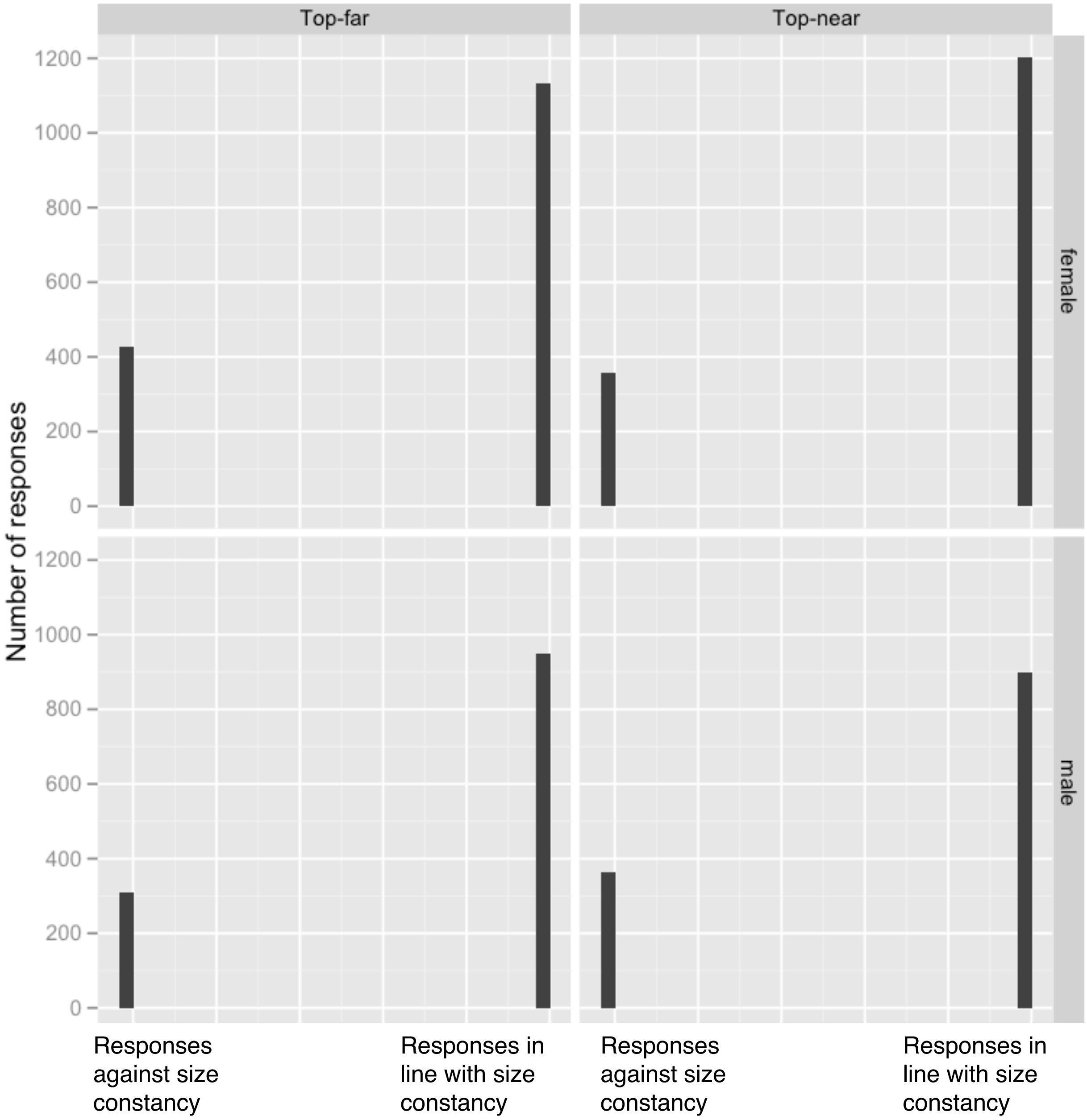


Figure 11: Number of responses in line with and against size constancy, by condition and sex

3.4. Effect of semantic size

Next we tested for an effect of semantic size on participants' response. The dependent variable was whether a participant responded "top". Subject was fitted as a random effect; Word 1 and Word 2 were associated with zero variance and removed from the model. Fixed effects were a three way interaction between condition, reported depth, and sex; and semantic size (big, small, or no size) for the top and bottom word. Semantic size was not significant for either top word (p=0.73) or bottom word (p=0.58). To check for a more general effect of semantic size (differences perhaps due to being a concrete noun), we therefore ran an additional model using item (size-neutral or semantic size) as a fixed effect along side the condition-reported depth-sex interaction. There was no significant difference between size-neutral and semantic size words(p=0.43).

Finally, we used a mixed effects model to check for any effect of semantics which may have existed within the semantic size data, whilst not reaching significance in the overall data set. Subject was fitted as a random effect; fixed effects were an interaction between condition, sex, and reported depth; as well as the following semantic variables: Semantic size, log frequency, imageability ratings, concreteness ratings, and semantic categories. We also checked for an effect of log frequency which may not have been evident in the size-neutral data. There was no significant effect semantic size; either for top (p=0.89) and bottom word (p=0.89). There was no significant effect of imageability for either top (p=0.42) or bottom word (p=0.69). There was no significant effect of concreteness for both top (p=0.69), or bottom word (p=0.16). There was no significant effect of log frequency for either top (p=0.07) or bottom word (p=0.93). Planned comparisons showed that there was no significant effect of semantic category: animal (top word: p=0.81; bottom word: p=0.59); artefact (top word: p=0.74; bottom word: p=0.63); plant (top word: p=0.46; bottom word: p=0.71); food (top word: p=0.48; bottom word: p=0.39); building (top word: p=0.65; bottom word: p=0.36); geographical (top word: p=0.85; bottom word: p=0.67). In addition, age (p=0.56), dominant eye (p=0.6), dominant hand (p=0.79), Latin square group (p=0.53) and trial (p=0.16) were insignificant.

3.5. Effect of size constancy, revisted.

The above analyses showed there was no significant difference between the size-neutral and semantic size items. The initial size constancy models from section 3.2 were therefore fitted to both the semantic size data, and the combined data set, to quantify the size constancy effect across all data.

3.5.1. Semantic size data

In the semantic size data, there were overall significantly more “top” responses in the top-far condition than the top-near condition (p<2e-16). Males responded “top” more often than females in both top-far condition and top-near condition, although neither of these trends was significant (p=0.39 and p=0.6, respectively). There was a highly significant effect of reported depth: participants who reported depth were significantly more likely to answer “top” in the top-far condition, compared with those who did not report depth (p=0.01); and significantly less likely to answer “top” in the top-near condition than those who reported no depth (p<0.00). The three way interaction of these terms was also significant, and significantly improved the fit of the model (p<0.00). In the top-far condition, males who reported no depth were significantly less likely to respond “top” compared with males who reported depth (p<0.02); females who reported no depth were more likely to respond “top” compared with females who reported depth, but not significantly so (p=0.54). In the top-near condition, males who reported no depth were significantly more likely to respond “top” than males who reported depth (p=1.05e-06). Females who reported no depth were more likely to respond “top” than females who reported depth, but this was not significant (p=0.56).

3.5.2. Combined data

In the combined data, there were overall significantly more “top” responses in the top-far condition than the top-near condition (p<2e-16). Males responded “top” more often than females in both top-far condition and top-near condition, although neither of these trends was significant (p=0.22 and p=0.29, respectively). There was a highly significant effect of reported depth: participants who reported depth were significantly more likely to answer “top” in the top-far condition, compared with those who did not report depth (p<0.00); and significantly less likely to answer “top” in the top-near condition than those who reported no depth (p=9.14e-09). The three way interaction of these terms was also significant, and significantly improved the fit of the model (p=7.316e-12). In the top-far condition, males who reported no depth were significantly less likely to respond “top” compared with males who reported depth (p<0.02); females who reported no depth were more likely to respond “top” compared with females who reported depth, but not significantly so (p=0.54). In the top-near condition, males who reported no depth were significantly more likely to respond "top" than males who reported depth (*p*=1.05e-06). Females who reported no depth were more likely to respond "top" than females who reported depth, but this was not significant (*p*=0.56).

Table 3: Probability of responding "top" across conditions, sex, and reported depth, in all data

	Top-far condition	**Top-near condition**
Size neutral data		
Males		
Reported depth	81% ***	26% ***
Reported no depth	58% ***	36% ***
Females		
Reported depth	71%	17% ***
Reported no depth	70%	23%
Semantic size data		
Males		
Reported depth	81% *	24%
Reported no depth	65%	34% **
Females		
Reported depth	75%	23%
Reported no depth	69%	18%
Combined data		
Males		
Reported depth	82% **	26% *
Reported no depth	62% *	35% ***
Females		
Reported depth	73%	23% ***
Reported no depth	70%	17%

Table note. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4. Discussion

4.1. Effect of size constancy

Analysis of all three data sets (size-neutral, semantic size and combined) revealed a strong effect of the size constancy illusion, mediated by sex and reported depth. This is in spite of Richards' (1977) claim that monocular contours are a necessary part of the stereopsis mechanism. As neither pictorial depth cues nor binocular convergence were available to participants during the task, and we therefore conclude that the size constancy illusion in this study was elicited as a consequence of binocular disparity. As far as we are aware, this is the first time that binocular disparity has been used to elicit the size constancy illusion in lexical stimuli.

The effect of reported depth on size constancy is of particular interest, for two reasons. Firstly, analysis revealed that, overall participants who reported no depth were significantly more likely to respond "top" in the top-far condition than the top-near condition; that is, they exhibited a strong effect of size constancy, regardless of the fact that they did not report noticing the depth effect which triggered the size constancy mechanism. Ward, Porac, Coren, & Girgus (1977) also found that participants did not report noticing any depth for a range of visual illusions which are often said to rely on size constancy scaling, including the Ponzo illusion; however the participants still perceived the upper line in the Ponzo illusion as longer. This is sometimes taken as evidence against size constancy explanations of such illusions (e.g. Coren & Girgus, 1978). However, in the present study the only difference between the top-near and top-far conditions was whether the top or bottom word was presented with greater relative disparity; therefore, the only obvious explanation of why participants should respond "top" significantly more often in the top-far compared with top-near condition, is that in the top-far condition, it is the top word which forms an uncrossed disparity (beyond the horopter), and so appears further away. Purghe and Coren (1992) note that depth cues can operate at several levels, leading to a distinction between registered depth (depth that is registered as stimuli are processed, and of which an individual need not be conscious); and phenomenal, or perceived depth (the individual's subjective, conscious experience of depth). Several studies have shown that, particularly in the case of visual illusions, registered and perceived depth are dissociable (e.g. Ward et al. 1977; Gillam, 1980; see Coren, 1990 for a more complete treatment of the subject). We therefore conclude that participants who did not report observing a depth effect nonetheless registered the depth effect during encoding, and that this formed the basis for the size constancy illusion.

The second point of interest about those who reported no depth is that, although they did perform in line with size constancy as outlined above, they nonetheless did so *less* than those who reported depth. Moreover this trend, although in the same direction for both sexes, was significant for males but not for females. The lack of an interaction between condition and sex when reported depth is not taken into account, indicates that this is not simply the result of males versus females, where the males happened to be less likely to report depth. Rather, it appears that whether an individual *perceives* depth, as opposed to simply *registering* it, impacts on the degree to which they experience the size constancy illusion so that, while registered and perceived depth may indeed be dissociable, the two together provide a stronger basis for the size constancy illusion than registered depth alone. This may have implications for the extent to which visual illusions and perceptual constancies are dependent on attention (e.g Fang, Boyaci, Kersten, & Murray, 2008; Murray, & He, 2006).

The fact that the interaction between reported depth and condition was significant in males but not females may perhaps be a result of terms of differences in hemispheric processing. There is evidence that males are more strongly lateralized than females in a variety of tasks (see McGlone, 1980, for a review) including face processing (e.g. Godard & Fiori, 2010); language processing (e.g. Kansaku, Yamura, & Kitazawa, 2000), mental rotation (e.g. Johnson, McKenzie, & Hamm, 2002), and visual illusions (Rasmjou, Hausmann, & Gunturkun, 1999). These results seem to bear out earlier suggestions that certain neuropsychological mechanisms involved in verbal and spatial awareness might be located in contralateral hemispheres in males, but the same hemisphere in females (e.g. Lansdell, 1962). Of particular interest to this study is the proposal that susceptibility to visual illusions may be more strongly lateralized in males than in females (Rasmjou et al. 2009). It is argued that the right hemisphere is more susceptible to visual illusions than is the left hemisphere (e.g. Clem & Pollack, 1975; Houlard, Fraisse, & Hecaen, 1976; but see, for example, Bertelson & Morais, 1983 for conflicting findings); therefore, the higher degree of lateralization in males leaves them more susceptible to visual illusions, thus potentially explaining the significant interaction we found in males, but not females.

4.2. Effect of height in the visual field

We hypothesised that the effect of the size constancy illusion would be stronger in the top-far condition than in the top-near condition, since the top-far condition represented the pattern of disparities which we are more likely to encounter in every day life (further things tend to be higher up). However, our results showed that this was only the case for male subjects; for female subjects, the strength of the illusion was significantly higher in the top-near condition. These results were quite striking: even given the differences in lateralization between males and females, which may have explained significance in males versus a null result in females, this does not seem to explain why females should display a significant result in the *opposite* direction.

One possible, but highly speculative, explanation concerns the distinction between the ventral and dorsal streams in processing. Previc (1990) argues that the upper visual field is processed by the ventral stream, while the lower visual field is processed by the dorsal stream. The dorsal stream is associated with, for example, grasping and manipulating objects which would tend to be within arms' reach (Goodale, Milner, Jakobson, & Carey, 1991), which may explain why dorsal areas display preferential activation for near versus far stimuli (e.g. Quinlan & Culham, 2007). We have already seen (section 1.2.1) several authors argue that binocular disparity is more effective as a depth cue for objects within arms' reach (e.g. Arsenault & Ware, 2004). Furthermore, Greene and Gentner (2001) suggest that binocular disparity is primarily processed by the dorsal stream. Therefore, although we may expect more uncrossed disparities to occur in the upper visual field, binocular disparity may have provided more precise online depth information in the lower visual field. Under this explanation then, the males paid more attention to prior knowledge about expected disaprities; whilst females paid more attention to the actual online information provided by the disparities, leading to improved performance in the top-near condition relative to the males.

4.3. Effect of semantic size

Analysis of the three data sets (size-neutral, semantic size, and combined) shows no significant differences between words with a semantic size, and those without. This suggests that knowledge of the real-world size of an object does not interact with the size constancy illusion, despite studies which in which such knowledge interferes with size judgements about objects (e.g. Haber & Levin, 2001), and about words which refer to those objects (e.g. Rubinstein & Henik, 2002). This null result is emphasised by the fact that analysis on the semantic size data revealed no effect of any of the semantic variables (semantic size, imageability, concreteness, semantic category), nor of BNC frequency rating. The absence of a frequency effect, in particular, indicates that the size constancy illusion is operating at a very low level of processing, given that word frequency is well-known to affect tasks which involve higher level processing (see Monsell,1991, for a review).

Fischmeister & Bauer (2006) note that monocular depth cues cues help us to understand perceptual grouping. The stimuli in this experiment were deliberately presented quite close to the fixation cross (35 pixels). This was done in order that both words and fixation crosses should within at least parafoveal vision, and allow participants to compare the size of both words whilst fixating on the cross. It may that placing the stimuli so close together, coupled with the absence of monocular depth cues, resulted in participants engaging in some form of perceptual grouping in which the top and bottom words were not sufficiently distinct for their contrasting semantics to come into play. However, such a post-hoc explanation remains highly speculative.

Three further explanations for the absence of a semantic size effect are possible. Firstly, we note that no semantic size pair contained two words of the same semantic category (building, animal, plant, artefact, food, geographical landmark). This is not surprising given that foods, for example, are unlikely to be larger than an average human; similarly, excluding atypical examples such as toy castles, we are unlikely to encounter buildings which meet our operational definition of “small” (i.e. smaller than an average human). Shoben and Wilson (1998), investigated the effect of role of categorization when making judgements about relative size using lexical stimuli. They proposed that people use semantic category as a context in which relative size can be judged: the superordinate category (e.g. buildings) is activated, and divided into two subcategories containing the large and small exemplars of that category (i.e. large buildings, and small buildings). Since our word pairs always referred to objects of different semantic categories, it may be that participants lacked an appropriate context in which knowledge of the object’s real world size could interact with the task processing. Both Setti et al. (2009) and Rubinstein & Henrik (2002) used stimuli from the same semantic category (animals). We might therefore seek to replicate their results using stimuli from different semantic categories. A null effect in such studies may imply that shared semantic category is indeed necessary for an interaction between semantic size and judgements about the relative sizes of words; a non-null result may imply that semantic size can affect judgements about the relative size of lexical stimuli, regardless of semantic category, but that this effect is restricted to higher-level processing rather than the low-level processing involved in visual illusions.

Secondly, we noted in section 1.2.1, above, that binocular disparity appears most effective as a depth cue at distances in which we would normally be able to grasp objects in the real world (e.g. Arsenault & Ware, 2004). This may imply of one two things (or both): binocular disparity is a more efficient depth cue at shorter viewing distances; or binocular disparity is a more efficient depth cue for objects we expect to be able to see at shorter viewing distances (i.e. smaller objects). Berryhill & Olsen (2009) note that a considerably greater viewing distance is required in order to perceive an entire house, compared with the viewing distance required to perceive an entire human face. This leads to the possibility that binocular disparity is a more effective depth cue for "small" words, which refer to objects that can be viewed at closer distance; than for "big" words. Therefore, it is possible that there *was* an effect of semantic size, in which "big" words were more likely to be judged as appearing longer due to their semantics, but that this was offset by an increased efficiency of the size constancy illusion in "small" words; such that the two effects cancelled one another out.

A third and final possibility, is that despite the pretest, the stimuli may simply have been inadequate to activate a sufficiently strong knowledge of semantic size in participants. The restrictions of word length (only six letter words were considered as stimuli) and syllable numbers meant that many of the words had a BNC frequency rating of below 100; and that several words (e.g. *armada*, *convoy*) referred not to a single object, but rather a group of objects together. Moreover, although we attempted to restrict the stimuli to unambiguous words, the limits of the English language, coupled with the word length and syllables constraints, meant that many of the stimuli we used could function as verbs as well as concrete nouns (e.g. *garden*, *button*, *bubble*...). Although the BNC frequency ratings of such items as verbs were, in all cases considerably lower than their ratings as nouns, it is perfectly possible that, devoid of context, and interspersed with size-neutral items - many of which were verbs - the size neutral words may have acted as primes such that participants also interpreted at least some of the noun/verb ambiguous semantic size items as verbs. Such an outcome may be predicted by strong claims about grammatical class as a fundamental organizational tool for language in the brain (e.g. Hillis & Caramazza, 1991). However, claims for such a strong claim have been widely disputed in the literature (e.g. Martin & Chao, 2001; Damasio, Tranel, Grabowsli, Adolphs, & Damasio, 2004). Vigliocco, Vinson, Arciuli, and Barber (2008) found a priming effect of grammatical class in lexical decision task only when it the primes were presented in minimal phrasal sort of context (e.g ***to** forbid*, rather than simply *forbid*). Although the task in this case was quite different (relative size judgement versus lexical decision task) such results cast at least some doubt on the contention that the size-neutral verbs primed participants to interpret semantic size items as verbs, thus failing to activate knowledge of the noun's semantic size. Moreover, any attempt to interpret the null effect of semantic variables as an artefact of the stimuli is at least partly compromised by the association of zero variance with Word 1 and Word 2 when entered into the model as random effects.

4.4 Further research

The above discussion has highlighted several areas in which further research would be beneficial, in order to clarify the results obtained in the current study. We end by suggesting some further ways in which this research paradigm might be expanded.

We have argued (section 3.4) that the null effect of semantic variables found in this study cannot easily be attributed to our choice of stimuli. As such, we expect these findings to generalise over other potential stimuli in English, and indeed other alphabetic languages. On the other hand, we feel it is possible that an effect of semantics may be evident in languages such as Chinese and Japanese, which use logographic rather than alphabetic script. Many Chinese characters developed more or less directly from pictograms, and as such retain an element of the object's physical appearance; this visual iconicity may provide for an increased access to, or activation of, the word's semantic size.

Moreover, imaging studies suggest that Chinese readers use display increased right hemisphere activation in reading, compared to readers of alphabetic script (e.g. Tan, Liu, Perfetti, Spinks, Fox, & Gao, 2001). We speculate that this may result in an increased possibility of semantic size interacting with the size constancy illusion (also thought to be right-hemisphere dominant; see section 3.2), especially in males. Note, however, that this assumes that we are correct in explaining the stronger effect of size constancy in males than in females through higher levels of cortical lateralization in males, who therefore have an higher reliance on right hemispheric processing during the task.

The use of Chinese and Japanese text would also allow us to investigate vertical disparities. Given the debate about the role of vertical disparities in stereopsis, a reasonable next step would be to attempt to replicate these research findings using vertical disparities in script that is read from top to bottom, and compare this with the size constancy illusion created by horizontal disparities. This would require a different experimental paradigm than use of a mirrored stereoscope, but the results could potentially shed light on the extend to which vertical disparities can be used in stereopsis (although not necessary on the extent to which they are used in normal viewing).

The differences between males and females, and those who did and did not report depth suggest that future studies may find it beneficial to control for sex and perceived depth across conditions or Latin Square groups. While this is easily done in the case of sex, it is less easy to see how this may be achieved for perceived depth. Presumably it would involve submitting participants to some sort of pretest or pilot study involving stereoscopic depth perception, but this poses potential problems: we would not want participants to engage in the same task as the actual experiment; and yet if different tasks were used, it would be difficult to guarantee that a participant who did not report depth in the first task would not report it in the second. However, if these obstacles could be overcome, comparative studies of those who did and did not report depth in the same task might yield interesting results, particularly, if EEG and imaging technologies could be used to determine activation levels and locations.

Appendix A

Table 4: Breakdown of participant details

Subject	*Age*	*Sex*	*Dominant eye*	*Dominant hand*
1	22	male	right	right
2	24	female	right	right
3	23	female	right	right
4	22	female	left	right
5	19	female	right	right
6	28	female	left	right
7	22	male	left	left
8	22	female	left	right
9	21	female	right	right
10	22	male	right	right
11	19	female	right	right
12	22	female	right	right
13	23	male	left	right
14	22	female	right	right
15	23	male	right	right
16	21	female	right	right
17	36	male	right	right
18	22	female	right	right
19	23	female	right	right
20	21	female	left	right
21	21	female	left	left
22	22	male	right	right
23	23	female	left	right
24	21	female	left	right

25	19	male	right	right
26	21	male	left	left
27	24	male	right	right
28	21	female	left	right
29	19	female	right	right
30	25	female	right	right
31	28	female	left	right
32	32	female	left	right
33	19	female	right	right
34	23	male	right	right
35	24	female	right	right
36	22	female	right	right
37	24	male	right	right
38	21	male	right	right
39	25	female	right	right
40	22	male	right	right
41	23	male	right	right
42	32	male	left	right
43	21	male	right	right
44	20	male	right	left
45	19	male	right	right
46	24	male	right	right
47	18	male	left	right

Appendix B

Table 5: Responses to pre-test questionnaire

Item	Response (N=13)				
	Large	Small	Unsure	Unknown word	TOTAL
Armpit		12	1		13
Arbour	5		1	7	13
Almond		13			13
Arcade	13				13
Armada	13				13
Armory	8			5	13
Azelea		3		10	13
Asylum	12		1		13
Avenue	13				13
Bakery	13				13
Banana		13			13
Bangle		13			13
Bauble		12		1	13
Basket		10	3		13
Bazaar	13				13
Beaker		13			13
Bedsit	13				13
Beetle		13			13
Belfry	13				13
Bikini		12	1		13
Bobbin		12		1	13
Bottle		13			13

Bridge	13				13
Brooch		13			13
Bubble		13			13
Bucket		12	1		13
Bureau	10		2	1	13
Bullet		13			13
Button		13			13
Camera		13			13
Campus	13				13
Canary		13			13
Candle		13			13
Canopy	13				13
Canyon	13				13
Carafe		12		1	13
Carrot		13			13
Cashew		13			13
Casino	13				13
Catkin		8		5	13
Castle	13				13
Cavern	13				13
Celery		13			13
Cellar	13				13
Chalet	13				13
Chapel	13				13
Cherry		13			13
Chisel		12	1		13
Cicada		12		1	13
Church	13				13
Cinder		9		4	13
Cinema	13				13

Circus	13				13
Clinic	13				13
Closet	13				13
Clover		13			13
Convoy	13				13
Conker		13			13
County	13				13
Cygnet		12		1	13
Crayon		13			13
Coupon		13			13
Crocus		13			13
Damson		10		3	13
Dagger		13			13
Dahlia		13			13
Digger	13				13
Dingle	2			11	13
Dinghy	13				13
Domino		13			13
Dragon	13				13
Earwig		13			13
Embryo		13			13
Enzyme		13			13
Estate	12		1		13
Eraser		13			13
Faucet		13			13
Flower		13			13
Forest	13				13
Friary	13				13
Galaxy	13				13
Garden	13				13

Gazebo	13				13
Ghetto	13				13
Glider	13				13
Goblet		13			13
Grotto	13				13
Grouse		13			13
Guitar		13			13
Hearse	13				13
Helmet		13			13
Hornet		13			13
Hostel	13				13
Insect		13			13
Island	13				13
Jungle	13				13
Kernel		11	1	1	13
Kettle		13			13
Kitten		13			13
Larder	12		1		13
Ladder	10		3		13
Lagoon	13				13
Lentil		13			13
Limpet		13			13
Locust		12		1	13
Locket		13			13
Maggot		13			13
Market	13				13
Magpie		13			13
Meteor	12		1		13
Marble		12	1		13
Marina	13				13

Meadow	13				13
Minnow		13			13
Mitten		13			13
Monkey		13			13
Morgue	13				13
Mosque	13				13
Muffin		13			13
Museum	13				13
Mussel		13			13
Nebula	12		1		13
Nettle		12	1		13
Needle		13			13
Nugget		13			13
Nutmeg		12	1		13
Napkin		13			13
Office	13				13
Orange		12	1		13
Orchid		13			13
Oyster		13			13
Pagoda	9			4	13
Palace	13				13
Parish	12		1		13
Parrot		13			13
Papaya		13			13
Peanut		13			13
Pebble		13			13
Pestle		11		2	13
Pellet		13			13
Pencil		13			13
Piazza	13				13

Pigeon		13			13
Pillar	13				13
Plaice		13			13
Planet	13				13
Plough	13				13
Poplar	13				13
Pocket		12	1		13
Potato		13			13
Cyprus	12		1		13
Priory	13				13
Prison	13				13
Quarry	11		2		13
Rabbit		13			13
Radish		13			13
Rafter	13				13
Rattle		13			13
Raisin		13			13
Rocket	10		3		13
Resort	13				13
Rosary		12	1		13
Ravine	13				13
Rodent		12	1		13
Runway		13			13
Sandal		13			13
Scarab		11		2	13
School	12		1		13
Saucer		13			13
Seesaw	13				13
Settee	13				13
Stable	13				13

Shrimp		13			13
Stream	11		2		13
Sleigh	13				13
Spider		13			13
Street	13				13
Studio	13				13
Suburb	13				13
Thread		12	1		13
Subway	13				13
Tassel		11	1	1	13
Tavern	13				13
Temple	13				13
Toilet	5	6	2		13
Tissue		13			13
Throne	10	3			13
Tomato		13			13
Tunnel	13				13
Turret	13				13
Valley	13				13
Walnut		13			13
Walrus	13				13
Wigwam	13				13
Waggon	10		1	2	13

Appendix C

Table 6: Word pairs and their BNC frequency ratings.

Semantic size pairs				Size neutral pairs			
Word 1	BNC frequency	Word 2	BNC frequency	Word 1	BNC frequency	Word 2	BNC frequency
forest	8679	mussel	201	method	9091	satire	196
jungle	984	radish	213	aiming	1003	rigour	203
armada	57	dahlia	59	myopic	71	abated	99
wigwam	12	beaker	181	abduct	30	candid	155
planet	2365	raisin	166	carbon	2462	fasten	148
bakery	299	papaya	17	notify	350	acidly	81
subway	154	napkin	277	florid	124	misled	292
dragon	370	bottle	5808	citing	344	murder	5854
tavern	378	limpet	49	evoked	352	barter	59
meadow	1104	spider	860	rested	1112	sticky	842
turret	176	beetle	510	donate	226	menace	514
dinghy	448	minnow	84	melted	556	haggle	57
prison	7177	clover	223	manner	6063	scenic	273
garden	13909	button	2428	agreed	14692	riding	2529
convoy	853	monkey	1008	cancel	892	legend	1245
valley	5401	pencil	1400	listen	5785	invest	1589
galaxy	985	camera	804	viable	970	poetic	746
cinema	2100	canary	271	biopsy	806	apathy	272
mosque	398	brooch	321	flawed	381	preach	303
tunnel	2692	insect	2120	export	2693	titles	2145
temple	2364	bucket	1401	excuse	3089	lively	1472
suburb	999	candle	1589	absurd	966	rhythm	1521
island	2120	cherry	990	talent	2095	motive	1042

circus	728	peanut	334	latent	655	refund	343
pillar	1026	hornet	75	amused	1046	mimics	82
runway	572	walnut	431	syntax	581	chilly	357
museum	6800	banana	936	memory	7604	exotic	1043
bridge	7532	shrimp	280	smiled	7607	glowed	312
chapel	2297	magpie	180	intend	2074	dreary	272
belfry	378	pellet	570	depart	446	midday	555
chalet	213	coupon	767	occult	226	waking	730
digger	159	pebble	468	darken	104	coldly	561
cavern	231	almond	570	rotate	236	parted	750
office	29943	tissue	2621	across	25202	bitter	2502
lagoon	265	dagger	347	stigma	275	grubby	307
settee	358	kitten	224	revise	362	wicker	226
walrus	62	helmet	856	neuter	35	rotten	807
bedsit	95	maggot	216	juggle	104	jagged	298
gazebo	44	eraser	33	emnity	104	levity	33
glider	593	saucer	499	outing	786	gamble	538
ghetto	310	rabbit	1960	gladly	299	divide	1773
rafter	188	lentil	76	fickle	128	panics	54
marina	398	tomato	1460	arable	433	ritual	1456
castle	5263	orchid	397	causes	4624	divert	433
hostel	669	bullet	1247	vector	658	polite	1174
arcade	377	muffin	82	unwise	413	graven	15
bazaar	203	bubble	799	mellow	233	locate	872
avenue	1841	potato	2524	remedy	1663	agenda	2350
clinic	2228	pigeon	861	namely	2164	typing	630
grotto	121	crocus	66	pallid	121	wallow	61
canyon	262	locket	81	primal	277	forego	70
seesaw	14	goblet	188	abides	14	injure	200
palace	4683	earwig	25	yellow	4553	peruse	31

campus	721	mitten	33	diving	686	lusted	13
church	24038	plaice	85	turned	24667	taunts	86
ravine	161	carrot	858	lament	164	vested	786
casino	255	flower	7267	elicit	245	active	7290
poplar	174	domino	176	disuse	122	caveat	120
cellar	944	bangle	47	admire	787	utters	40
closet	223	kettle	923	morbid	191	tribal	719

References

Allison, R. S., & Howard, I. P. (2000). Temporal dependencies in resolving monocular and binocular cue conflict in slant perception. *Vision research*, *40*(14), 1869-85.

Allison, R. S., Gillam, B. J., & Vecellio, E. (2009). Binocular depth discrimination and estimation beyond interaction space. *Journal of Vision*, *9*, 1-14.

Arsenault, R., & Ware, C. (2004). The importance of stereo and eye coupled perspective for eye-hand coordination in fish tank VR. *Presence: Teleoperators and Virtual Environments, 13,* 549–559.

Backus, B. T., Fleet, D. J., Parker, a. J., & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *Journal of neurophysiology*, *86*(4), 2054-68.

Baker, D. H., & Graf, E. W. (2009). On the relation between dichoptic masking and binocular rivalry. *Vision research*, *49*(4), 451-9.

Barbet, I., & Fagot, J. (2002). Perception of the corridor illusion by baboons (Papio papio). *Behavioural and Brain Research, 132,* 111-115.

Barbet, I., & Fagot, J. (2007). Control of the corridor illusion in baboons (Papio papio) by gradient and linear-perspective depth cues. *Perception*, *36*(3), 391-402.

Barlow, H.B., Blakemore, C., & Pettigrew, J.D. (1967). The neural mechanism of binocular depth discrimination. *J Physiol (Lond)*, *193,* 327–342, 1967.

Bates, D., Maechler, M., & Dai, B. (2008). lme4: Liner mixed-effects models using S4 classes. R package version 0.999375-22. Retrieved January 15, 2010, from http://lme4.r-forge.r-project.org/

Berryhill, M. E., & Olson, I. R. (2009). The representation of object distance: evidence from neuroimaging and neuropsychology. *Frontiers in human neuroscience*, *3*, 1-9.

Bertelson, J., & Morais, J. (1983). A Ponzo-like illusion left and right of fixation: a failed prediction. *Neuropsychologia*, *21*, 105-109.

Bishop, P.O. (1994). Size constancy, depth constancy and vertical disparities: a further quantitative interpretation. *Biological Cybernetics, 71*, 37–48.

Blake, R., & Boothroyd, K. (1985). The precedence of binocular fusion over binocular rivalry. *Perception & psychophysics*, *37*(2), 114-24.

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature reviews. Neuroscience*, *3*(1), 13-21.

Blakemore, B. C. (1970). The range and scope of binocular depth discrimination in man. *J. Physiol., 211*, 599-622.

Blythe, H. I., Liversedge, S. P., & Findlay, J. M. (2010). The effective fusional range for words in a natural viewing situation. *Vision research*, *50*, 1559-1570.

Breitmeyer, B., Julesz, B., & Kropfl, W. (1975). Dynamic random-dot stereograms reveal up-down anisotropy and left-right isotropy between cortical hemifields. *Science, 187,* 269-270.

Bruce, V., Green, P.R., & Georgeson, M.A. (2003). *Visual Perception*. Hove, UK: Psychology Press.

Bruner, J. S., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology, 42,* 33 – 44.

Brysbaert, M. (2007). *"The language-as-fixed-effect fallacy": Some simple SPSS solutions to a complex problem (version 2.0)*. Royal Holloway, University of London. Technical report.

Clem, R.K., & Pollack, R.H. (1975). Illusion magnitude as a function of visual field exposure. *Perception and Psychophysics, 17*(5), 450-4.

Collewijn, H. & Erkelens, C. J. (1990). Binocular eye movements and the perception of depth. Reviews of Oculomotor Research, 4, 213–261.

Coren, S., & Girgus, J. (1978). Seeing is deceiving: the psychology of visual illusions. Hillsdale, NJ: Erlbaum.

Cornell, E. D., Macdougall, H. G., Predebon, J., & Curthoys, I. S. (2003). Errors of binocular fixation are common in normal subjects during natural conditions. *Optometry and Vision Science, 80*, 764–771.

Cumming, B. G., & Parker, a. J. (1999). Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *19*(13), 5602-5618.

Cumming, B. G., & Parker, a. J. (2000). Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the Macaque. *The Journal of Neuroscience*, *20*(12), 4758-4767.

Damasio, H., Tranel, D., Grabowski, T. J., Adolphs, R., & Damasio, A. R. (2004). Neural systems behind word and concept retrieval. *Cognition, 92,* 179–229.

van Ee*, R.,* & Erkelens, C.R. (2010). Stereo-vision: Head-centric coding of retinal signals. Current Biology, 20, 567-568

Fang, F., Boyaci, H., Kersten, D., & Murray, S. O. (2008). Attention-dependent representation of a size illusion in human V1. *Current biology : CB*, *18*(21), 1707-12.

Fischer, M. H. (2001). Number processing induces spatial performance biases. *Neurology, 57*, 822-826.

Fineman, M. B. (1981). Complexity of context and orientation of figure in the corridor illusion. *Perceptual and Motor Skills, 53,* 11-14.

Fischmeister, F. P., & Bauer, H. (2006). Neural correlates of monocular and binocular depth cues based on natural images: a LORETA analysis. *Vision research*, *46*(20), 3373-80.

Foley, J.M., & Richards, W. (1972). Effects of voluntary eye movements and convergence on the binocular appreciation of depth. *Perception and Psychophysics, 11*, 423-427.

Fujita, K., Blough, D. S., & Blough, P. M. (1991). Pigeons see the Ponzo illusion. *Animal Learning and Behaviour, 19,* 283-293.

Fujita, K., Blough, D.S., & Blough, P. M. (1993). Effects of the inclination of context lines on perception of the Ponzo illusion by pigeons. *Animal Learning and Behavior, 21*, 29-34.

Gårding, J., Porrill, J., Mayhew, J. E., & Frisby, J. P. (1995). Stereopsis, vertical disparity and relief transformations. *Vision research*, *35*(5), 703-22.

Gillam, B. (*1980*). Geometrical illusions. *Scientific American, 242*, 102-111.

Glanville, A. D. (1933). The psychological significance of the horopter. *The American Journal of Psychology, 45*, 592-627.

Godard, O., & Fiori, N. (2010). Sex differences in face processing: Are women less lateralized and faster than men? *Brain and Cognition, 73*, 167-175.

Goodale, M. A., Milner, A. D., Jakobson, L. J., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature,* 349**,** 154– 156.

Greenberg, J. H. (1966). *Language universals, with special reference to feature hierarchies*. The Hague: Mouton.

Greene, E., & Gentner, S. (2001). Further consideration of size illusions in random dot stereograms. *Perceptual and motor skills*, *93*(1), 205-12.

Gregory, R. L. (1966). Eye and brain. London: World University Library.

Haber, R. N., & Levin, C. a. (2001). The independence of size perception and distance perception. *Perception & psychophysics*, *63*(7), 1140-52.

Harris, J. M., & Wilcox, L. M. (2009). The role of monocularly visible regions in depth and surface perception. *Vision research*, *49*(22), 2666-85.

Harris, M. G. (1994). Visual motion caused by self motion. In A. T. Smith & R. J. Snowden (Eds.), *Visual Detection of Motion*. London: Academic Press.

Hartung, B., Schrater, P. R., Bülthoff, H. H., & Franz, V. H. (2005). Is prior knowledge of object geometry used in visually guided reaching ? *Journal of Vision*, 504-514.

Helmholtz, H. (1925). Helmholtz's treatise on physiological optics. In J. P. C. Shouthall (Ed. & Trans.), *Handbuch der physiologischen optik (vol. 3)*. New York: Optical Society of America. (Original work published 1909).

Hillis, J. M., & Banks, M. S. (2001). Are corresponding points fixed? *Vision Research, 41*, 2457-2473.

Hoffman, D. M., & Banks, M. S. (2010). Focus information is used to interpret binocular images. *Journal of Vision*, *10*, 1-17.

Holway, A.H., & Boring, E.G. (1941). Determinants of apparent visual size with distance variant. American Journal of Psychology, 54, 21-37.

Houlard, N., Fraisse, P., Hecaen, H. (19676). Effects of unilateral hemispheric lesions on two types of optico-geometric illusions. *Cortex,12*(3), 232-240.

Howard, H. J. (1919). A test for the judgement of distance. *American Journal of Opthalmology, 2,* 656-675.

Howard, I.P., & Rogers, B.J., 1995. *Binocular vision and stereopsis*. Oxford University Press, New York.

Hudson, W. (1960). Pictorial depth perception in sub-cultural groups in Africa. *Journal of Social Psychology 52,* 183-208.

Hudson, W. (1962). Pictorial perception and educational adaptation in Africa. *Psychologia Africana, 9*, 226-239.

Ittelson, W. H. (1951). Size as a Cue to Distance. *American Journal of Psychology, 64*, 54-67.

Jakobson, R., & Halle, M. (1956). *Fundamentals of language*. The Hague: Mouton.

Jaschinski, W., Jainta, S., & Kloke, W. B. (2010). Objective vs subjective measures of fixation disparity for short and long fixation periods. *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)*, *30*(4), 379-90.

Johnson, B. W., McKenzie, K. J., & Hamm, J. P. (2002). Cerebral asymmetry for mental rotation: Effects of response hand, handedness and gender. *NeuroReport, 13,* 1929–1932.

Jones, R. K., & Lee, D. N. (1982). Why two eyes are better than one: The two views of binocular vision. *Journal of Experimental Psychology: Human Perception and Performance*, *7,* 30–40.

Julesz B. (1963). Stereopsis and binocular rivalry of contours. *J Opt Soc Am*. *53*, 994-999.

Julesz, B. (1964). Binocular Depth Perception without Familiarity Cues: Random-dot stereo images with controlled spatial and temporal properties clarify problems in stereopsis. *Science*, *145*(3630), 356-362.

Julesz, B. (1971). Foundations of Cyclopean Perception, Chicago University Press, Chicago.

Julesz, B., & Miller, J. E. (1975). Independent spatial-frequency-tuned channels in binocular fusion and rivalry. *Perception*, *4*(2), 125-143.

Kansaku, K., Yamaura, A., & Kitazawa, S. (2000). Sex differences in lateralization revealed in the posterior language areas. Cerebral *Cortex, 10*, 866–872.

Kaufman, L. (1964). On the nature of binocular disparity, *American Journal of Psychology, 11*, 393-402.

Kersten, D., & Murray, S. O. (2010). Vision: When Does Looking Bigger Mean Seeing Better? *Current Biology*, *20*(9), R398-R399. Elsevier Ltd. doi: 10.1016/j.cub.2010.03.021.

Klatzky, R. L. (1989). Can You Squeeze a Tomato? The Role of Motor Representations in Semantic Sensibility Judgments. *Journal of Memory and Language*, *28*, 56-77.

Koehler, W. and Wallach, H. (1944). Figural After-effects: An Investigation of Visual Processes. *Proceedings of the American Philosophical Society, 88,* 269-357.

Koenderink, J. J., van Doorn, a. J., & Kappers, a. M. (1996). Pictorial surface attitude and local depth comparisons. *Perception & psychophysics*, *58*(2), 163-73.

Lansdell, H. (1962). A sex difference in effect of temporal lobe neurosurgery on design preference. *Nature, 194,* 852–854.

Lehmkuhle, S., & Fox, R. (1980). Effect of depth separation on metacontrast masking. *Journal of experimental psychology. Human perception and performance*, *6*(4), 605-21.

Leibowitz, H.W., Shina, K., & Hennessy, R.T. (1972). Oculomotor adjustments and size constancy. *Perception and Psychophysics, 12,* 497-500.

McGlone, J. (1980).Sex differences in human brain asymmetry: a critical survey. *Behavioral and Brain Sciences, 3,* 215-262.

McKee, S. P., Welch, L., Taylor, D. G., & Bowne, S. F. (1990). Finding the common bond: Stereoacuity and the other hyperacuities. *Vision Research, 30*, 879–891.

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)*, *204*(1156), 301-28.

Marshall, J. a., Burbeck, C. a., Ariely, D., Rolland, J. P., & Martin, K. E. (1996). Occlusion edge blur: a cue to relative visual depth. *Journal of the Optical Society of America. A, Optics, image science, and vision*, *13*(4), 681-8.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology, 58*, 25-45.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology, 11,* 194–201.

Mather, G. (2006). *Foundations of perception*. Hove, U.K.: Psychology Press, Erlbaum.

Matthews, N., Meng, X., Xu, P., & Qian, N. (2003). A physiological theory of depth perception from vertical disparity. *Vision research*, *43*(1), 85-99.

Mon-Williams, M., & Tresilian, J. R. (2000). Ordinal depth information from accommodation. *Ergonomics, 43*, 391-404.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. Humphreys (Eds.), Basic processes in reading: Visual word recognition (pp. 148-197). London: Lawrence Erlbaum Associates.

Morgan, M. (2003). *The space between our ears: How the brain represents visual space*. London: Weidenfeld & Nicolson.

Murray, S.O., Boyaci, H., and Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience, 9*, 429–434.

Murray, S.O., & He, S. (2006). Contrast invariance in the human lateral occipital complex depends on attention. *Current Biology,16,* 606–611.

Newhouse, M., & Uttal, W. R. (1982). Distribution of stereoanomalies in the general population. *Bulletin of the Psychonomic Society*, 48-50.

Nikara, T., Bishop, P. O., & Pettigrew, J. D. (1968). Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *6*(4), 353-72.

Ogle, K. N. (1932). Analytical treatment of the longitudinal horopter; its measurement and application to related phenomena, especially to relative size and shape of ocular images. *J. opt. Soc. Amer, 22,* 665-728.

Ogle, K. N. (1950). *Researches in binocular vision*. Philadelphia: Saunders.

Ooi, T. L., Wu, B., & He, Z. J. (2001). Distance determined by the angular declination below the horizon. *Nature*, *414*(6860), 197-200.

Orban, G. a., Janssen, P., & Vogels, R. (2006). Extracting 3D structure from disparity. *Trends in Neurosciences*, *29*(8), 466-73.

O'Shea, R. P., Blackburn, S. G., & Ono, H. (1994). Contrast as a depth cue. *Vision Research, 34,* 1595-1604.

den Ouden, H. E., van Ee, R., & de Haan, E. H. (2005). Colour helps to solve the binocular matching problem. *The Journal of physiology*, *567*(Pt 2), 665-71.

Panum P. L. (1858). *Physiologische Untersuchungen iiber das Sehen mit zwei Augen*. (Kiel: Schwerssche Buchhandlung)

Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature reviews. Neuroscience*, *8*(5), 379-91.

Patterson, R. (1997). Visual processing of depth information in stereoscopic displays. *Science*, *17*, 69-74.

Patterson, R., & Fox, R. (1984). The effect of testing method on stereoanomaly. *Vision Research*, *24*(5), 403-408.

Pettigrew, J. D., Nikara, T., & Bishop, P. O. (1968). Binocular interaction on single units in cat striate cortex: simultaneous stimulation by single moving slit with receptive fields in correspondence. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *6*(4), 391-410.

Poggio, G. F., Gonzalez, F., and Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *The Journal of Neuroscience, 8*, 4531-4550.

Poggio, G.F., & Talbot, W.H. (1981). Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *J Physiol,* 315, 469 – 492.

Pollack, P. (1955). A note on monocular depth perception. The American Journal of Psychology, 68, 315-318.

Previc, F.H. (1990). Functional specialization in the lower and upper visual fields in humans: its ecological origins and neurological implications. *Behavioural and Brain Sciences, 13,* 519-575

Purghé, F., & Coren, S. (1992). Amodal completion, depth stratification, and illusory figures: a test of Kanizsa's explanation. *Perception, 21*(3), 325-335.

Qian, N. (1997). Binocular disparity and the perception of depth. *Neuron*, *18*(3), 359-68.

Quinlan, D. J., and Culham, J. C. (2007). fMRI reveals a preference for near viewing in the human parieto- occipital cortex. *Neuroimage, 36*, 167–187.

R Development Core Team. (2008). R: A language and environment for statistical computing. Retrieved January 15, 2010, from R Foundation for Statistical Computing: http://www.R-project.org

Rasmjou, S., Hausmann, M., & Güntürkün, O. (1999). Hemispheric dominance and gender in the perception of an illusion. *Neuropsychologia*, *37*(9), 1041-7.

Read, J. C., & Cumming, B. G. (2006). Does depth perception require vertical-disparity detectors? *Journal of vision*, *6*(12), 1323-55.

Richards, W. (1970). Stereopsis and stereoblindness. *Experimental brain research. Experimentelle Hirnforschung*. *Expérimentation cérébrale*, *10*(4), 380-8.

Richards, W. (1971). Anomolous stereoscopic depth perception. *Journal of the Optical Society of America, 61,* 410.

Richards, W. (1977). Stereopsis with and without monocular contours. *Vision Research, 17, 967-969*.

Ross, H.E. & *Plug*, C. (1998). The mystery of the moon illusion: Exploring size perception. *Human Performance in Extreme Environments, 3*(1), 34-36.

Rubinsten, O., & Henik, A. (2002). Is an ant larger than a lion? *Acta psychologica*, *111*(1), 141-54.

Schreiber, K. M., Hillis, J. M., Filippini, H. R., Schor, C. M., & Banks, M. S. (2008). The surface of the empirical horopter. *Journal of Vision*, *8*, 1-20.

Sedgwick, H. A. (1986). Space perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 1. Sensory processes and perception* (pp. 21.1-21.57). New York: Wiley.

Sereno, S. C., O'Donnell, P. J., & Sereno, M. E. (2009). Size matters: bigger is faster. *Quarterly journal of experimental psychology (2006)*, *62*(6), 1115-22.

Servos P, Goodale MA, Jakobson LS (1992) The role of binocular vision in prehension: a kinematic analysis. *Vision Research, 32,* 1513-1521.

Servos, P., Goodale, M. A., & The, A. (1994). Binocular vision and the on-line control of human prehension. *Experimental Brain Research*, (April 1993), 119-127.

Setti, A., Caramelli, N., & Borghi, A. (2008). Conceptual information about size of objects in nouns. *European Journal of Cognitive Psychology*, *21*(7), 1022-1044.

Shoben, E. J., & Wilson, T. L. (1998). Categorization in judgments of relative magnitude. *Journal of Memory and Language, 38*, 94-111.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*, 153–156.

Stevens, K. A., & Brookes, A. (1988). Integrating stereopsis withmonocular interpretations of planar surfaces. *Vision Research, 28,* 371-386.

Su, Y., He, Z. J., & Leng, T. (2009). Coexistence of binocular integration and suppression determined by surface border information. *PNAS*, 1-6.

Tam, W. J., & Stelmach, L. B. (1998). Display duration and stereoscopic depth discrimination. *Canadian Journal of Experimental Psychology, 52*(1), 56-61.

Tan, L.H., Liu, H.L., Perfetti, C.A., Spinks, J,A., Fox, P.T., & Gao, J.H. (2001). The neural system underlying Chinese logograph reading. *NeuroImage, 13*, 836-846.

Thomas, O.M., Cumming, B.G., and Parker, A.J. (2002). A specialization for relative disparity in V2. *Nature Neuroscience, 5*, 472–478.

Timney, B., & Keil, K. (1996). Horses are sensitive to pictorial depth cues. *Perception, 25,* 1121-1128.

Uttal, W. R., Davis, N. S., & Welke, C. (1994). Stereoscopic perception with brief exposures. *Perception & psychophysics*, *56*(5), 599-604.

Vigliocco, G., Vinson, D. P., Arciuli, J., & Barber, H. (2008). The role of grammatical class on word recognition. *Brain and Language*, *105*, 175-184.

Viguier, A., Clement, G., Trotter, Y. (2001). Distance perception within near visual space. *Perception, 30*, 115-124.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences, 7,* 483–488.

Ward, L. M., Porac, C., Coren, S., and Girgus, J. S. (1977). The case for misapplied constancy scaling: Depth associations elicited by illusion configurations. *American Journal of Psychology, 90*, 609-620.

Wesp, R., Peckyno, A., McCall, S., & Peters, S. (2000). Object recognition may distort size perception. *Perceptual & Motor Skills, 90,* 803-809.

Westheimer, G. (1979). Brain Involved in Stereoscopic Acuity. *Experimental Brain Research*, *597*, 585-597.

Wheatstone, C. (1838). Contributions to the physiology of vision. Part the First. On some remarkable, and hitherto unobserved, phaenomena of binocular vision. *Philosophical Transactions of the Royal Society of London, 128*, 371-394.

Wolfe, J.M., *et al*. (2009) *Sensation and Perception* (2nd edition). Sunderland, MA: Sinauer Associates Inc.

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

Zhang, Z., Cantor, C. R., & Schor, C. M. (2010). Perisaccadic Stereo Depth with Zero Retinal Disparity. *Current Biology*, 1176-1181.

Zwann, R. A. (2004). The immersed experience: Toward an embodied theory of language comprehension. In B. Ross, (Ed.), The Psychology of Learning and Motivation Vol. 44, New York: Academic Press.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Do language comprehenders routinely represent the shapes of objects? Psychological Science, 13, 168–171.

Zwaan, R.A., & Yaxley, R.H. (*2003*). Spatial iconicity affects semantic-relatedness judgments. *Psychonomic Bulletin & Review, 10,* 954-958